Minimal Information to Determine Affine Shape Equivalence

Johan Wagemans, Luc Van Gool,
and Christian Lamote
University of Leuven

David H. Foster
University of Manchester Institute
of Science and Technology

Participants judged the affine equivalence of 2 simultaneously presented 4-point patterns. Performance level (d') varied between 1.5 and 2.7, depending on the information available for solving the correspondence problem (insufficient in Experiment 1a, superfluous in Experiment 1b, and minimal in Experiments 1c, 2a, 2b) and on the exposure time (unlimited in Experiments 1 and 2a and 500 ms in Experiment 2b), but it did not vary much with the complexity of the affine transformation (rotation and slant in Experiment 1 and same plus tilt in Experiment 2). Performance in Experiment 3 was lower with 3-point patterns than with 4-point patterns, whereas blocking the trials according to the affine transformation parameters had little effect. Determining affine shape equivalence with minimal-information displays is based on a fast assessment of qualitatively or quasi-invariant properties such as convexity/ concavity, parallelism, and collinearity.

When an object is viewed from different positions, its projected shape on the retina varies, yet its perceived shape often remains the same. A door, for example, looks rectangular even though its projected shape is generally trapezoidal. This phenomenon, called *shape constancy*, has long puzzled perceptual scientists (e.g., Helmholtz, 1857/1962; Koffka, 1935; Rock, 1983). According to Gibson (e.g., 1950, 1979), the visual system interprets different views of the same object as having the same shape because it is sensitive to their projective equivalence, that is, their congruence or equivalence under the group of projective transformations.

Shape Equivalence

Projective Equivalence

Surprisingly little empirical work has been devoted to a direct test of the visual system's sensitivity to projective equivalence as a possible basis of shape constancy. In fact, even the experimental establishment of the phenomenon of shape constancy itself came late (e.g., Lappin & Preble, 1975; but see also Stavrianos, 1945, and Thouless, 1931, 1934, for interesting historical precursors). In three studies of the perceptual relevance of projective invariance, Niall critically reviewed the available evidence (Niall, 1992; Niall & Macnamara, 1989, 1990). He concluded that all of the earlier principal experiments either involved projections of highly regular and familiar objects (e.g., Attneave & Frost, 1969; Perkins, 1972) or merely demonstrated shape constancy and assumed projective invariance as the explanation without actually testing it (e.g., Johansson, 1975; Johansson, von Hofsten, & Jansson, 1980). Subsequently, Cutting (1986) examined the perceptual usefulness of the cross-ratio, the prototypical projective invariant, in judgments of the rigidity and flatness of a translating or rotating planar surface. With displays of four parallel lines, it was shown that the perceived departure from rigid motion was proportional to the degree of projective distortion (i.e., deviation from the cross-ratio). Nevertheless, it was concluded that the cross-ratio could not always be effective in specifying form (see Cutting, 1987, and Niall, 1987, for more technical discussion).

In one of Niall's studies on sensitivity to projective equivalence (Niall & Macnamara, 1989), observers were asked to complete drawings of a quadrilateral structure (a two-dimensional [2-D] glass panel of a building) viewed from an oblique angle. Naive observers as well as students

of architecture produced drawings that deviated systemati-
cally from projective equivalence. In a second study (Niall
& Macnamara, 1990), the projective thesis was tested more
directly: Participants were presented with projectively con-
gruent or noncongruent shapes (pentagons with a star to
indicate their three-dimensional [3-D] orientation) and were
asked to select the projectively equivalent pairs. The five-
point cross-ratios of the selected comparison figures differed
significantly from those of the standard figures in an active
drawing task as well as in a passive recognition task. In a
third study (Niall, 1992), the poor sensitivity to projective
equivalence of static shapes was shown to extend to planar
shapes (quadrilaterals) and solid forms (prisms) that were
continuously rotating in depth around a fixed rotation point.
Collectively, these data reflect a markedly inaccurate estima-
tion of projective properties that led Niall to conclude that
sensitivity to projective equivalence seems an implausible
basis for shape constancy.

Affine Equivalence

In the work presented here, we take one step back and
examine whether human observers can judge affine equiva-
lence between simple planar shapes. More specifically, we
ask whether the minimal information from which it is
possible to determine affine shape equivalence is sufficient
for such judgments. Because affine transformations are more
restrictive than projective transformations (e.g., parallelism
is preserved), affine shape equivalence might well be easier
to determine than projective shape equivalence. In practice,
orthographic or parallel projection (which involves affine
transformations) is a good approximation to perspective or
polar projection (which involves projective transformations)
when the depth range of the objects is small relative to the
viewing distance. Psychophysical research that has com-
pared the two has shown that perceptual performance is
often equivalent (e.g., Börjesson & Lind, 1996; Braunstein,
Liter, & Tittle, 1994; Lappin & Fuqua, 1983; Todd, 1984;
Wagemans & Tibau, 1999).

Perspective Equivalence

A related strategy was followed by Pizlo (1994) in his
analysis of shape constancy. Instead of moving from affine
to projective—which is the next broadest group in Klein's
(1908/1939) hierarchy of transformation groups (see Cut-
ting, 1983, 1986, Chapter 5, for a nontechnical introduc-
tion)—Pizlo focused on perspective transformations (or
perspectivities) as an intermediate case. Although they do
not form a group and therefore do not have invariants in the
classical sense, perspective transformations are interesting
because they seem most relevant to vision. Whereas projec-
tive transformations are arbitrary combinations of perspec-
tivities, the geometry that is most relevant to vision involves
only one perspective and a constant distance between the
center of projection and the image plane (Pizlo, Rosenfeld,
& Weiss, 1997). In addition, Pizlo and Rosenfeld (1992)
have shown that perspective transformations have so-called
pseudo- or *quasi-invariants*. This notion was introduced by

Binford and Levitt (1993) to describe properties that may
vary with the transformation but do so within a small range
of values over a large range of transformations. In a series of
studies (e.g., Pizlo, 1994; Pizlo & Salach-Golyska, 1995),
Pizlo showed that human observers are sensitive to perspec-
tive equivalence and that some specific predictions derived
from an algorithmic approach to perspective judgments
(Pizlo & Rosenfeld, 1992) could be corroborated.

In sum, we wanted to investigate how well human
observers can determine affine shape equivalence in patterns
with minimal information because it may shed some new
light on the use of invariants under groups of transforma-
tions as a basis for shape constancy.

Human Performance Versus Mathematical Constraints

Another reason to focus on the issue of minimal informa-
tion to determine affine shape equivalence is that similar
questions of minimal information have been productive in
other areas of research. For example, the computational
theorem that three distinct views of four noncoplanar points
are both necessary and sufficient for obtaining a unique
interpretation of an object's 3-D structure (formulated by
Ullman, 1979) has been successfully exploited in several
psychophysical studies (e.g., Braunstein, Hoffman, & Pol-
lick, 1990; Braunstein, Hoffman, Shapiro, Andersen, &
Bennett, 1987; Domini, Caudek, & Proffitt, 1997; Doner,
Lappin, & Perfetto, 1984; Lappin, Doner, & Kottas, 1980;
Todd, Akerstrom, Reichel, & Hayes, 1988; Todd & Bressan,
1990). The general conclusion from this research is that the
ability of human observers to perceive structure from motion
is sometimes more powerful than expected on the basis of
some computational models (see also Todd, 1985). Thus,
two views of four noncoplanar points are sufficient for some
perceptual tasks like rigidity judgments (subsequently also
shown to be possible mathematically by Bennett, Hoffman,
Nicola, & Prakash, 1989, and, independently, by Huang &
Lee, 1989). Experiments with other phenomena such as the
stereokinetic effect, motion parallax, and depth from dispar-
ity also suggest that the human visual system is able to
derive estimates of depth magnitudes in situations where it
seems impossible on the basis of the laws of projective
geometry alone (e.g., Caudek & Proffitt, 1993; Durgin,
Proffitt, Olson, & Reinke, 1995; Proffitt, Rock, Hecht, &
Schubert, 1992).

Conclusion

In the light of this accumulating evidence that human
observers can often do more than expected on the basis of a
computational analysis of the mathematical constraints,
Niall's results showing insensitivity to projective equiva-
lence in shape constancy are the more unexpected. Yet, as
Niall and Macnamara (1990, p. 658) admit, they "have not
proved once and for all that the projective thesis never
provides an explanation of shape constancy, since, naturally,
[they] have not examined all possible conditions in which
the thesis might obtain." Because the available evidence is

mixed (negative in Niall's research, positive in Pizlo's research, and mixed in Cutting's) and because research from other areas suggests that our visual system is more rather than less resourceful given the geometric laws, an alternative approach to the question of shape equivalence is needed. Because the mathematics of the problem of affine shape equivalence is central to this work, some technical background is first provided.

Affine Transformations

Hierarchy of Geometries

As noted earlier, different geometries can be put in a hierarchy, from Euclidean geometry to topology (Klein, 1908/1939). When one moves up in the hierarchy, more aspects of a shape are lost and, consequently, the aspects remaining unchanged or invariant become more abstract. The geometrical classification is not discussed here (a good nontechnical introduction with pictorial examples can be found in Michaels & Carello, 1981, Chapter 2, pp. 30–37, and a more mathematical treatment in Foster, 1975, and Van Gool, Moons, Pauwels, & Wagemans, 1994). It is, however, useful to show the effect of a few common transformations on shape equivalence to indicate the position of affine transformations in this hierarchy. Figure 1 shows a pentagon (A) and the resulting shape after rotation (B), size scaling (C), shear and compression (D), a single perspective projection (E), and an arbitrary combination of perspective projections (F). These figures exemplify the following transformations, respectively: Euclidean, similarity, affine, perspective, and projective. Some of the properties lost in going from (A) to (F) are the distances in (A), which are preserved only in (B) (i.e., distance is a Euclidean but not an affine invariant),
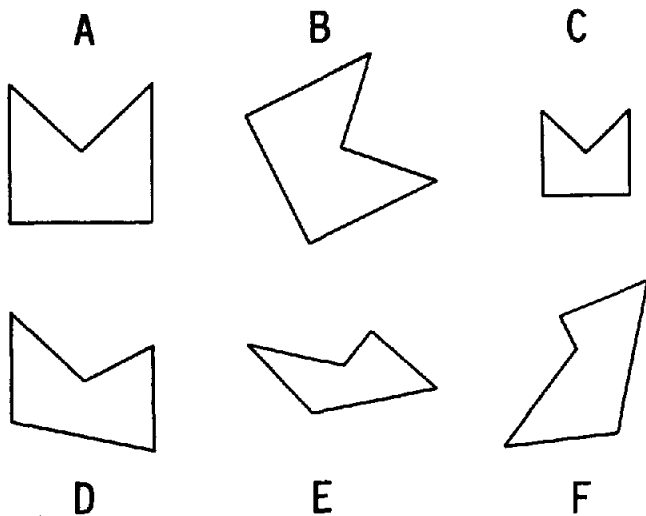

Figure 2. Minimal information needed to determine shape equivalence under different geometries (represented in three rows, from top to bottom): similarity, affine, and projective. For patterns in the left column, there is insufficient information to determine shape equivalence, and within this column, the two patterns on the right are equivalent to the one on the left (under the transformation group considered in each row). For patterns in the right column, there is sufficient information to determine shape equivalence, and within this column, the pattern on the left is the standard, the one in the center is a transformed version of it, and the one on the right cannot be obtained by such a transformation (under the transformation group considered in each row).

and the parallelism of the pentagon in (A)–(D), which is destroyed by the perspective as well as projective transformations in (E) and (F) (i.e., parallelism is an affine but not a projective invariant).

