Research Report

WHEN DOES HUMAN OBJECT RECOGNITION USE A VIEWER-CENTERED REFERENCE FRAME?

Michael J. Tarr[a] and Steven Pinker[b]

[a]*Department of Psychology, Yale University and* [b]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*

Abstract—*How do people recognize an object in different orientations? One theory is that the visual system describes the object relative to a reference frame centered on the object, resulting in a representation that is invariant across orientations. Chronometric data show that this is true only when an object can be identified uniquely by the arrangement of its parts along a single dimension. When an object can only be distinguished by an arrangement of its parts along more than one dimension, people mentally rotate it to a familiar orientation. This finding suggests that the human visual reference frame is tied to egocentric coordinates.*

Object constancy, the ability to recognize an object despite changes in its retinal image produced by displacements and rotations, is an important problem in both human vision and computer vision systems (Marr, 1982; Rock, 1983). A prominent proposal by Marr and Nishihara (1978) is that the visual system first aligns a coordinate system on an input object based on its axes of symmetry, elongation, or movement, describes the arrangement of the object's parts within that system (resulting in the same description regardless of the object's orientation relative to the viewer), and matches the description against memory representations stored in the same format. An alternative is that the input is transformed into a canonical orientation and then is matched against a representation in memory of the appearance of the object in that orientation (Rock, 1974; Tarr & Pinker, 1989). We present data from experiments designed to determine if and when people use such mechanisms. The experiments rely on the discovery by Shepard and his collaborators (Shepard & Cooper, 1982; Shepard & Metzler, 1971) that humans possess an analogue visual transformation process, "mental rotation." The principal empirical signature of mental rotation is that people take more time to classify a shape that is oriented farther from the upright, and take more time to match two objects that differ by greater orientation differences. Other evidence confirms that this chronometric pattern reflects an incremental rotation process. For example, during the interval between stimulus presentation and the response, the subject can quickly classify a probe stimulus displayed at an intermediate orientation; the optimal intermediate orientation changes continuously during the stimulus-response interval. There is also evidence from single-cell recordings in the monkey motor cortex for an analogous transformation process in the motor planning system (Georgopoulos, Lurito, Petrides, Schwartz, & Massey, 1989).

The fact that mental rotation exists, however, does not mean that it is used to recognize objects. Most mental rotation tasks require the subject to discriminate shapes from their mirror-images. It is possible that people generally use object-centered coordinate systems to recognize shapes, but that such coordinate systems can be of either handedness, so that objects and their mirror-images have equivalent representations. If coordinate systems are not explicitly labeled as right-handed or left-handed, mental rotation would be needed when handedness must be discriminated (and only then). Input shapes would be rotated into alignment with the up–down axis of the perceiver, so that the right and left sides of the shape would align with the right and left sides of the person, which are explicitly labeled as "right" and "left," making the handedness discrimination possible (Corballis, 1988; Hinton & Parsons, 1981).

In fact when subjects are simply required to name objects, orientation effects on response time greatly diminish (Corballis, Zbrodoff, Shetzer, & Butler, 1978). However, such results are inconclusive. If shapes can be discriminated on the basis of orientation-independent local features, such as a curved segment present in only one object, subjects could name objects via this shortcut. Furthermore, the alphanumeric characters typically used are highly overlearned and might be stored in multiple representations, each specific to an orientation, so input shapes at any of these orientations could be matched directly in constant time (Jolicoeur, 1985; Tarr & Pinker, 1989).

We present data from experiments that avoid these problems. Subjects learned names for three novel shapes (a subset of those shown in Fig. 1A), each studied only at the upright orientation. The seven shapes were composed of similar configurations of line segments in different spatial arrangements, so no local feature could serve as a unique cue. No shape was the mirror-image of any other and each shape had a clearly marked base and vertical axis, minimizing the time needed to locate the shape's intrinsic axis and bottom. Subjects saw the shapes on a CRT at different orientations and identified them by pressing one of three buttons labeled with the shape names. On 25% of the trials one of the other four shapes in Figure 1A was presented, and subjects pressed a foot pedal.[1]

Results, shown in Figure 2A (test orientations at 0°, 45°, −90°, and 135°),

Correspondence and reprint requests to: Michael J. Tarr, Department of Psychology, Yale University, PO Box 11A Yale Station, New Haven, CT 06520.

