Research Report

THE VISUAL SYSTEM'S MEASUREMENT OF INVARIANTS NEED NOT ITSELF BE INVARIANT

Johan Wagemans,[1] Luc Van Gool,[2] and Christian Lamote[1]

[1]*Laboratory of Experimental Psychology and* [2]*ESAT-MI2, University of Leuven, Belgium*

Abstract—*When two shapes that differ in orientation or size have to be compared or objects have to be recognized from different viewpoints, the response time and error rate are systematically affected by the size of the geometric difference. In this report, we argue that these effects are not necessarily solid evidence for the use of mental transformations and against the use of invariants by the visual system. We report an experiment in which observers were asked to give affine-invariant coordinates of a point located in an affine frame defined by three other points. The angle subtended by the coordinate axes and the ratio of the lengths of their unit vectors systematically affected the measurement errors. This finding demonstrates that the visual system's measurement of invariants need not itself be invariant.*

An important problem for visual perception is how to establish a constant visual world from the continuously changing available information. For objects to be recognized, for example, the visual system must somehow deal with the changing projections depending on the point of observation. There is considerable debate about how this is done (see Tarr, 1995, for a review). According to one approach to shape constancy, the perceptual system makes use of features of the projected image or attributes of the optic array that remain unchanged, or invariant, under changes in viewpoint (e.g., Gibson, 1950, 1979). Despite some psychophysical evidence supporting this position (e.g., Cutting, 1986; Pizlo, 1994) and the current popularity of invariants in computer vision (e.g., Mundy & Zisserman, 1992; Van Gool, Moons, Pauwels, & Wagemans, 1994), the dominating belief seems to be that object recognition cannot be based on invariants because objects are harder to recognize from some viewpoints than from others. Typically, increasing recognition latencies and error rates are observed with an increasing orientation difference between a previously learned or standard orientation of an object and a subsequently viewed version of it (e.g., Cooper, 1976; Jolicoeur, 1985; Jolicoeur & Landau, 1984). These results have been interpreted as solid empirical evidence for an alternative class of theories according to which different views of objects are matched through a mental transformation or normalization process (e.g., Tarr & Pinker, 1989, 1990; Ullman, 1989). This view makes extensive reference to the way Shepard and Cooper (1982) interpreted the effects of orientation disparity in handedness discrimination tasks, namely, as evidence for mental rotation.

In this report, we argue that the frequently observed effects of parametric differences on the difficulty of the matching task (as measured by response times and error rates) need not be solid evidence against the visual system's use of invariants. For three-dimensional (3-D) objects, Biederman and Gerhardstein (1993) argued that the effects of viewpoint might be caused by the occlusion of different parts of an object or by the disappearance of nonaccidental properties, which are critical to determine the part category to which each part belongs (see also Farah, Rochlin, & Klein, 1994; Tarr & Bülthoff, 1995). For two-dimensional objects, the problem may be even more basic. Consider Figure 1a, which presents the projections of two planar shapes. No information is available on their 3-D orientation and position (together referred to as *pose*). If we assume pseudo-orthographic projection (no perspective), could these projections have resulted from the same shape? According to the mental transformation approach, the visual system is capable of simulating in 3-D space paths that correspond to combinations of 3-D rotations and translations of one projection, and then deciding whether there is a path that works out well and yields the other projection. In that case, the two projections are affine equivalent, which means that one can be mapped onto the other by a plane affine transformation. According to the invariants-based approach, in contrast, the visual system is capable of finding features that are invariant under the group of transformations that relate both images, which in this case are affine invariants.

In a recent series of experiments, participants were asked to match dot-pattern versions of these patterns (with dots at the vertices, one of which was marked as a reference point) under affine transformations (Wagemans, Van Gool, Lamote, & Foster, 1996). Results demonstrated that the task could be done reasonably well (i.e., from 75% to 95% correct identifications, depending on the conditions), even with patterns that contained minimal information, but the evidence was mixed regarding the theoretical controversy between the mental transformation approach and the invariants-based approach. On the one hand, the elimination of one of the transformation components (i.e., tilt) did not result in any appreciable improvement of general performance level (i.e., around 90% correct identifications in both conditions). It is difficult to reconcile this result with the use of mental transformations. On the other hand, performance was modulated rather strongly by some of the affine transformation parameters in some of the experiments (e.g., response times increased from 2 s at 0° to 3.5 s at 180° rotation and error rates increased from 7% at 0° to 20% at 60° slant). One would not expect these effects from an invariants-based approach.