Minimal Information

As a corollary to this hierarchical organization of transformations, the minimal information from which it is possible to determine shape equivalence varies systematically. Thus, because distance is the basic invariant of Euclidean transformations, pairs of points are sufficient to specify *shape* in a Euclidean way: Two point pairs are Euclidean equivalent if and only if they are separated by the same distance. Three points are needed in similarity geometry (see Figure 2, top row): Any pair of points is indistinguishable because there is always a combination of size scaling with rotation and translation that will make the pairs congruent; if there are three points, two of them will define a reference frame, and the third can be tested against the similarity transformation computed from the reference pair (assuming that no reflections are allowed). Four points are needed in affine geometry (see Figure 2, middle row): All triangles are affinely equivalent, and three points define an affine reference frame (Koenderink & van Doorn, 1991; Lamdan, Schwartz, & Wolfson, 1988). This procedure forms the mathematical basis of Ullman's (1989) alignment scheme and so-called *recognition polynomials*, a computational technique to recognize shapes from different viewpoints (Bennett, Hoffman, & Prakash, 1993). Finally, five points are the minimum needed to determine projective shape equivalence (see Figure 2, bottom row).

Figure 1. A pentagon (A) and the resulting shape after rotation (B), size scaling (C), shear and compression (D), a single perspective projection (E), and an arbitrary combination of perspective projections (F), representing different geometries (Euclidean, similarity, affine, perspective, and projective, respectively). Properties are progressively lost in going from (A) to (F).

In sum, we use the notion of *minimal information* not in the sense of least complexity but to indicate that four points are necessary to determine shape equivalence under arbitrary affine transformations. In fact, it is quite likely that determining shape equivalence is more difficult between our minimal shapes than with natural shapes in more realistic environments. Even so, if we can establish that affine shape equivalence can be determined from four points, we will at least have shown that the visual system can rely on what is mathematically available when all other sources of information (e.g., about depth or about object identity) are removed. To show this was the major goal of the present study.

Decomposition of Affine Transformations

All that now remains to be done before describing the experiments is to specify the particular parameterization or decomposition of the affine transformation group used in this study. This is not trivial (see Wagemans, De Troy, Van Gool, Wood, & Foster, 1994, and Wagemans, Vanden Bossche, Segers, & d'Ydewalle, 1994, for a fuller discussion). Mathematically speaking, an affine transformation can be implemented as a multiplication of the (x, y) coordinates of the pattern by a 2×2 matrix, $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$. The four parameters of the matrix (i.e., a, b, c, and d) can be chosen arbitrarily, as long as they are real numbers and the resulting determinant differs from zero.

There are several ways in which this general affine transformation can be decomposed into sequences of simpler transformations. That these different decompositions are not equivalent follows from the fact that the affine group is not a direct product of one-dimensional subgroups. There is no way to parameterize these transformations as a sequence of one-parameter transformations without the order in which these are carried out being important for the precise amounts (e.g., rotation angles, scale factors) needed of each. Changing the order will automatically imply changing the relative amount of each one-parameter transformation. The exceptions to this rule are rare. Rotation and scaling can be combined in any order without changing the effect. Not accidentally, these are the transformations that have been studied empirically to test the plausibility of the mental transformation approach (e.g., Bundesen & Larsen, 1975; Larsen & Bundesen, 1978; Shepard & Cooper, 1982). It is hard to understand how the perceptual system would deal with more complicated groups, such as the affine transformations tested here, where the decomposition strategy matters and where a prohibitively large number of essentially different compositions exist for each transformation. Much of the problem finds its origin in the noncommutativity of matrix multiplication in the general case. Consider the following example. Take a book and hold it with its front cover toward you. Then rotate it clockwise about the optical axis through 90° and rotate it over 45° about the horizontal axis. The resulting image is different from what would be obtained if the order of these rotations were reversed.

In the present experiments, a decomposition for general affine transformations was used as a point of departure, but only a subset of the resulting parameters were varied

experimentally to avoid a combinatorial explosion of the number of trials (see Appendix A for the details of the derivations). An effort was made to vary those parameters that maximally capture the typical affine (i.e., non-Euclidean) deformations. An affine transformation can be interpreted as describing the effect of orthographically projecting a planar shape oriented arbitrarily in depth onto an image plane or screen. The 3-D orientation of the plane containing the patterns can be specified by two parameters, slant and tilt (see Figure 3); within this plane, patterns were allowed to have an arbitrary orientation. Hence, the affine transformation under which shape equivalence had to be judged by observers consists of a rotation in the image plane and a slant and tilt of the plane (Figure 4), applied in that order (see Appendix A for the computational formulas).

Notice that tilt as defined here might be found to differ slightly from definitions elsewhere (e.g., Stevens, 1983). Tilt is sometimes defined through the orientation of the projection of the normal of the object plane onto the image plane. Our definition uses the orientation of the line of intersection of these planes. The resulting tilts may differ by 90°. Notice also that all of the affine transformations tested here have a positive determinant; hence, no reflections were allowed. In Experiment 1, the situation was simplified by assuming zero tilt angles. If these transformations defeated the ability of observers to determine shape equivalence, there would be no point in taking the experiments further to include nonzero tilts.

Experiment 1

The main purpose of Experiment 1 was to determine whether observers are able to decide on affine shape
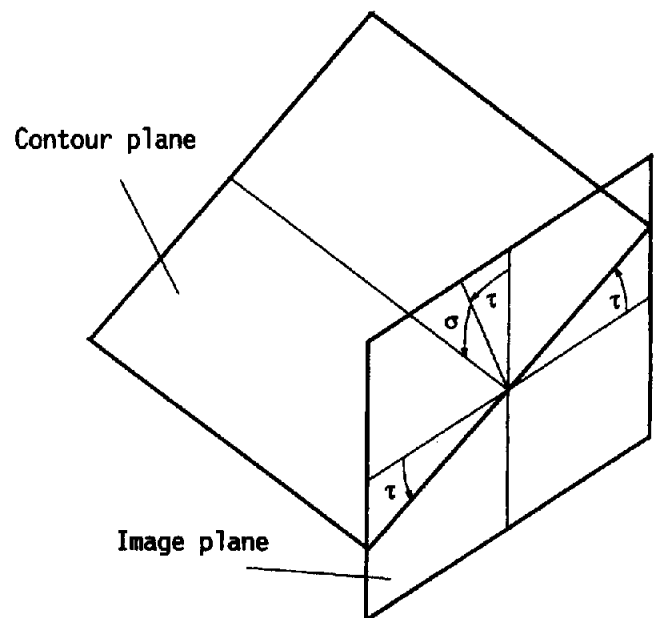

Figure 3. The 3-D orientation of the plane containing a pattern can be specified by two parameters, slant angle σ and tilt angle τ.
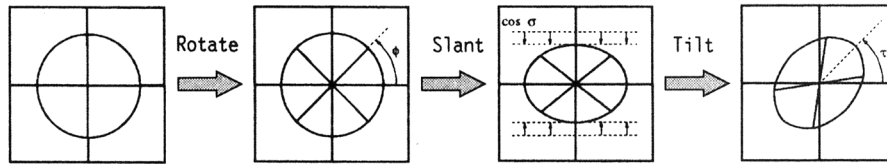
Figure 4. Decomposition of an affine transformation in three steps (from left to right): (1) rotation in the plane; (2) slant about the horizontal coordinate axis, equivalent to compression along the vertical axis; and (3) tilt.

equivalence given minimal information. As explained earlier, four points are sufficient to determine affine shape equivalence. Participants were therefore presented with patterns consisting of four points in a random configuration. In *same* trials (i.e., with affine equivalence), the second pattern was derived from the first by multiplying its (x, y) coordinates by a matrix whose entries represented specific parameter values for the rotation ϕ and slant σ. In other words, the imaginary planar surface on which the points were plotted was rotated about the optical axis and subsequently slanted about the horizontal axis. Hence, the patterns could be considered the same but viewed from a different angle. In *different* trials (i.e., without affine equivalence), the patterns that had to be compared were unrelated; a similar affine transformation was applied to a second set of four randomly selected (x, y) coordinates. This transformation was applied to avoid possible cues based on the relative distances between points because they are smaller in the transformed patterns than in the untransformed ones. Participants had to respond *same* to the first type of pattern pairs and *different* to the second. Performance level was measured

across a large number of trials to determine whether minimal information was sufficient; response times (RTs) were recorded to enable an assessment of the effect of the size of the affine transformation.

Three variants of the same experiment were run. In Experiment 1a, all four points in a point pattern were black. This implies that observers had to solve the correspondence problem in order to match the point patterns: They had to discover which point in the second pattern corresponded to a particular point in the first pattern. Because all the points were black, participants could detect nonintended matches between points in one pattern with points in the other pattern by matching basis points that were labeled differently by the computer. As a result, some *different* trials could have been considered affine equivalent simply because of nonintended correspondences (see Figure 5 for an example). Pizlo (1994) has called this the *shape ambiguity* problem, which is really the complement of the shape constancy problem. Whereas constancy has to do with two different views of one single object being interpreted as images of the same object, ambiguity refers to the fact that two different objects

A   B   C

Figure 5. An example of a *different* trial which could give rise to a false positive (*same*) response because the two independently generated point patterns could be made (close to) affine equivalent by assuming incorrect correspondences. (A) An affine reference frame attached to the first randomly generated point pattern in which one of the points is taken as the origin and two others as defining vectors OX and OY. The coordinates of the remaining point, P, can then be expressed as a linear combination of the OX and OY vectors (i.e., 1.54 times OX and 0.35 times OY). (B) and (C) Two copies of the affine transformed version of the second randomly generated pattern. In (B) the incorrect correspondence establishes an affine reference frame within which the fourth point, P', happens to lie only a few pixels away from its correct position. In (C) the intended correspondence is shown together with the location of the fourth point when the second pattern is affine related to the first (i.e., the open circle). Under these correspondences, P'' deviates from its affine-equivalent position.

sometimes give rise to the same image. In Experiment 1a, the latter was possible as observers could take any of the several different correspondences as a possible match.