1. Nine subjects learned the shapes by tracing them and then drawing them from memory. Different subsets of shapes were taught to different subjects. The 3 target shapes were shown 8 times in the orientations 0°, 45°, −90°, and 135°, and the 4 distractors were shown 2 times at these orientations, for a total of 128 trials, preceded by 12 practice trials.

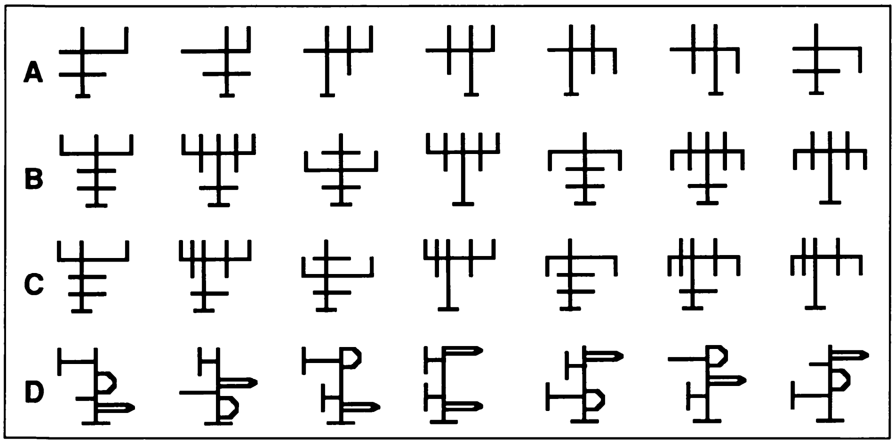

Fig. 1. Shapes presented to subjects for identification. (A) Asymmetrical shapes. (B) Symmetrical shapes. (C) Skewed-symmetrical shapes. (D) Bilaterally redundant shapes.

suggest that people employed mental rotation to recognize the shapes: Recognition time was linearly related to orientation from the upright, and the .95 confidence interval for the obtained slope of 2.42 ms/deg (an estimate of the rotation rate) includes the slope values obtained in the Cooper and Shepard experiments,[2] but does not include zero.[3,4]

The fact that people show orientation effects even when the task does not require handedness to be assigned is evidence that mental rotation, not the computation of an object-centered viewpoint-independent description, is the mechanism used.[5] An alternative is that subjects attempted to determine the handedness of the stimuli despite its irrelevance to the task, anticipating that there might be mirror-image distractors. This alternative can be eliminated by an experiment (originally reported as Condition 0/105/−150 of Experiment 3 in Tarr & Pinker, 1989) in which subjects saw both handedness versions of the shapes, and were required to ignore the difference, identifying each object and its mirror image by the same response.[6] Here handedness information is by definition irrelevant to the task. Nonetheless, orientation effects were found once again for standard versions of the shapes (Fig. 2B; test orientations at 0°, 105°, and −150°; slope for standard versions: 3.65 ms/deg).[7]

Under some circumstances, however, orientation-invariant recognition does occur. The same experimental procedure run with symmetrical shapes (Fig. 1B), which cannot be assigned a handedness, shows that people can recognize such shapes equally quickly at all orientations[8] (Fig. 2C; test orientations at 0°, 45°, −90°, and 135°; slope: 0.63 ms/deg). This is not an effect of geometric symmetry itself. Additional experiments show that the crucial property is that one side be redundant with the other, so only a single side need be examined to discriminate among shapes within the set. When the shapes are skewed so that they are not symmetrical (Fig. 1C), there is still no effect of orientation (Fig. 2D; test orientations at 15° and 120° or −15° and −120°; slope: 0.29 ms/deg). Even when there is no similarity between the shapes' right and left sides, but the arrangement of parts on each side is unique to that shape so only a single side need be examined to identify it, as in Fig. 1D, there are no effects of orientation (Fig. 2E; test orientations at 15° and 120° or −15° and −120°; slope: 0.18 ms/deg).[9,10]

2. Mean slopes ranged from 1.61 ms/deg to 3.06 ms/deg in the summary of experiments compiled by S. Shepard and D. Metzler (1988).