Although the perceptual effects of the transformation parameters seem to argue against the invariant nature of the visual processing of shape equivalence, they do not rule out that invariants are used. Let us try to clarify this point for a particular type of invariants, affine coordinates, which could be used to solve the problem of affine shape equivalence illustrated in Fig-

Address correspondence to Johan Wagemans, Laboratory of Experimental Psychology, University of Leuven, Tiensestraat 102, B-3000 Leuven, Belgium; e-mail: johan.wagemans@psy.kuleuven.ac.be.

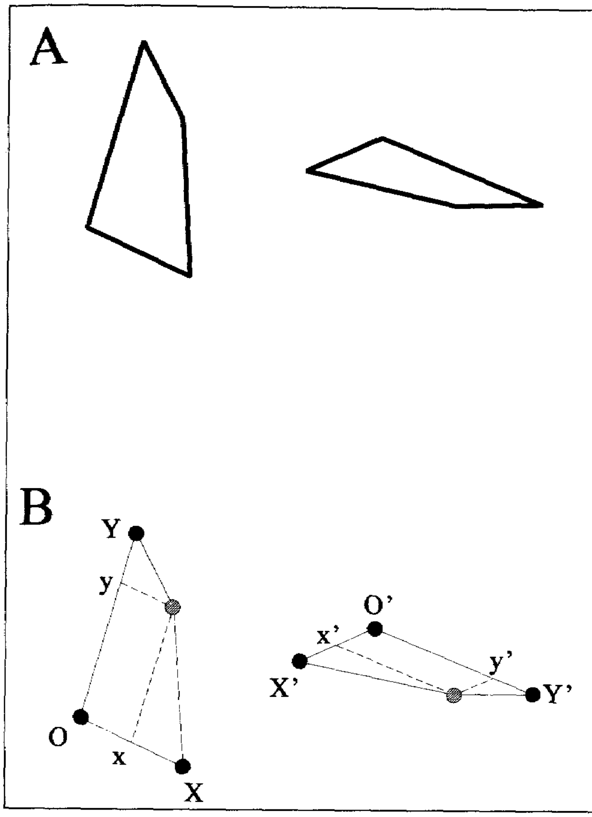

Fig. 1. Two simple shapes related by an affine transformation (a) and a demonstration of how affine-invariant coordinates can be used to determine the affine shape equivalence of such patterns (b). See the text for more details.

ure 1a. A triple of points suffices to define an affine coordinate frame (Koenderink & van Doorn, 1991; Ullman, 1989). One of the points plays the role of origin, while the other two define the coordinate axes and the unit lengths to be applied. Any additional point can then be given affine coordinates, following the construction of Figure 1b. Consider first the quadruple of dots drawn at the left. Suppose we take the dot in the lower left corner as the origin O, the one on the right as X, and the one in the upper left corner as Y. Draw a line from the hatched dot to OX, parallel to OY, and another line from the hatched dot to OY, parallel to OX. This yields two coordinates, x and y, that can be expressed as fractional numbers, Ox/OX and Oy/OY, respectively (0.50 and 0.75 for the example in Fig. 1b). These coordinates are affine invariant: The same fractions are obtained for all affine-equivalent patterns (e.g., in the pattern on the right of Fig. 1b, $O'x'/O'X'$ is also 0.50 and $O'y'/O'Y'$ is also 0.75). Because the coordinates are defined relative to the OX and OY lengths, OX and OY are called unit vectors. The geometric construction underlying the definition of affine-invariant coordinates makes use of two well-known affine-invariant properties, namely, parallelism of lines and relative distances between three collinear points.

The fact that such affine coordinates are affine invariants need not imply that their extraction from a pattern will always take the same amount of time or be equally accurate. First, there is the problem of selecting the same points as basis and using them in the same role (i.e., as origin or as defining the axes). Even with minimal patterns, the facilitation of finding the basis-point correspondences enhanced performance of subjects in detecting affine shape equivalence (Wagemans et al., 1996). In realistic, more complex shapes, the problem of choice is, of course, much larger. Second, it is fair to suspect that the extraction of the affine coordinates will be easier for the pattern at the left in Figure 1 than for the pattern at the right, which is a particularly oblique view.