In Experiment 1b, shape ambiguity was avoided by adding color. In this experiment, participants were given all the information they needed to solve the correspondence problem perfectly. Each point was identified with a unique color that was preserved in the affine-transformed pairs. The probability that nonintended correspondences would still yield close-to-affine matches in *different* trials is much smaller under these conditions. Therefore, rejecting random pairs should become as easy as identifying affine-equivalent ones. The literature on visual search, however, suggests that identity and position cannot be encoded simultaneously without conscious effort (e.g., Atkinson & Braddick, 1989; Green, 1992; Johnston & Pashler, 1990; Sagi & Julesz, 1985; Tsal & Lavie, 1993). Moreover, research on apparent motion has shown that color is not as important as shape. For example, Kolers and von Grünau (1976) reported that observers resolved differences in shape between the initial and final stimuli by perceiving a gradual shape transformation, whereas color differences were resolved abruptly, somewhere in mid-air. Therefore, solving the correspondence problem by considering each point's unique color and determining whether an affine transformation can bring the patterns into congruence could require considerable scrutiny and time. If observers were not willing to make this effort on each trial, then performance might not increase as much as it should.

Finally, Experiment 1c provided the most stringent test of the sufficiency of minimal information to determine affine equivalence between two planar patterns of four points. The addition of a unique color to each point in Experiment 1b actually provided too much information: If only one element out of the four could be matched efficiently, invariant ordering cues would in most cases give the other correspondences away (because reflections were excluded). As indicated in Figure 5, three points constitute an affine reference frame within which the position of the fourth may be computed. By adding color to only one point, this point will be easily identified as the anchor point or origin of the affine reference frame (*O* in Figure 5). This is what was done in Experiment 1c: One of the points in each pattern was colored blue and the others black. The use of only two colors should avoid the high attentional demands associated with the four colors of Experiment 1b (e.g., Johnston & Pashler, 1990; Tsal & Lavie, 1993). The only cases where this colored point would be insufficient as an anchor point are those where it happens to be the central point in a concave configuration. Because on average 6% of the trials contained two concave configurations, this occurred in only 1.25% of the trials.

Method

Participants. Thirty undergraduate students from the Department of Psychology at the University of Leuven volunteered to participate in partial fulfillment of a course requirement. They were naive about the purpose of the experiment and had normal or corrected-to-normal vision. Ten participants were randomly assigned to each of the three variants of Experiment 1.

Stimuli. In each trial, two patterns, each constructed on the basis of four (x, y) coordinates, pseudorandomly selected in a circular area with a diameter of 264 pixels (10.8 cm), were presented on the screen, one in the left half, the other in the right half, separated by a center-to-center distance of 364 pixels (14.8 cm). One of the two patterns was always untransformed (i.e., the original pattern); the other one was a transformed version of either the same original pattern or a different one. The left–right position of the untransformed pattern was chosen randomly from trial to trial. There were two constraints on the random selection of the (x, y) coordinates to obtain more-or-less homogeneous distributions: First, one (x, y) coordinate had to be selected in each quadrant of an imaginary circle. The orientation of the orthogonal axes defining the quadrants was always selected at random. Second, strong proximities were avoided by rejecting possible (x, y) coordinates closer than 7 pixels (2.8 mm). The pseudorandom selection of four (x, y) coordinates was repeated for each original pattern in all of the trials of all of the experiments. The point patterns were created by plotting small, filled circles (radius 3 pixels) on the selected (x, y) coordinates. The points were easy to distinguish from the light gray background (with 50 cd/m²) and, when colored differently, also from one another. Black (with 15 cd/m²), blue (with 27 cd/m²), red (with 32 cd/m²), and yellow (with 106 cd/m²) were selected from the standard VGA color palette.

The affine transformations tested in this experiment were a combination of rotation and slant, applied in that order. In principle, all cases are covered by letting the slant angle σ vary between 0° and 90° and the rotation angle ϕ vary between 0° and 360°. Because the compression in the image as a result of slanting a planar surface in depth is given by the cosine of the slant angle, the range was restricted to avoid degenerate cases. The *slant angle* σ had one of five values: 0°, 15°, 30°, 45°, and 60°. Because the number of levels of the rotation factor had to be limited, the *rotation angle* ϕ was given one of nine values: 0°, 22.5°, 45°, 67.5°, 90°, 112.5°, 135°, 157.5°, and 180°. Each of these rotation values occurred equally often and was positive (clockwise) in about half of the trials (chosen randomly) and negative (counterclockwise) in the rest. Figure 6 shows what happens to a pattern of four points under some of these affine transformations. Figure 7 shows how color may help to determine shape equivalence.

Apparatus. The experiment was fully automated and run on a computer with a VGA graphics board. The point patterns were
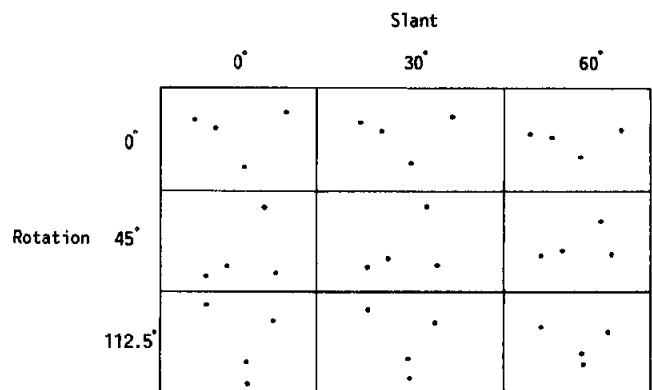

Figure 6. An example of a pattern with four pseudorandomly generated points and the resulting image after application of the affine transformations used in Experiments 1–3. The columns represent different slant angles ($\sigma = 0°$, 30°, and 60°), and the rows different rotation angles ($\Phi = 0°$, 45°, and 112.5°).
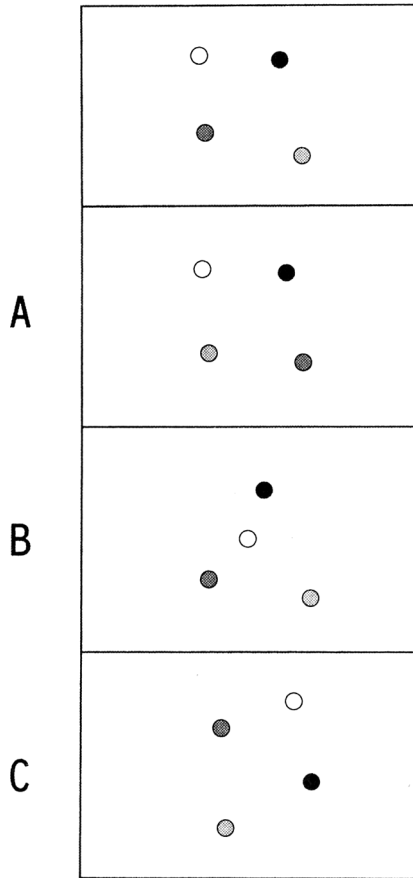
Figure 7. An example of a randomly generated point pattern (i.e., the standard shown at the top) and the way color can be used to determine shape equivalence under an affine transformation. (A) This pattern is not congruent to the standard pattern because the order of the colors differs. (B) This pattern is affine equivalent on the basis of the color correspondences but not on the basis of point positions. (C) This pattern is affine equivalent on the basis of both the color correspondences and the point positions.

presented as black, blue, red, or yellow points on a gray background on a color screen with a 70 Hz temporal resolution and a 640 × 480 spatial resolution. Participants were seated at a distance of 114 cm on a chair with adjustable height so that their eyes were level with the center of the screen. At that viewing distance, the individual points, the patterns defined by them, and the whole stimulus area (patterns and separating space) subtended 6.13 arcmin, 5.39°, and 12.76° of visual angle, respectively. The room was moderately lit. A headrest was used to stabilize the head. Viewing was binocular.

Task and procedure. Participants were told that they would be shown two patterns side-by-side on the screen. In some trials, the patterns would have the same shape, although possibly viewed from a different position, whereas in other trials the patterns would be unrelated. Participants were instructed to consider the patterns in each display as the *same* in the former case and *different* in the latter. Thus, affine shape equivalence was explained by referring to the notion of different views on a plane with arbitrary 3-D orientation, which was also explicitly demonstrated with a sheet of paper held at different 2-D and 3-D orientations. Participants

responded by pressing one of two buttons on a response panel; half of the participants used the right button to respond *same* (i.e., affine equivalent) and the left for *different* (i.e., not affine equivalent) and vice versa for the other half. Participants were familiarized with the task in a block of 150 practice trials, which were made up in the same way as the experimental ones but with different patterns.

Because pilot work had indicated that most participants used two fixations in viewing the stimuli (i.e., one for each of the two patterns presented simultaneously), a fixation cross was not used. Each trial consisted simply of the presentation of the two patterns, one on each side of the screen. Viewing time was unlimited: The patterns remained on the screen until the participant pressed one of two response buttons on a panel connected to the computer. Response times (RTs) were recorded to the nearest ms from stimulus onset until button press.

In half of the trials, an original pattern was paired with a transformed version of the same pattern; in the other half, an original pattern was paired with a transformed version of a different pattern (i.e., with freshly generated random coordinates). No single pattern was used in more than one trial (unless by accident the same set of four coordinates was chosen twice, which was extremely unlikely). Each of the 45 parameter combinations (9 rotation angles × 5 slant angles) of the affine transformation occurred 10 times: five times for *same* trials and 5 times for *different* trials. The total sequence of trials was divided into three blocks of 150 trials that could be run without a break. Each block, taking about 10 to 15 min, was a random sample from the total sequence of trials with the constraint that it contained an equal number of *same* and *different* trials. All participants received three blocks of trials in one session with only minor (e.g., 2–3 min) breaks between blocks. No trial-by-trial feedback was given, but after each series of 150 trials, the general level of performance was reported to the participant as the percentage of correct responses.

Results

General performance. The average percentage of correct responses computed across participants and across both types of trials was 79.5% in Experiment 1a, 85.4% in Experiment 1b, and 81.2% in Experiment 1c (see Table 1).

Table 1
Performance Levels in Experiment 1

	Experiment 1a		Experiment 1b		Experiment 1c	
Index	M	SD	M	SD	M	SD
% CRT	79.45	4.89	85.44	6.94	81.16	4.75
% CI	85.54	10.31	96.04	3.62	89.49	8.15
% CR	73.44	6.73	74.84	14.47	72.82	15.61
% same	56.03	7.21	60.59	7.94	58.34	11.53
d'	1.79	0.38	2.70	0.68	2.08	0.27
RT$_{CI}$	2442	1010	3517	1074	3032	957
RT$_{CR}$	2791	1295	5052	2454	3710	1172
% CR*	90.57	6.58	84.58	15.64	88.29	7.98

Note. Each row represents data for a different performance index (from top to bottom): % CRT = total percentage correct responses; % CI = percentage correct identifications; % CR = percentage correct rejections; % same = percentage *same* responses; d' = discrimination index; RT$_{CI}$ = average response time (in ms) for the correct identifications; RT$_{CR}$ = average response time (in ms) for the correct rejections; and % CR* = percentage correct rejections of trials with convexity cue. Each entry represents the mean and standard deviation over 10 participants.