3. Error rates for the four orientations were 5%, 3%, 6%, and 6%.

4. Orientation effects in this experiment cannot be attributed to the prominent vertical axis of the shapes being aligned with subjects' retinal or head axis during the initial teaching of the shapes. Tarr and Pinker (1989) found effects of orientation on naming times even when shapes were taught at an orientation of 15°, which did not coincide with the subjects' retinal or head-defined upright (Experiments 3 and 4; Condition 15/120).

5. There is no paradox in the suggestion that people know the direction in which to rotate an object before they have recognized it. For example, if three noncollinear landmarks can be extracted from the input shape independently of orientation, and analogous landmarks are indicated in memory representations, the optimal axis and direction of rotation can be computed, though the degree of shape match for the rest of the object cannot be assessed until the transformation is executed. See Ullman (1989) and Tarr and Pinker (1989) for discussion.

6. The procedure was identical to that of Experiment 1 except that 13 subjects were run, both handedness versions of each shape were shown (consequently half of the trials presented mirror images of the shapes), and a different and smaller set of orientations were used. For the data on the recognition of mirror image versions, which are not relevant here, see Experiment 3 of Tarr and Pinker (1989).

7. Orientation effects in this experiment cannot be attributed to subjects' not having had sufficient practice to realize that each shape and its mirror-image were to be treated equivalently. After the trials reported, an additional 1408 trials were administered, followed by 768 trials in which the shapes were shown at 24 new orientations separated by 15° increments. We found comparable effects on recognition time of the difference between the stimulus orientation and the nearest well-learned orientation (slope for standard versions = 4.14 ms/deg). This shows that rotation was still necessary for shapes in new orientations even after extensive practice at treating the shape and its mirror image as equivalent. See Tarr and Pinker (1989) for details.

8. The slope of the line shown in Figure 2C (0.63 ms/deg) is significantly different from the slope of the line shown in Figure 2A (2.42 ms/deg; $F(1, 19) = 5.18, p < .05$) and from the slope of the line shown in Figure 2B (3.65 ms/deg; $F(1, 23) = 12.2, p < .01$).

9. Apart from the specified changes in the stimuli, the use of new orientations, and slight variations in the number of trials, the method was unchanged from previous experiments.

10. The orientation-independence effect holds not only for the set of orientations shown in Figure 2C–E, but for a larger set of orientations, presented to the subjects for the first time after they had undergone many more trials (>1000). Slopes for new orientations were: 0.53 ms/deg (symmetrical shapes); 0.57 ms/deg (skewed-symmetrical); 1.07 ms/deg (bilaterally redundant).