Because we did not know of any empirical evidence demonstrating such effects directly, in our experiment we instructed subjects explicitly to give affine-invariant coordinates. To disentangle the point search and coordinate measurement problems as much as possible, we used patterns consisting of four points only, and three points were indicated explicitly and unambiguously as basis points (assuming no reflections). We also manipulated the configurations systematically to investigate whether the angle subtended by the coordinate axes and the projected unit lengths affected the accuracy of the subjects' measurements. If this were the case, the results would constitute good empirical support for the influence of object pose on the estimation of affine coordinates and for the more general thesis that the visual system's measurement of affine invariants need not itself be invariant. It would also follow that the frequently reported effects of parametric differences between shapes on the difficulty to assess their shape equivalence need not per se reflect the use of mental transformations.

METHOD

Subjects

Fifteen naive undergraduate psychology students at the University of Leuven were recruited in partial fulfillment of a course requirement.

Stimuli

All patterns presented to the subjects contained four dots, one blue, one red, and two black (see Fig. 2a for examples in black and white). The blue dot (shown by open circles in Fig. 2a) indicated the origin of the coordinate system; the line segments from it to the black dots defined the unit vectors of the OX and OY coordinate axes. The vector closest to horizontal had to be taken as the OX axis; the one closest to vertical was the OY axis. The red dot (shown by hatched circles in Fig. 2a) was located randomly in the parallelogram defined by the OX and OY unit vectors, except that locations within a 10-pixel zone around the axes were avoided.

Three different orientations and lengths of the OX and OY unit vectors were used (see Fig. 2b): (a) OX was either horizontal (0°) or 30° away from horizontal, either clockwise or counterclockwise; (b) OY was either vertical (90°), 60°, or 120°; (c) OX was 225 pixels (7.5 cm) long or 75 pixels shorter or longer; and (d) OY was 150 pixels (5 cm) long or 50 pixels

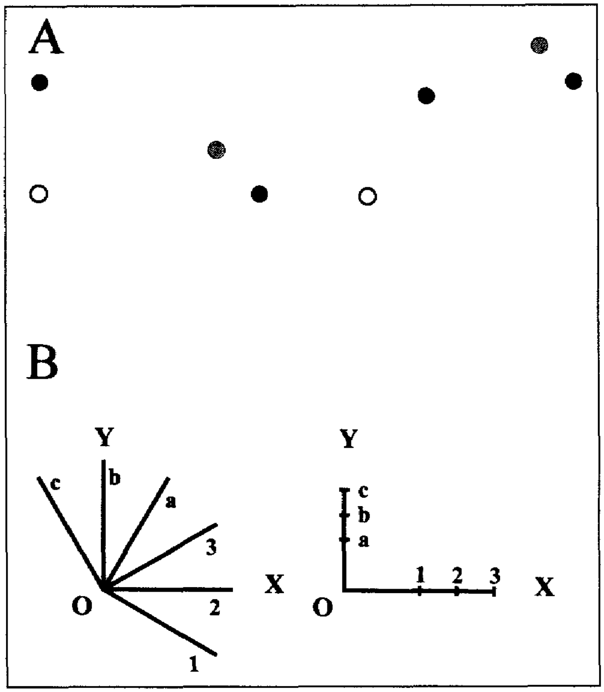

Fig. 2. The experimental stimuli. Two example stimuli (shown here in black and white) are illustrated in (a). Subjects had to estimate the coordinates of the point indicated by the hatched circle in relation to the affine frame defined by the three other points. In (b), the diagram on the left shows the three possible orientations of the *OX* segment (1, 2, and 3) and the *OY* segment (a, b, and c). The diagram on the right shows the three possible lengths of *OX* and *OY* (indicated by numbers and letters, respectively). In combination, the orientation variables define the internal angle of the frame (1a = 90°, 1b = 120°, 1c = 150°, 2a = 60°, etc.), whereas the length variables define the aspect ratio of the frame (1a = 1.50, 1b = 1.00, 1c = 0.75, 2a = 2.25, etc.).

shorter or longer. The combination of these conditions resulted in 81 (3 × 3 × 3 × 3) different affine reference frames.