In all three variants of this experiment, the mean percentage of correct identifications (*same* responses on *same* trials) was higher than the mean percentage of correct rejections (*different* responses on *different* trials). A *t* test computed on dependent samples reached statistical significance for Experiment 1a and 1b, $t(9) = 2.66$ and 4.22, ($p < .05$ and $.01$, respectively), whereas it was only marginally significant in Experiment 1c, $t(9) = 2.29$ ($p < .10$). This difference was partly based on a higher rate of *same* responses (e.g., 56.0% in Experiment 1a). Somewhat surprisingly, the addition of color did not remove this response bias. Although it was anticipated that the problem with shape ambiguity in point patterns with only black points would be reduced, it was actually a little higher in Experiment 1b (60.6%) and Experiment 1c (58.3%). Interestingly, this *same* bias was more pronounced in those participants with poorer performance levels, a result that is confirmed by the large negative correlation between average performance level and response bias for each participant, $r(9) = -.89$ and $-.81$, in Experiments 1b and 1c, respectively ($p < .01$ in both cases). These participants may not have used color as often as the others to reject *different* pairs. When response biases were eliminated, performance was still reasonably high (average $d' = 1.8$, 2.7, and 2.1, in Experiments 1a, 1b, and 1c, respectively).

The average RT followed the same pattern: RTs for the correct identifications were smaller than for the correct rejections, especially for Experiment 1b. As anticipated, the process of using color in Experiment 1b was slow and effortful.

Analysis of variance (ANOVA) on RTs and error rates. To enable an assessment of the effects of trial type, rotation, and slant, an ANOVA was performed on the RTs of the correct-response trials (averaged across at most five repeated measures). In the few cases where none of the responses to repeated measures for a specific trial were correct, the general average for that trial type (e.g., *different*) for that participant was used. This is a conservative procedure that

decreases the chance of obtaining statistically significant effects. Because a complete within-subjects design was used, each effect in this and the next experiments was analyzed with the subject-interaction as error term. Table 2 summarizes the effects. Trial type was always at least marginally significant. As with the general performance levels, this effect was attributable to *same* trials being responded to faster and more accurately than *different* trials.

Parametric effects were rather unstable across subexperiments and across dependent measures, but generally the following trends emerged. When a main effect of rotation or slant reached statistical significance, it was usually also involved in a reliable interaction with trial type. These interactions occurred because the affine transformation parameters had only an effect on *same* trials, not on *different* trials. This result is illustrated in Figure 8 for rotation and in Figure 9 for slant (with averages across all trials from all three subexperiments). These figures also show that rotation had a more pronounced effect on RTs (Panel A), whereas error rates (Panel B) were more influenced by differences in slant. Finally, it should be observed that slant had a highly nonlinear effect: RTs increased only when slant was 45° or 60°, whereas error rates were virtually identical (between 5% and 10%) for all slants except for 60° (20%). This suggests that performance is affected more by the compression effect in the plane than by the slant angle σ in depth (i.e., it varies more linearly with cos σ).

Discussion

More than 85% of the affine-equivalent point patterns were recognized as such when all points were black and more than 95% when each point had its own color. Although far from perfect, given that participants required almost 2.5 s to do the task when all points were black and more than 3.5 s when all points were colored, the level of performance does show that observers are capable of determining affine shape equivalence at much higher than chance level on the basis of

Table 2
Results of Analyses of Variance on Response Times and Error Rates in Experiment 1

Effect	df	Experiment 1a	Experiment 1b	Experiment 1c
Response times				
Tr	1, 9	3.93 (<.08)	9.34 (<.05)	4.59 (<.07)
R	8, 72	<1	8.15 (<.0001)	2.70 (<.05)
S	4, 36	5.13 (<.005)	<1	<1
Tr × R	8, 72	3.27 (<.005)	8.32 (<.0001)	4.22 (<.0005)
Tr × S	4, 36	5.96 (<.001)	4.43 (<.005)	1.38 (>.25)
R × S	32, 288	<1	<1	<1
Tr × R × S	32, 288	<1	<1	1.08 (>.35)
Error rates				
Tr	1, 9	6.91 (<.05)	17.82 (<.005)	5.20 (<.05)
R	8, 72	1.88 (<.08)	1.61 (>.13)	3.09 (<.005)
S	4, 36	3.98 (<.01)	<1	3.93 (<.01)
Tr × R	8, 72	1.43 (>.20)	1.82 (<.09)	1.74 (>.10)
Tr × S	4, 36	13.79 (<.0001)	5.79 (<.005)	5.31 (<.005)
R × S	32, 288	<1	<1	<1
Tr × R × S	32, 288	1.99 (<.005)	<1	<1

Note. Each entry shows the *F* value with *p* value in parentheses. The corresponding degrees of freedom (*df*) are shown in the second column. Tr = trial; R = rotation; S = slant.

minimal information. Participants were less good at reject-ing pairs in which the patterns were not affine related (less than 75% correct). Contrary to expectations, the number of false positives was not reduced by introducing color as a correspondence cue. This result suggests that the strategy that most participants used was to check the color of the elements for those patterns that seemed affine equivalent rather than first eliminating the pairs that were not affine equivalent by the color of their elements. As in visual search and apparent motion, shape seems to have priority over color. As might have been expected from that work, observers were also very slow in using color as a correspon-dence cue: On average, it took them over 5 s to correctly reject pairs of patterns that could not be made affine equivalent.

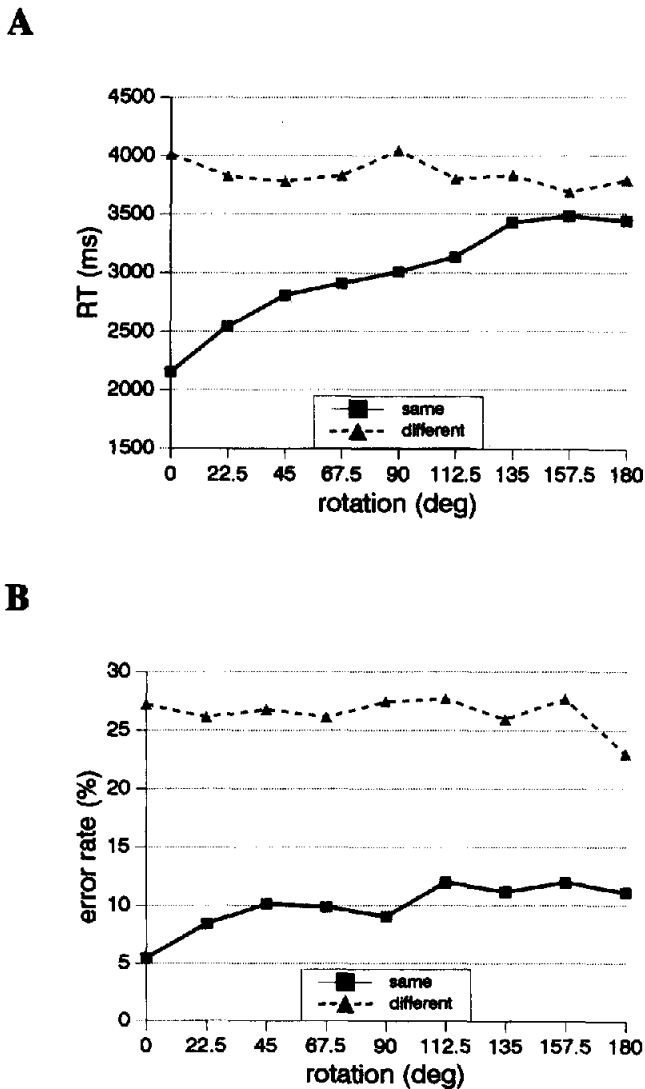The performance levels obtained in Experiment 1c were

A

B

Figure 9. The effect of slant on (A) response times (RTs in ms) and on (B) error rates (in %) for *same* trials (full line) and *different* trials (dashed line) in Experiments 1a–c.

A

B

Figure 8. The effect of Rotation on (A) response times (RTs in ms) and on (B) error rates (in %) for *same* trials (full line) and *different* trials (dashed line) in Experiments 1a–c.

between those of Experiments 1a and 1b; the ordering corresponds to the amount of information in the patterns (2 colors in 1c vs. 1 color in 1a and 4 colors in 1b). The ordering held for general performance level (81% vs. 80% and 85%), as well as for the proportion of correct identifica-tions of affine shape equivalence (89% vs. 85% and 96%) and more subtle variables such as response bias (58% vs. 56% and 60%), discrimination index d' (2.08 vs. 1.79 and 2.70), and the time to make a correct response (± 3 s vs. 2.5 and 3.5 s for *same* trials; ± 4 s vs. 3 and 5 s for *different* trials). These results suggest that task performance depends critically on the available information.

Experiment 2

In Experiment 1, the affine transformation under which shape equivalence had to be judged consisted of a combina-

tion of rotation and slant. Observers were able to perform the task, but it might be argued that the set of affine transformations was too restricted and that the ability to judge affine equivalence would be weaker if more general affine transformations were used. Testing this hypothesis was the main purpose of this experiment. The patterns, after being rotated and slanted, were now also tilted; that is, they were now rotated completely arbitrarily in 3-D space.

Two versions of the same experiment were run, both with the colors used in Experiment 1c (i.e., one blue and three black points). In Experiment 2a, viewing time was unlimited (as in Experiment 1), whereas in Experiment 2b it was limited to 500 ms. This relatively brief exposure constitutes a more demanding test of whether human perceivers can determine affine shape equivalence with minimal information.

Method

Twenty new participants were drawn from the same participant pool as in Experiment 1 (10 for Experiment 2a and 10 for 2b). Apparatus, task, and procedure were the same as in Experiment 1. Because the information contained in the four-point displays was properly minimal with one blue and three black points, this color combination was used again here. The only difference in stimuli between this experiment and Experiment 1c was that a third component, tilt, was added to the affine transformation (after rotation and slant; see Figure 4 again and Figure 10 for examples). In Experiment 2, the *tilt angle* τ had one of five values: 0°, 45°, 90°, 135°, and 180°. As before, each of these levels occurred equally often and was positive in about half of the trials (chosen randomly) and negative in the rest. In combination with rotation and slant, this parameterization yielded 225 conditions. In order that the total number of trials in this experiment was the same as in the previous one, each condition occurred only twice, that is, once for each trial type (*same* and *different*). Because of the absence of repeated measures here, we tested the effect of each factor combination by averaging across the third factor. The data were analyzed in two

Figure 10. An example of how the addition of a tilt component to an affine transformation affects shape equivalence between a randomly generated four-point pattern (shown on the left) and its transformed versions (shown on the right). Below each transformed version, the affine parameters (rotation, slant, and tilt) are indicated in degrees.