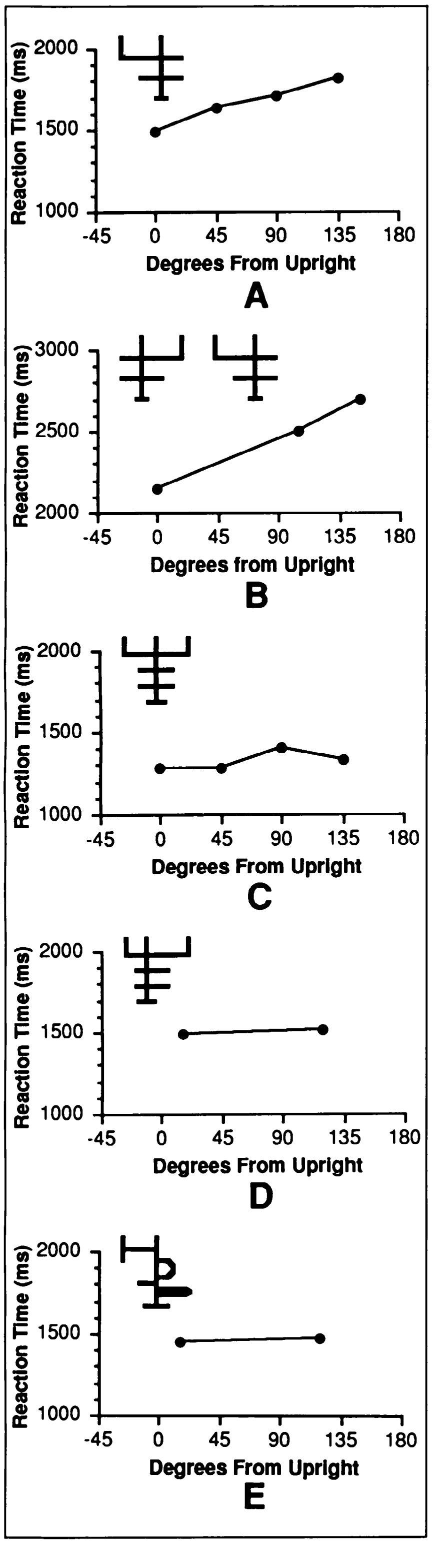

Fig. 2. Response times to identify shapes as a function of orientation. (A) Asymmetrical shapes. (B) Asymmetrical shapes and their mirror images. (C) Symmetrical shapes. (D) Skewed-symmetrical shapes. (E) Bilaterally redundant shapes.

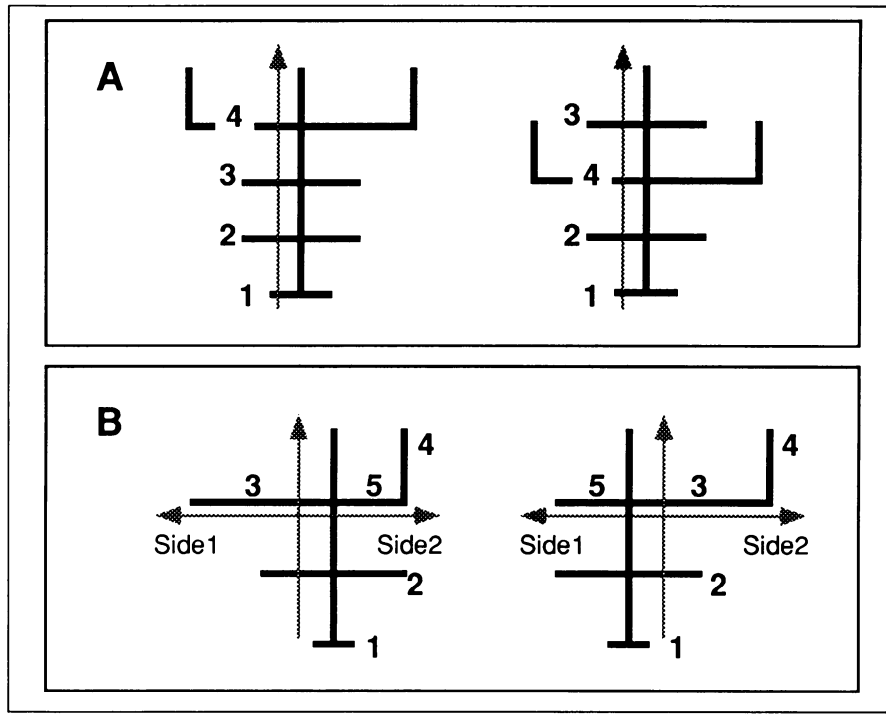

Fig. 3. One-dimensional and two-dimensional descriptions of shapes. (A) A pair of shapes where one-dimensional descriptions are sufficient for distinguishing between them. (B) A pair of shapes where one-dimensional descriptions are insufficient to make the discrimination, necessitating the use of two-dimensional descriptions.