Task and Procedure

Subjects were instructed to inspect each pattern carefully and to locate the red dot by means of affine-invariant coordinates. This procedure was explained as follows. Subjects had to indicate to what percentage of the unit length on *OX* the red dot would project, with projection proceeding along a line through the point and parallel to *OY*. This was the *X*-coordinate. Similarly, a *Y*-coordinate had to be given. This procedure was illustrated carefully with two examples on the blackboard in a classroom; one example showed an orthogonal frame in the standard horizontal-vertical orientation, and the other showed an oblique frame with *OX* much longer than *OY*. The experimenter demonstrated the geometrical procedure by drawing some construction lines on the patterns, as in Figure 1b. The classroom also contained 15 computers, each with a 486 processor and an SVGA screen with a 800 × 600 resolution. Although the instructions were given collectively, each subject performed the experiment individually on a computer that was separated from the neighboring one by at least 75 cm. Each subject received the 81 configurations in a different random order and with a different random location of the fourth dot for each configuration. Subjects were instructed to solve the task purely visually.

When the experiment was initiated on the computer, the first pattern appeared in the middle of the screen while a text line underneath asked for a percentage on *X*. As soon as the subject entered a number in the computer, this text line was replaced by a second one asking for a percentage on *Y*, while the dot pattern remained on the screen. As soon as this second percentage was entered as well, the dot pattern disappeared and another one was shown. Nine practice trials were given to familiarize subjects with the task. These trials were followed by feedback on mean deviations from the correct *X*- and *Y*-coordinates and the standard deviation of the measurement errors. Subjects then had the opportunity to ask questions before the experimental trials began. These trials were administered in series of 10. After each series, subjects received feedback about their performance on those trials. The total experiment lasted for about half an hour.

RESULTS

Two dependent variables were measured, the percentage of error on the *X* estimate and the percentage of error on the *Y* estimate. For each dependent variable, the absolute deviations from the true values were entered into the data analysis. One obvious typing error was removed from the data files. The effects of the orientation and length of the *OX* and *OY* unit vectors were analyzed using two higher order variables that have a more direct meaning in terms of affine distortions. The two orientation variables were combined into one, *internal angle* ($OX \measuredangle OY$), and the two length variables were also combined into one, *aspect ratio* (*OX*/*OY*).

On average, subjects' estimates deviated from the true values by 5.8% for *X* and 7.5% for *Y*. Though this performance was not bad, it shows that the visual system does not measure affine-invariant coordinates perfectly. An important point is that the configurations used were only a small subset of the possible quadrilaterals generated randomly (e.g., those used by Wagemans et al., 1996). Larger errors would be quite likely with a less constrained set of shapes, especially because our task did not require the establishment of correspondences.

The internal angle had a large and systematic effect on the percentage of error for both *X* and *Y*, $F(8, 112) = 10.48$, $p < .0001$, and $F(8, 112) = 16.98$, $p < .0001$, respectively. As Figure 3 demonstrates, these effects imply that it becomes increasingly difficult to measure affine-invariant coordinates the more the internal angle deviates from 90°.

The aspect ratio also had a systematic, but somewhat smaller, effect on the accuracy of the *X* and *Y* estimates, $F(8, 112) = 2.14$, $p < .05$, and $F(8, 112) = 2.94$, $p < .01$, respectively. As Figure 4 demonstrates, the percentage of error decreases with increasing aspect ratio for *X* but increases with increasing aspect ratio for *Y*. That is, it becomes easier to estimate the *X* coordinate as *Y* becomes small compared with *X*,