Table 3
Performance Levels in Experiment 2

Index	Experiment 2a		Experiment 2b	
	M	SD	M	SD
% CRT	80.92	5.06	76.70	2.77
% CI	90.75	5.03	75.55	4.67
% CR	71.08	12.81	77.85	6.70
% same	59.86	8.31	48.75	5.03
d'	1.97	0.25	1.48	0.19
RT$_{CI}$	2413	800	1260	369
RT$_{CR}$	3262	1880	1248	288
% CR*	85.09	13.95	88.19	7.75

Note. Each row shows data for a different performance index (from top to bottom): % CRT = total percentage correct responses; % CI = percentage correct identifications; % CR = percentage correct rejections; % same = percentage *same* responses; d' = discrimination index; RT$_{CI}$ = average response time (in ms) for correct identifications; RT$_{CR}$ = average response time (in ms) for correct rejections; and % CR* = percentage correct rejections of trials with convexity cue. Each entry represents the mean and standard deviations over 10 participants.

different ways: first, by averaging across tilt angles and performing ANOVAs on these averages with rotation and slant as the major variables (this analysis allows a comparison with the results from Experiment 1) and second, by averaging across rotation angles and performing ANOVAs on these averages with slant and tilt as the major variables (this analysis allows an assessment of this new component).

Results

General performance. The general performance in Experiment 2a was very similar to that in the comparable Experiment 1c. The overall average percentage of correct responses was 80.9% (see Table 3). As before, the average percentage of correct *same* responses was higher than the average percentage of correct *different* responses (90.8% vs. 71.1%), $t(9) = 3.75; p < .01$. Despite the addition of color to one of the points, there was still a bias toward responding *same* (59.9%). As in the previous experiments, there was a large negative correlation between average performance level and response bias for each participant, $r(9) = -.82$, $p < .01$. The average d' was 1.97 ($SD = 0.26$). This pattern of results changed in Experiment 2b, when viewing time was limited to 500 ms. The average percentage of correct responses was 76.7%, which is relatively high when one compares the 500 ms exposures with the 2–3 s taken by observers in Experiment 2a. For the first time, the average percentage of correct *same* responses was not higher than the average percentage of correct *different* responses (75.6% vs. 77.9%), $t(9) < 1$. Also for the first time, there was no response bias (48.8% *same* responses). Detectability of affine equivalence as measured by d' remained satisfactory (average $d' = 1.48$).

ANOVA on RTs and error rates. The several ANOVAs we performed to assess the effects of the different transformation parameters revealed a rather complicated pattern of results (see Table 4) that can be summarized as follows. First, RTs were generally less affected than error rates.
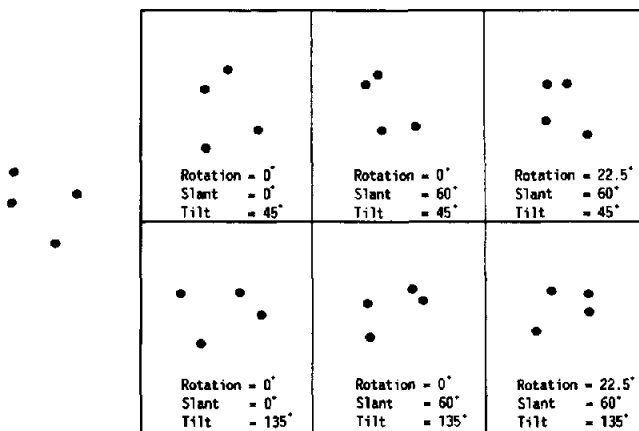
Table 4
Results of Analyses of Variance (ANOVA) on Response Times and Error Rates in Experiment 2

ANOVA	df	Experiment 2a	Experiment 2b
Response times			
ANOVA-1			
Tr	1, 9	4.45 (<.07)	<1
R	8, 72	<1	<1
S	4, 36	2.47 (<.07)	1.27 (>.25)
Tr × R	8, 72	1.68 (>.10)	1.74 (>.10)
Tr × S	4, 36	3.68 (<.05)	5.36 (<.005)
R × S	32, 288	<1	<1
Tr × R × S	32, 288	1.61 (<.05)	<1
ANOVA-2			
T	4, 36	1.47 (>.20)	<1
Tr × T	4, 36	<1	1.08 (>.35)
T × S	16, 144	<1	<1
Tr × T × S	16, 144	<1	<1
Error rates			
ANOVA-1			
Tr	1, 9	14.07 (<.005)	<1
R	8, 72	1.37 (>.20)	<1
S	4, 36	8.18 (<.0001)	43.15 (<.0001)
Tr × R	8, 72	1.20 (>.30)	2.98 (<.01)
Tr × S	4, 36	8.98 (<.0001)	31.77 (<.0001)
R × S	32, 288	<1	1.20 (>.20)
Tr × R × S	32, 288	<1	<1
ANOVA-2			
T	4, 36	<1	6.95 (<.0005)
Tr × T	4, 36	3.65 (<.05)	1.42 (>.20)
T × S	16, 144	1.53 (<.10)	1.33 (>.15)
Tr × T × S	16, 144	2.16 (<.01)	1.02 (>.40)

Note. Each entry shows the F value with p value in parentheses. The corresponding degrees of freedom (df) are shown in the second column. ANOVA-1 is based on averages across Tilt, while ANOVA-2 is based on averages across Rotation. To avoid redundancy, only the effects of Tilt are reported for ANOVA-2. Tr = trial; R = rotation; S = slant; T = tilt.

Second, Experiment 2b yielded stronger parametric effects than Experiment 2a. Third, the most robust effect, across all measures and subexperiments, was the Trial Type × Slant interaction. As in Experiment 1, this effect was caused by a strong effect of slant for *same* trials only. More specifically, more errors and longer RTs occurred when slant was 45° or 60°.

Discussion

Despite the extra transformation, the general levels of performance obtained in Experiment 2a were virtually identical to those of Experiment 1c (with equal color cues): The overall percentage of correct responses was 81% in both, the proportion correct identifications was 90% in both, discrimination index d' was 2 in both, and so on. If anything, participants responded more quickly in Experiment 2a than in Experiment 1c, $t(18) = 1.6$, $p > .10$, for the correct identifications, and $t(18) < 1$, for the correct rejections. Thus, it appears that an increased complexity of transformation does not make it more difficult to judge shape equivalence under that transformation.

Observers in previous experiments usually needed more than 2 s to respond that two patterns with four points were equivalent under an affine transformation and more than 3 s to reject random pairs. This may seem to imply that they were using a slow and effortful procedure that seems unnatural for normal shape perception. Yet, Experiment 2b (with 500 ms exposures) indicated that observers still perform reasonably well when such scrutiny is made impossible. The general performance levels obtained in this experiment were satisfactory; thus, the average percentage of correct responses was above 75% and d' was 1.5, despite the demanding nature of the task. That the percentage of correct responses to *same* pairs was only 75%, compared with 90% in Experiment 1c, $t(18) = 7.00$, $p < .001$, may be partly attributable to a larger bias toward responding *same* in Experiment 1c, $t(18) = 3.62$, $p < .01$. Moreover, the decrease in performance levels is small, given the reduction in viewing time (0.5 s vs. 3 s minus response-initiation time). This suggests that the checking procedures that participants probably used in the previous experiments were not necessary; their first impressions were often correct already. A single glance seems enough to determine affine shape equivalence between two patterns of four points.

Possible Strategies or Sources of Information

The main findings of this study so far are easy to summarize. In several experiments in which pairs of patterns of four points were presented side-by-side on a computer screen, observers were able to judge whether they were equivalent under an affine transformation. General performance levels (d') varied between 1.5 and 2.7 depending on the color information available for solving the correspondence problem (insufficient in Experiment 1a, superfluous in Experiment 1b, and minimal in Experiments 1c, 2a, 2b) and on the time available for verifying the first visual impression (unlimited in Experiments 1 and 2a and 500 ms in Experiment 2b). However, in comparable conditions, performance was essentially the same regardless of the complexity of the affine transformation (rotation and slant in Experiment 1 and rotation, slant, and tilt in Experiment 2). In sum, observers are sensitive to affine shape equivalence even with minimal stimulus information, with general affine transformations, and with limited viewing time.

Given that minimal information is sufficient for observers to determine affine shape equivalence, what mechanisms might they have used to achieve this performance? Although a definitive statement requires further research, the present results provide some clues already. We first discuss some possible sources of information or strategies that may have been used in Experiments 1 and 2 in the light of the available evidence. We then present Experiment 3 in an attempt to further dissociate these mechanisms.

Mental Transformations

As has already been noted, determining shape equivalence under affine transformations might be achieved by mentally undoing the transformations. Because the affine

relation between the patterns implies an arbitrary 3-D relation between the two planes, using mental transformations as a strategy to determine affine shape equivalence would amount to establishing a motion path in 3-D space, as if the observer imagined walking between the two stimulated viewing positions. Given the explosion of possible parameterizations of the affine group discussed in the section on *Affine Transformations*, such a strategy, applied exhaustively, is not very plausible on purely theoretical grounds.

The present data often showed significant nonlinearities, especially in the effect of slant. A significant increase in RTs and error rates was observed only when the slant angle σ reached 60° (giving a compression of 50%). Smaller slant angles had little effect (see Figure 9). This seems to contrast with the linear effects on RTs obtained in experiments with simple planar rotation or size scaling: "The typical, most intriguing result found in many mental rotation studies is the almost perfect linear increase of RT with angular stimulus disparity" (Wohlschläger & Wohlschläger, 1998, p. 397).

Furthermore, the introduction in Experiment 2 of an additional tilt component in the affine transformations under which shape equivalence had to be determined did not depress performance below that of Experiment 1 in which there was no tilt. Although it may seem natural to expect mental transformations to become more complex when the geometric transformations they have to undo become more complex, there are at least two possible mental-transformation strategies that do not have this property. First, it is possible, in principle, that the geometric transformation that is used in making the stimuli is decomposed mentally into fewer components or parameters. Thus, if observers were mentally to undo the rotation and tilt components that are used in Experiment 2 in one single *tilt-minus-rotation transformation*, one might find that the addition of tilt does not affect performance much. Second, as proposed by Parsons (1987) for the case of 3-D Shepard-like objects, it is possible that observers always mentally rotate the affine-transformed point patterns about an axis (unique to each 3-D orientation difference) that simultaneously corrects for all orientation differences in each of the separate components (i.e., rotation, slant, and tilt in our case). Again, if this so-called *mental rotation along the shortest path* were the mental-transformation procedure used by our observers, then similar performance levels might be obtained in Experiments 2a and 1c.