Why is mental rotation needed to recognize asymmetrical shapes, even when handedness is irrelevant to recognition, but not needed to recognize shapes whose two sides are redundant? For the symmetrical, skewed-symmetrical, and bilaterally redundant shapes, it is sufficient to keep track of the one-dimensional ordering of parts on either side of the shape from bottom to top. For example, "two small crossbars underneath a longer crossbar with upright bends" is sufficient to discriminate the first shape from the second shape in Figure 3A. This suggests that perceivers can assign a one-dimensional vector to a shape's axis defining a top-to-bottom ordering of parts equally quickly regardless of the shape's orientation. In contrast, shapes requiring rotation have parts whose locations must be specified along two dimensions simultaneously. For example, to identify the first shape in Figure 3B, the perceiver must encode the fact that the top crosssbar is shorter on one side of the shape and longer on the other, and that the right-angle upward bend is on the side with the short crossbar segment. Absolute handedness information is not required: it does not matter whether the first side is remembered as the right side and the second as the left or vice-versa. But discriminating between sides is required: it matters that the side with the long crossbar segment is remembered as being a *different* side than the side with the bend. This is necessary in order that the shape not be confused with the second shape in Figure 3B, which also has a crossbar that is longer on one side than the other, but in which it is the side with the longer crossbar segment that has the bend. Thus the mere requirement that two sides be kept distinct is enough to require that subjects mentally rotate.

This result suggests that the part of the visual system subserving object recognition lacks an object-centered 2D (and presumably 3D) coordinate system. The reference frame aligned with the viewer's egocentric upright, with its body-defined up–down and right–left directions, is the only one in which two dimensions are simultaneously specified. There is an object-centered mode of description, but it is insufficient to represent the arrangement of parts along two dimensions simultaneously; all that can be specified is the order of parts along a single foreaft dimension (or at most, the distance of each part from the midline along an independent medial-lateral "half-dimension," but with no specification of separate sides).

These data do not imply that all misoriented objects are recognized through mental rotation. For highly familiar objects, multiple orientation-specific representations can be directly matched against the input (Jolicoeur, 1985; Tarr & Pinker, 1989), and for many objects, sets of distinctive features or parts may suffice for identification. Even different objects composed of the same parts may be distinguished without mental rotation if the objects differ in how their parts are arranged along a single dimension. But determining the 2D and 3D relational structure of an object appears to require that the object be represented in a familiar orientation with respect to the viewer's upright. It is unclear how many cases of object recognition in natural settings require the computation of multidimensional spatial relations. Jolicoeur (1985) found that pictures of everyday objects are recognized more slowly as they are misoriented farther from the upright. Perhaps this is because many common shapes, although symmetrical about one axis, are not symmetrical about their other axes. For example, to recognize quadrupeds depicted in side view line drawings, one must encode properties of the head, tail, and limbs, which in the general case are fully distinguished only by their positions both along the foreaft axis and above or below it. This suggests that mental rotation may not be an uncommon strategy for recognizing misoriented complex objects.

Acknowledgments— We thank Jigna Desai and Greg Wolff for assistance and Irving Biederman and Pierre Jolicoeur for their helpful comments. Supported by NSF Grant BNS 8518774.

REFERENCES

Corballis, M.C. (1988). Recognition of disoriented shapes. *Psychological Review, 95*(1), 115–123.

Corballis, M.C., Zbrodoff, N.J., Shetzer, L.I., & Butler, P.B. (1978). Decisions about identity and orientation of rotated letters and digits. *Memory & Cognition, 6*, 98–107.

Georgopoulos, A.P., Lurito, J.T., Petrides, M., Schwartz, A.B., & Massey, J.T. (1989). Mental rotation of the neuronal population vector. *Science, 243*, 234–236.

Hinton, G.E., & Parsons, L.M. (1981). Frames of reference and mental imagery. In J. Long & A. Baddeley (Eds.), *Attention and performance IX*. Hillsdale, NJ: Erlbaum.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13*(4), 289–303.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.

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

Rock, I. (1974). The perception of disoriented figures. *Scientific American, 230*(Jan), 78–85.

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.

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

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703.

Shepard, S., & Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 3–11.

Tarr, M.J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology, 21*(2), 233–282.

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

(RECEIVED 9/28/89; ACCEPTED 12/5/89)