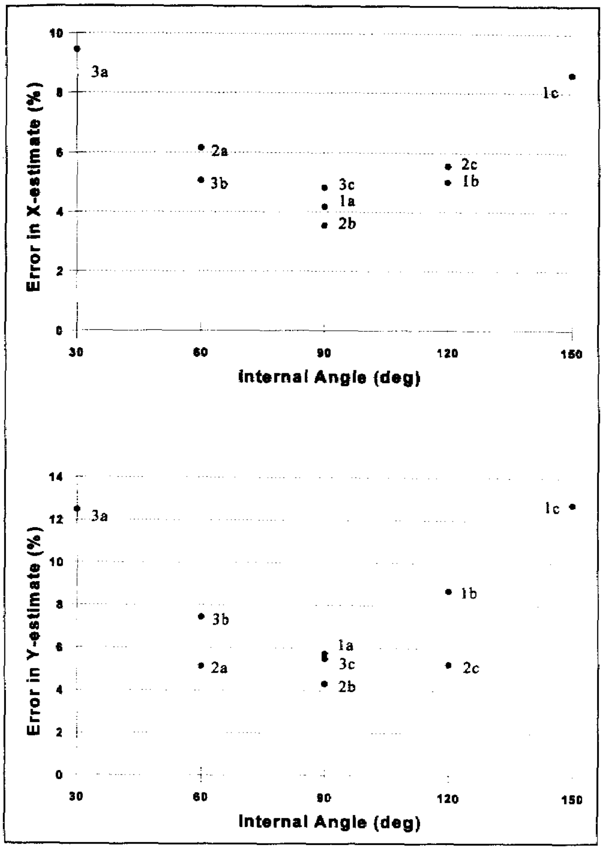

Fig. 3. Effects of the internal angle on the percentage of error in the *X*-estimate (top) and the *Y*-estimate (bottom). The number-and-letter codes in the graphs refer to the conditions as defined in Figure 2b (on the left).

whereas estimating *Y* coordinates becomes more difficult as *X* becomes large compared with *Y*.

DISCUSSION

When observers are asked to estimate the affine-invariant coordinates of a fourth point located within an affine reference frame defined by three other points, the accuracy depends on the particular configurations of the points. More specifically, the errors become larger as the internal angle of the frame deviates more from 90° and as the length of the unit vector under consideration becomes smaller relative to the other vector. Although the task in this experiment was quite different from a matching or a recognition paradigm, the implications for that line of research are clear. Even when subjects are explicitly asked to give affine-invariant coordinates, the size of the errors made depends on characteristics of the configurations related to affine deformation parameters such as shear and deformation. This finding supports the idea that the effects on response times and error rates of parametric differences between two shapes in a matching task, or between a presented picture of an object and a previously learned or standard model in a recognition experiment, do not necessarily reflect a mental transformation process that is supposed to undo the physical transformation. The results of this experiment demonstrate that these effects could also be caused by the measurement characteristics of the visual system. Regardless of how big or how robust these parametric effects are, they do not constitute solid evidence for the use of mental transformations or against the use of invariants.

If the measurement of affine-invariant coordinates is imperfect in ideal circumstances such as those in the task used in this experiment, how could such coordinates ever play a role in determining shape constancy or viewpoint-invariant recognition in more demanding circumstances requiring, for example, that features be found or that correspondences be established? In other words, the fact that the errors in the estimates of affine-invariant coordinates varied systematically with certain aspects of the pattern configuration may undermine the classic argument in favor of the mental transformation approach, but certainly does not appear to be good news for the invariants-based approach either.