To test whether these specific mental-transformation procedures were used by our observers, post hoc ANOVAs and linear regressions were performed with the appropriately recoded variables. The details of this procedure, as well as some of the statistical results, are reported in Appendix B. The major findings are the following. First, a combined tilt-minus-rotation did not have a systematic effect on RTs. Second, RT was not a linear function, not even a monotonically increasing function, of the shortest 3-D angle in space. The shortest 3-D angle in space produced statistically reliable effects but the peaks and troughs of the resulting functions varied considerably with slant. Although there are many possible reasons to find nonlinearities even under a mental transformation account (e.g., Tarr, 1995), it seems

fair to conclude that the evidence for the use of mental transformations in space as the most important mechanism to determine affine shape equivalence between point patterns with minimal information is not very strong. It is important to understand this conclusion correctly: We do not claim that mental transformations are never used, only that performance in our experiments was probably not strongly dependent on mentally undoing the affine transformations relating the *same* shapes by rotating them in 3-D space. We return to this point in the General Discussion.

Numerical Invariants

The fact that general performance levels in Experiment 2a were not lower than those in Experiment 1c, despite the addition of a tilt component to the transformation, is consistent with the notion that judgments of affine shape equivalence are based on invariants, because invariance holds for the whole group of transformations. It is therefore possible to determine whether two patterns presented side-by-side on the screen are affine equivalent by computing one or more affine invariants for one pattern and determining whether the same values result for the other pattern. A good example of useful affine invariants are so-called *affine coordinates*. In addition to illustrating shape ambiguity and how color can solve it, Figure 5 shows how three points define an affine reference frame and how the fourth point retains affine-invariant coordinates expressed within this frame. When one of the points in a random four-point pattern is taken as the origin, *O*, two other points, *X* and *Y*, define two vectors, *OX* and *OY*. The coordinates of the remaining point, *P*, can then be expressed as a linear combination of the *OX* and *OY* vectors (i.e., 1.54 times *OX* and 0.35 times *OY*). In the other pattern, which is an affine transformed version of the first, the same point *P* after transformation, *P'*, has identical affine coordinates expressed in terms of the two transformed vectors (i.e., 1.54 *O'X'* and 0.35 *O'Y'*). Two important observations in relation to these affine-invariant coordinates should be made.

First, this alignment scheme exploits parallelism as an affine-invariant property as well as the fact that an affine transformation preserves relative distances between three collinear points. In other words, the ratio between two adjacent segments on a given line is affine invariant (i.e., the three-point analogue to the four-point cross-ratio for the projective case; see Van Gool et al., 1994, for an elaboration of this point). Lappin and Fuqua (1983) have already demonstrated that observers can accurately decide whether the middle point of three collinear points rotating in a slanted plane is centered exactly between the other two points in 3-D space (see also Wagemans & Tibau, 1999).

Second, although the measurements themselves are invariant, the processes that are used to provide the measurements do not have to be independent of specific affine transformations. For example, the alignment strategy assumes that correct matches can be found between the points in the original pattern and the corresponding points in the transformed version. Establishing these correspondences might take time and perhaps more so when the affine transforma-

tion has larger parametric values (e.g., larger rotation, slant, and tilt angles). Moreover, Figure 5 suggests that some aspects of the particular configuration might affect the difficulty of the measurements. For example, the position of the fourth point relative to the two vectors will be more difficult to establish when the points are farther apart and when the angles between the two vectors deviate more from 90°. Obviously, these distances and angles are changed by the affine transformation. Therefore, although these metric aspects are irrelevant to the invariants themselves, they might affect the calculation of the invariants by the visual system. In fact, that is exactly what was demonstrated in a study in which observers were explicitly instructed to give affine-invariant coordinates; hence, mental transformations were excluded because no comparisons had to be made (Wagemans, Van Gool, & Lamote, 1996).

It is not easy to provide more direct support for the use of invariants unless one explicitly manipulates them (which was not done in our experiments). In principle, one could design experiments in which the discrimination would not be between perfectly affine-related pairs and completely unrelated pairs (the present design) but between perfectly affine-related pairs and pairs in which the affine-related one would be systematically distorted. As can be seen in Figure 5, distortion of one of the affine coordinates (e.g., along the OX vector) while keeping the other unchanged (i.e., on OY) is possible by repositioning the fourth point, P, on a line parallel to the line through O and X (and vice versa for the distortion of the other affine coordinate). If those distortions were harder to discriminate from perfectly affine-related pairs than more random perturbations (i.e., in both directions at the same time), this might constitute more direct evidence for the use of these coordinates by the human visual system. Within the framework of the present experiments, it is difficult to determine whether such numerical invariants were influential in determining observers' performance.

In sum, not much can be said in favor of or against the use of invariants on the basis of the present data except to note that there are few other cues available with minimal displays of the kind used in the present experiments. One special class of cue is discussed next.

Qualitatively Invariant and Quasi-Invariant Properties

We noticed that observers found some trials easier than others, irrespective of the amount of affine distortion. For example, when one pattern in the display was convex and the other was concave, it was easy to reject them as *different:* No affine transformation exists that can change a convex quadrilateral into a concave one, and vice versa. Also, if the point pattern happened to have two points close together or three points almost collinear, this property was easy to verify for the other pattern: If it was also present, participants tended to respond *same;* if it was not, they tended to respond *different* (likewise for parallel configurations). Convexity, proximity, collinearity, and parallelism are referred to as *qualitative invariants* (or *nonaccidental properties;* see Biederman, 1987; Lowe, 1987; Wagemans, 1992, 1993) to

distinguish them from the numerical invariants discussed in the preceding section. This notion may be clarified by considering the properties of convexity/concavity, parallelism, and collinearity in turn.

Convexity/concavity. The number of sides of the smallest polygon enclosing four points, the *convex hull,* is preserved by all affine transformations. The convex hull is a quadrilateral with a convex pattern of four points and a triangle with a concave pattern; this property is perfectly invariant. But it is such a general property of shape that it is of little help in identification (it has little discriminatory value): Not all convex quadrilaterals are the same, nor all concave ones. In the present experiments, however, convexity/concavity could have been used to reject at least some *different* pairs (for an example, see Figure 2, row *affine,* column *sufficient*). To determine whether participants used such a cue, the trials in which the convexity/concavity cue was available were isolated and the proportion of correct rejections computed. They are shown in the bottom row in Tables 1 and 3 with the general performance data (% CR* = percentage correct rejections of trials with convexity cue). In all experiments, this rejection rate was higher than the average rejection rate computed across all trials; more specific quantitative tests (t tests between dependent samples) confirmed this observation.

Three other properties associated with the use of convexity/concavity are worth noting. First, participants who seemed to have used it (high % CR*) tended to have a lower bias toward responding *same* and shorter RTs, a result that was confirmed statistically by the appropriate correlation coefficients. These participants may thus have used a "quick-and-dirty" strategy to reject some trials instead of a time-consuming process to determine the patterns' congruence. Second, the specific stimulus manipulations in the experiments seem to have led to different uses of the convexity/concavity cue. For example, in Experiment 1b, where participants were possibly distracted by the gamut of colors, interindividual differences associated with the convexity/concavity cue (i.e., % CR*) were much larger; however, in Experiment 2b, where participants were forced to rely on their first impressions, use of the convexity/concavity cue to reject *different* patterns was more common and, typically, the few participants who made less use of it performed more poorly. Third, in Niall's studies on projective shape equivalence (Niall, 1992; Niall & Macnamara, 1990), only concave quadrilaterals and pentagons were used; this restriction might have been a factor contributing to the different performance levels he recorded.

Parallelism and collinearity. If parallelism and collinearity happened to be perfectly represented in the original four-point pattern, they would have been perfectly preserved following affine transformation. As such, they are also perfectly invariant. But, like convexity/concavity, the properties of parallelism and collinearity are qualitative in the sense that they are generally insufficient to define shape adequately. Moreover, because in most random-point patterns parallelism and collinearity were not perfect but only approximate (i.e., small instead of zero angular deviations), these properties are better considered as quasi-invariant;
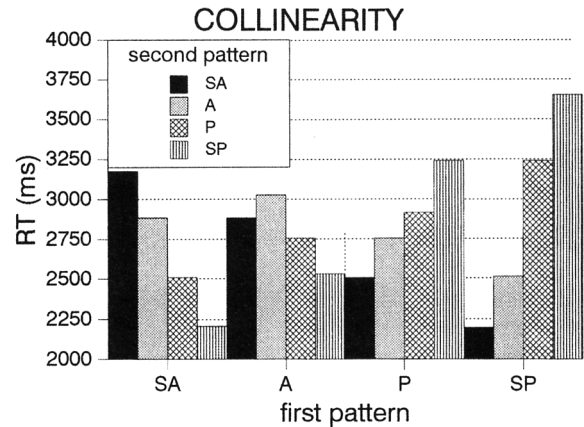
under some affine transformations, deviations from close-to-perfect parallelism or collinearity may become large and exceed the threshold for a visual judgment of parallelism or collinearity. Nevertheless, observers could have used perfect or imperfect regularities as qualitatively invariant or quasi-invariant properties in the affine shape-matching task. The nonlinearities in the effects of affine parameters on error rates and RTs are consistent with that proposal (e.g., 60° slants are more likely to destroy imperfect parallelism and collinearity). Quasi-collinearity has also been found to be a spatial primitive in texture discrimination (Caelli & Julesz, 1978) and collinearity a spatial primitive in shape-from-motion displays as well (Lappin, Ahlström, Craft, & Tschantz, 1995).

Post hoc analyses were performed to determine whether parallelism and collinearity played a role in observer performance. These analyses proceeded in three steps. First, algorithms were developed to quantify the degree of parallelism and collinearity in each pattern consisting of four points. To measure the degree of parallelism, virtual quadrangles were constructed and the angular deviations from all possible virtual parallelograms were measured. The smallest angular deviation was retained to express the strongest possible parallelism. Similarly, to measure the degree of collinearity, all triplets of points were considered. Collinearity was computed as a function of two parameters: (a) the perpendicular distance of the middle point (point 2) relative to the line segment connecting the other two points (points 1 and 3) and (b) the smallest angle separating line segment 1–2 from 1–3 or line segment 2–3 from 1–3. Again, only the smallest deviation from perfect collinearity in all four triplets was retained to express the degree of collinearity in the four-point patterns.

In a second step, an independent set of 10 observers was asked to rate (on a 6-point scale) the degree of parallelism and collinearity in 120 point patterns generated by the same algorithm as the one used to generate the experimental stimuli. In accord with the notion of a fuzzy geometry (Ferraro & Foster, 1994), parallelism and collinearity were treated as fuzzy regularities, and four fuzzy sets of patterns were determined in which these regularities were either strongly absent or just absent (SA and A, respectively) versus strongly present or just present (SP and P, respectively). This was done by calculating fuzzy membership functions on the basis of the observers' ratings and the associated computed values of parallelism and collinearity (see Kandel, 1986, for the mathematical techniques and Garmendia & Van Bockstaele, 1993, for this particular application).

In a third step, using the criteria provided by the independent raters, each pattern pair presented in the present experiments was coded as SA-SA, SA-A, SA-P, SA-SP, A-A, A-P, A-SP, P-P, P-SP, or SP-SP. The trials belonging to these 10 categories were brought together for each participant's data file and the proportion of correct responses as well as the average RT were entered in an ANOVA across participants. As an example, Figure 11 shows the results for collinearity in the *different* trials of Experiment 1a. The way to interpret this figure is the following. When collinearity
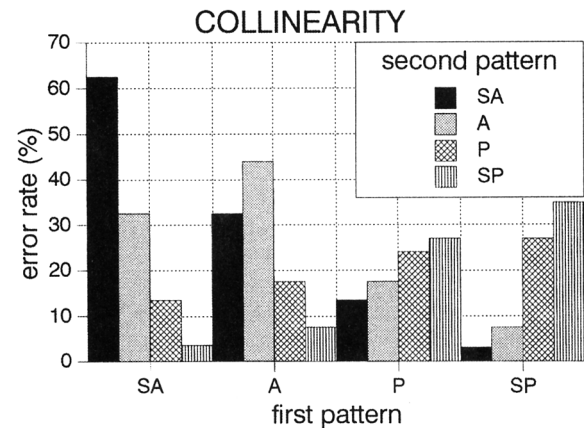
A

B

Figure 11. Performance levels for *different* trials of Experiment 1a as a function of the relationship between the patterns according to whether collinearity was strongly present (SP), present (P), absent (A), or strongly absent (SA). The top panel (A) shows the average RT (in ms) for correct responses, and (B) shows the error rate (in %).
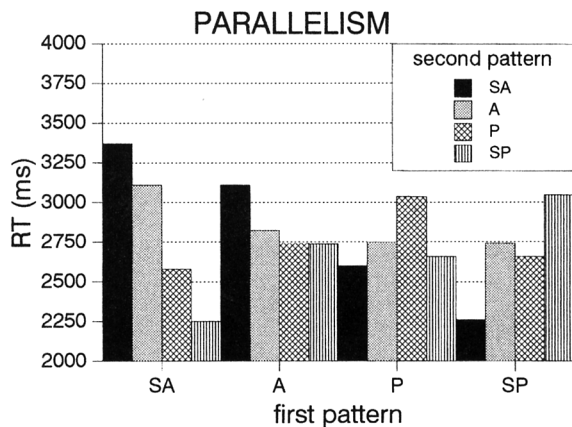
was strongly present in one of the two patterns (rightmost set of bars), it became easier to determine that the pair was probably *different* (i.e., not related by an affine transformation), as the other pattern showed increasingly less collinearity. In (A) it can be seen that RT became progressively shorter as the other pattern was categorized as SP, P, A, and SA, respectively. In (B) it can be seen that there was a similar decrease in error rate as well. The opposite pattern of results holds when the first pattern had a large deviation from perfect collinearity (i.e., SA in the leftmost columns) and intermediate results are found for the intermediate categories. Closely similar findings were obtained for parallelism (see Figure 12). The pattern of results as described was corroborated statistically by the overall ANOVAs as well as by the appropriate trend analyses.

Proximity. Affine transformations do not preserve absolute distances, so a property like proximity, which is obviously based on distance, is not truly affine invariant. Even so, if two points are very close together and distant from a third, this relative proximity—when compared to the distance to and between other points—is preserved under a wide range of viewpoints. It can, however, be destroyed under some affine transformations (e.g., at large values of slant angle σ). Although grouping by proximity is generally a powerful principle (e.g., Kubovy, Holcombe, & Wagemans, 1998; Kubovy & Wagemans, 1995), the effect of relative proximity in the present experiments was less pronounced than with the other qualitatively invariant properties because strong proximities were avoided in the patterns used here (see *Methods*). For example, fewer of the

A

B

Figure 12. Performance levels for *different* trials of Experiment 1a as a function of the relationship between the patterns according to whether parallelism was strongly present (SP), present (P), absent (A), or strongly absent (SA). The top panel (A) shows the average RT (in ms) for correct responses, and (B) shows the error rate (in %).

statistical tests in the post hoc analyses reached standard levels of reliability.

Tests of qualitative cues. Because the results of these post hoc analyses indicating the use of qualitatively invariant or quasi-invariant properties were essentially inferential, we have undertaken experiments to test their use explicitly. In previous work, Foster (1980a; Ferraro & Foster, 1984) has suggested that binary-valued cues (also called *discrete attributes*) such as collinearity determine maxima in performance in some simple tasks requiring point-discrimination (Foster, 1979), angle-discrimination (Foster, 1980b), and curved-line discrimination (Foster, 1983). Pursuing this approach, we have used a perturbation technique (Foster, 1980b) that progressively distorted pairs of *same* patterns, degrading each qualitatively invariant property. The results of these experiments (Kukkonen, Foster, Wood, Wagemans, & Van Gool, 1996) confirmed the special role of qualitative cues. For example, clear peaks and troughs in performance were obtained where patterns changed from parallel to nonparallel and from having collinear points to noncollinear points. This behavior agrees well with the notion that qualitative invariants correspond to certain singular values of numerical invariants (see Van Gool et al., 1994, for a further discussion). Thus, qualitative invariants may have a special status for the visual system because they simplify otherwise hard-to-compute numerical invariants.
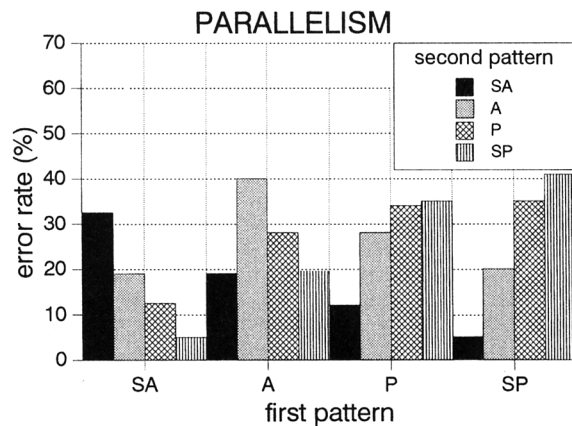
Experiment 3

In the light of the foregoing analyses, it seems that qualitative properties like convexity/concavity, parallelism, and collinearity played the most important role; numerical invariants or mental transformations were probably only used when qualitatively invariant or quasi-invariant properties were absent or to verify a first guess on the basis of their presence when viewing time was unlimited. The aim of Experiment 3 was to further substantiate this tentative conclusion. We do this by creating conditions that will affect the usefulness of qualitative and quantitative affine invariants and of mental transformations.

One way in which the use of *mental transformations* could be facilitated is to group together trials that share a particular geometric transformation. Once the observer knows the particular geometric transformation relating a pair of point patterns that are affine equivalent, it should be easy to verify in any given trial that this transformation is applicable and to decide *same* if it succeeds and *different* if it does not. Blocking the trials may also facilitate the establishing of corresponding points, which is needed to extract affine-invariant coordinates within an affine reference frame.

The question of whether it is possible to prepare for a specific orientation has been important in the established literature on mental rotation (e.g., Cooper & Shepard, 1973, 1975; Koriat & Norman, 1988) as well as in more recent work on orientation-dependent object identification (e.g., Gauthier & Tarr, 1997; McMullen, Hamm, & Jolicoeur, 1995; Murray, Jolicoeur, McMullen, & Ingleton, 1993). The general conclusion from this strand of research is that linear effects on RT of angular disparity between two comparison

shapes or between a target object for identification and the standard upright orientation stored in memory are flattened only when participants can prepare for a specific shape at a specific orientation; they cannot rotate an abstract frame of reference (for a review, see Gauthier & Tarr, 1997).

These results, however, should not be taken to predict that blocking will have no effect in our experimental paradigm. The reason we believe mental transformations in space are so unlikely as the major mechanism to determine affine shape equivalence is that the number of possible transformations to be tried is excessive. In Experiment 1, when our affine parameterization would have been used, there were 45 combinations to be tested, yielding 61 differently oriented axes if mental rotation through the shortest path in space were attempted. In Experiment 2, these numbers were 225 and 83, respectively. In sharp contrast, most of the literature on orientation priming in mental rotation or object identification has been restricted to simple planar rotations (i.e., 1 axis, with 5 to 11 rotation angles). Thus, it should help to reduce uncertainty about the affine parameters to be undone mentally or about the orientation of the rotation axis in space to be used for mental rotation of one shape into the other, if all trials were grouped in blocks of uniform transformations.

Even the established literature on mental rotation contains some clues that this blocking should have a considerable effect. For example, Metzler and Shepard (1974/1982) compared conditions with an unpredictable axis (i.e., blocks with picture-plane rotations and depth rotations mixed together) with fixed-axis conditions (i.e., pure blocks). Slopes were steeper in mixed blocks than in pure blocks (i.e., RTs were 20% slower). According to Metzler and Shepard (1974/1982, p. 49), this difference in slope reflects "the additional time required for prerotational search when the axis of rotation was not known in advance." If such effects are possible with uncertainty about only two rotation axes, it is quite logical to expect more dramatic effects with our large numbers of rotation axes, especially because our experiments included arbitrary axes (i.e., not aligned with canonical frames of references), which are known to be more difficult (e.g., Pani, 1993; Pani, William, & Shippey, 1995; Parsons, 1995).

One way in which the use of affine invariants would be made impossible is to use point patterns consisting of three points instead of four. As outlined in the section on Affine Transformations, all triangles are affine equivalent, and it is thus impossible to perform a shape-discrimination task with three-point patterns. But this mathematical constraint holds true only for the group of all affine transformations, with no restrictions on the parameterization of the real numbers in the 2 × 2 matrix of the general affine transformation. Because our sampling of affine parameters is limited, it may be that observers do have an appreciation of what constitutes affine shape equivalence but one that is based not on affine invariants as such but on similarity approximations. Consider, for example, the three-point patterns in the upper left box of Figure 2 (i.e., column sufficient, row similarity): The pattern on the left looks more similar to the pattern in the middle than the pattern on the right. Observers' judgment of

affine shape equivalence may be influenced by shape equivalence under similarity transformations.

Simpson (1986) introduced conditions with three collinear points rotating in depth to test the relevance of the cross-ratio, a projective invariant for four collinear points, as a source of information about objects in depth. He obtained equal performance levels in both conditions (i.e., with three and four points) and argued that the cross-ratio cannot be necessary to solve the task. On the basis of additional experimental findings, he concluded that performance was based instead on a perceptual analysis of the sinusoidally changing point positions (not unlike Todd's, 1982, trajectory analysis for interpreting structure from motion). Although Simpson's displays and task are quite different from ours, we considered his method of removing information that is essential to extract a particular type of invariant useful for our purposes too.