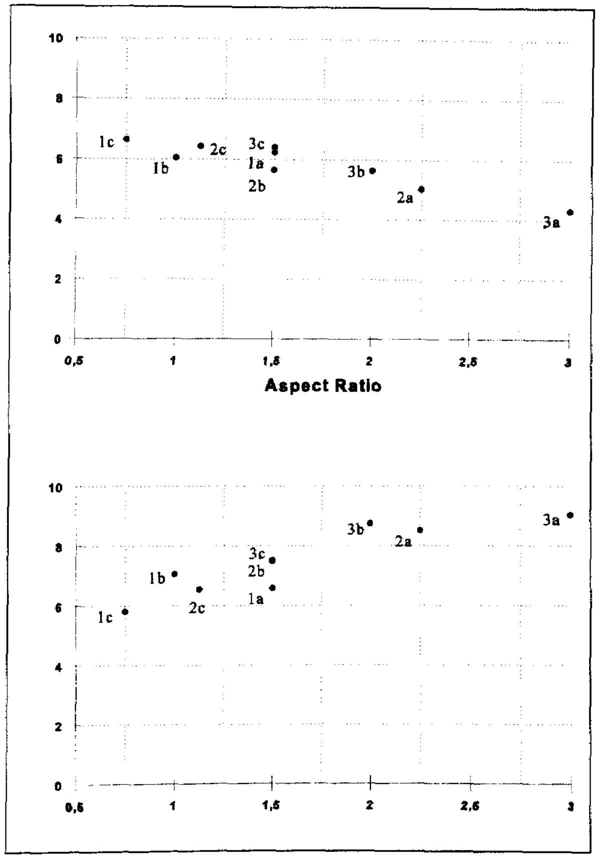

Fig. 4. Effects of the aspect ratio on the percentage of error in the *X*-estimate (top) and the *Y*-estimate (bottom). The number-and-letter codes in the graphs refer to the conditions as defined in Figure 2b (on the right).

However, in the experiment reported here, as well as in the affine matching experiments reported elsewhere (Wagemans et al., 1996), the information contained in the patterns is minimal in the sense that four points is the minimum for defining affine invariants at all, and, conversely, the affine structure of such patterns is fully determined by only a pair of affine-invariant coordinates, from which all other affine invariants for the patterns can be calculated (see Van Gool et al., 1994, for more mathematical background). In more realistic patterns, many more invariants exist, and it is quite likely that the visual system makes use of multiple sources of information when they are available (e.g., Bruno & Cutting, 1988; Cutting, 1986). Moreover, in many man-made objects, special features such as collinearity and parallelism are more likely to occur. The human visual system is much better equipped to detect such regularities or nonaccidental properties (Lowe, 1987; Wagemans, 1993). However, they tie in naturally with the other invariants because they often boil down to special cases of them (see Van Gool et al., 1994). In a more recent study, we have demonstrated more directly that special features like concavity, convexity, collinearity, and parallelism are used as qualitative cues in the discrimination of affine-transformed minimal patterns (Kukkonen, Foster, Wood, Wagemans, & Van Gool, 1996).

Acknowledgments—This research was supported by an Esprit Basic Research Action (No. 6448, VIVA for Viewpoint-Invariant Visual Acquisition). J.W. was supported by the National Fund for Scientific Research (NFWO-Belgium) and the Research Council of the University of Leuven (PDM-94/32). We would like to acknowledge the programming support by Andreas De Troy.

REFERENCES

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.

Bruno, N., & Cutting, J.E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General, 117*, 161–170.

Cooper, L.A. (1976). Demonstration of a mental analog of an external rotation. *Perception & Psychophysics, 19*, 296–302.

Cutting, J.E. (1986). *Perception with an eye for motion*. Cambridge, MA: MIT Press/Bradford Books.

Farah, M.J., Rochlin, R., & Klein, K.L. (1994). Orientation invariance and geometric primitives in shape recognition. *Cognitive Science, 18*, 325–344.

Gibson, J.J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

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

Jolicoeur, P., & Landau, M.J. (1984). Effect of orientation on the identification of simple visual patterns. *Canadian Journal of Psychology, 38*, 80–93.

Koenderink, J.J., & van Doorn, A.J. (1991). Affine structure from motion. *Journal of the Optical Society of America A, 8*, 377–385.

Kukkonen, H.T., Foster, D.H., Wood, J.R., Wagemans, J., & Van Gool, L. (1996). Qualitative cues in the discrimination of affine-transformed minimal patterns. *Perception, 25*, 195–206.

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

Mundy, J.L., & Zisserman, A. (Eds.). (1992). *Geometric invariance in computer vision*. Cambridge, MA: MIT Press.

Pizlo, Z. (1994). A theory of shape constancy based on perspective invariants. *Vision Research, 34*, 1637–1658.

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

Tarr, M.J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2*, 55–82.

Tarr, M.J., & Bülthoff, H.H. (1995). Is human object recognition better described by geon-structural descriptions or by multiple views? *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1494–1505.

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

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

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

Van Gool, L., Moons, T., Pauwels, E., & Wagemans, J. (1994). Invariance from the Euclidean geometer's perspective. *Perception, 23*, 547–561.

Wagemans, J. (1993). Skewed symmetry: A nonaccidental property used to perceive visual forms. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 364–380.

Wagemans, J., Van Gool, L., Lamote, C., & Foster, D.H. (1996). *Minimal information to determine affine shape equivalence*. Manuscript submitted for publication.

(Received 3/20/95; Accepted 6/29/95)