In sum, if mental transformations play a major role in determining affine shape equivalence, changing the design from mixed to blocked presentation of trials should result in considerably improved performance. On the other hand, if qualitative and quantitative affine invariants play a major role, changing the stimulus to three-point patterns should result in considerably reduced performance. Table 5 summarizes the rationale behind Experiment 3 in terms of the stimulus information available and plausible observer strategies. Note that some qualitatively invariant properties (e.g., quasi-collinearity, strong relative proximity) are still possible with three points.

Method

Participants. Eight graduate students from the Department of Psychology at the University of Leuven volunteered to participate. They were naive to the purpose and details of the experiment, and each had normal or corrected-to-normal vision. Four participants were assigned randomly to each of the two between-subjects conditions.

Apparatus and stimuli. This experiment was run on a computer with an SVGA graphics board. The point patterns were created in the same way as in Experiments 1 and 2, except for the following differences. They were presented as black or yellow points on a gray background on a color screen with a 50 Hz temporal resolution and an 800 × 600 spatial resolution. When the point patterns consisted of four points, three were black and one was

Table 5
Summary of the Rationale in Experiment 3

Stimulus patterns/ possible strategy	Mixed design	Blocked design
4 points		
Qualitative affine invariants	present	present
Quantitative affine invariants	present	present
Similarity approximations	possible	possible
Mental transformations	unlikely	more likely
3 points		
Qualitative affine invariants	fewer	fewer
Quantitative affine invariants	absent	absent
Similarity approximations	possible	possible
Mental transformations	unlikely	more likely

yellow to facilitate correspondence; when the patterns consisted of three points, two were black, and one was yellow. The affine transformation was parameterized as before with rotation, slant, and tilt applied in that order and with the following levels: Rotation φ was 0°, 45°, 90°, 135°, or 180°; slant σ was 0°, 41.41°, or 60° (i.e., the three levels were equidistant on a cosine scale); and tilt τ was 0°, 45°, 90°, 135°, or 180°. Participants responded by pressing one of the two horizontal arrow keys on the computer's keyboard.

Task and procedure. The task was the same as in Experiments 1 and 2. Observers were not mathematically informed, and they did not question the appropriateness of using three versus four points in determining affine shape equivalence. Four participants saw only three-point patterns; the remaining four saw only four-point patterns. Each of the 75 parameter combinations (5 rotation angles, 3 slants, and 5 tilts) of the affine transformation occurred 20 times: ten times for *same* trials and 10 times for *different* trials. Each of the 8 observers performed these 1,500 trials twice, once with a mixed design and once with a blocked design. Two observers in each group of four started with the mixed design and then continued with the blocked design, whereas the remaining two used the opposite order.

In the *mixed* design, all trials were randomly mixed together in a block of trials, as in Experiments 1 and 2. The total number of trials in this condition was divided into five blocks of 300 trials that could be run without a break. Short breaks of 2–3 min separated these blocks that were otherwise based on uniform samples from the total population of experimental trials. In the *blocked* design, trials were grouped together according to the particular affine transformation. Each sample of 20 trials with a particular combination of affine parameter values was preceded by an indication of those values (i.e., rotation, slant, and tilt angle in degrees) and an example trial. At the end of each block of 20 trials, observers could take a break if they wished but in general they continued.

Results and Discussion

Table 6 summarizes the most important indexes of general performance level in the different experimental conditions. As expected, the task was easier with four-point patterns than with three-point patterns: Average d' was 2.25 compared with 1.69, $F(1, 6) = 8.79$, $p < .05$, and average RT was 2,094 ms compared with 2,635 ms, $F(1, 6) = 1.71, p > .20$. More importantly, the effect of blocking trials according to affine parameter values yielded a smaller effect: d' increased somewhat (from 1.89 to 2.05), and average RT decreased somewhat (from 2,503 ms to 2,225 ms) but neither effect was statistically reliable, $F(1, 6) = 1.96, p > .20$, for d', and $F(1, 6) = 1.20, p > .30$, for RT.

Table 6
Performance Levels in Experiment 3

Stimulus patterns	Mixed	Blocked	Average
Discriminability index (d')			
4 points	2.069	2.425	2.247
3 points	1.708	1.679	1.694
	1.888	2.052	
Response times (ms)			
4 points	2.347	1.842	2.094
3 points	2.660	2.609	2.635
	2.503	2.225	

That performance levels were still satisfactory (i.e., d' around 1.7) with only three points may seem strange in the light of our earlier arguments about four points being the minimal number of points to establish affine shape equivalence. The reason is that the experiment did not test *general* affine equivalence but included only a subset from the space of affine parameters. Slant angle was limited to 60° and only three different levels were used. This limits the range of possible triangles that can correspond to the target three-point configuration. This also leaves a number of qualitatively invariant properties intact under a considerable subset of the transformations. For example, quasi-collinearities and strong relative proximities are still possible in three-point configurations, and they would be preserved under all but the most extreme slant.

We do not want to argue that blocking did not have an effect, whereas stimulus information did (for it may only be a matter of having more observations to make the blocking effect statistically reliable too), but it will be clear that having four points in a pattern has a larger impact on task performance than blocking the trials according to the affine transformation parameters. From the rationale set out earlier, this result suggests that judgments of affine shape equivalence with minimal-information displays are based more on affine invariants than on mental transformations. This conclusion is consistent with the post hoc analyses of task performance in the previous experiments.

General Discussion

Summary and Conclusions

This study was concerned with the ability of observers to determine whether two simultaneously presented planar shapes are projections of the same shape at different 3-D orientations under orthographic projection and with only minimal information available. In short, can observers determine the affine shape equivalence of two four-point patterns? In Experiment 1, where the set of affine transformations to be considered was restricted to combinations of planar rotation and slant about a horizontal axis, observers responded correctly in about 80–85% of the trials (and 85–95% of *same* trials), depending on whether color was added to one or more of the points to facilitate correct correspondences. When affine transformations were made more general by adding a third rotation component (tilt), so as to simulate completely arbitrary 3-D orientations, performance remained around 80% correct responses across all trials (and 90% correct identifications). When viewing time was limited to 500 ms, performance levels remained around 75% correct (for *different* trials as well as *same* trials), even though participants took 2–5 s when viewing time was unlimited. Response-bias free measures of performance level (d') ranged between 1.5 and 2.7. Although far from perfect, they demonstrate that observers are indeed relatively good at determining whether four-point patterns can be related by an affine transformation.

Extensive post hoc analyses of the data in Experiments 1 and 2 suggested that performance was most probably based

on a mixture of strategies. For example, some *different* trials were easy because in one pattern the four points formed a quadrilateral configuration, whereas in the other pattern they formed a triangular configuration with one point inside. Because affine transformations cannot deform a convex quadrilateral into a concave one, convexity/concavity is an affine-invariant property. In all conditions of Experiments 1 and 2, the correct-rejection rate of trials with this convexity cue was higher than that of *different* trials in general, especially when color did not distract attention away from qualitative shape properties and when insufficient time was available for more elaborate, fine shape judgments. Approximate parallelisms and collinearities, although only quasi-invariant (i.e., qualitatively invariant under most affine transformations except at the most extreme parameter values) were also found to be useful. Because these qualitative properties were associated with relatively faster responses, we have referred to their use as a quick-and-dirty strategy.

When these qualitatively or quasi-invariant properties were unavailable, performance was still better than chance. This suggests that a slower, more effortful procedure was also used by at least some observers in some conditions. Two alternative strategies were distinguished. First, affine shape equivalence could be determined on the basis of numerical affine invariants such as affine coordinates. Second, the ability to determine whether two simultaneously presented four-point patterns could be projections of the same pattern at two different 3-D orientations could be based on mentally transforming one into the other so as to undo the geometric transformation. To test these possible strategies underlying affine shape matching, Experiment 3 was performed. Here, three-point patterns were presented to make quantitative affine invariants uninformative, and trials were blocked according to specific 3-D orientations to facilitate the use of mental transformations. In general, performance was more influenced by the type of stimulus (three- versus four-point patterns) than by the presentation condition (mixed versus blocked trials), suggesting that performance was based more on the use of affine invariants than on the use of mental transformations. Similarities based on some qualitatively invariant properties were still possible.

In the remainder of the article, we discuss our findings in relation to other recent work. A conciliatory view results from that discussion.

Mental Transformations Versus Invariants

We do not deny the existence of mental transformations. Indeed, the original results obtained by Shepard, Cooper, and their colleagues constitute strong evidence that a mental operation analogue to a physical rotation can be used to judge whether two differently oriented objects are identical or mirror-reflected or whether a single shape presented at a novel orientation is a standard or mirror-reflected version (see Shepard & Cooper, 1982, for a collection of papers). In addition to the experimental separation of prestimulus and poststimulus rotation (Cooper, 1975; Cooper & Shepard, 1973) and the presentation of test objects on the path of mental rotation (Cooper, 1976), the linearity of the functions

relating RTs to angular disparity was crucial. Larsen and Bundesen (1998) emphasized this point, referring to additive-factors logic (Sternberg, 1998), when they discussed our article on the distinction between invariants as mathematical entities and invariants as measured by the visual system (Wagemans et al., 1996). However, in that article as well as in the present one, we did not argue against the mental rotation interpretation of classic, linear effects of angular disparity on RT. We argued only that effects of transformation parameters on task performance (error rate or RT) need not be taken as evidence against the use of invariants by the visual system.

In recent work on viewpoint-dependent effects in object recognition, it was argued that other normalization procedures such as alignment (e.g., Ullman, 1989) and view interpolation (e.g., Edelman & Bülthoff, 1992; Poggio & Edelman, 1990) do not predict a strict linear relationship between orientation and recognition performance (e.g., Hayward & Tarr, 1997; Tarr, Bülthoff, Zabinski, & Blanz, 1997). In our own laboratory, we have provided evidence that other normalization procedures are more likely than mental rotation along the shortest axis in space, even with solid 3-D objects rendered with full depth cues (Willems & Wagemans, 1998). This result makes the use of mental rotation along the shortest axis in space with sparse four-point configurations as used here even less likely, but it does not rule out other normalization procedures. Indeed, the nonlinear effects often obtained in the experiments presented here are compatible with that view.

Niall (1997) has demonstrated that mental rotation is neither necessary nor sufficient to explain changes in RTs for the simultaneous comparison of planar shapes depicted in depth. When two planar shapes were depicted as separated by a small and fixed angular slant difference in depth and then depicted as tilted in depth together (in so-called *tandem rotation*), RTs varied nearly linearly with the magnitude of the tilt in depth, even with constant angular differences. This result suggests that comparison of shapes in different 3-D orientations can vary as a function of slant and tilt without implying the use of mental transformations in 3-D space to undo the transformations. The nonlinear effect of slant obtained in the present study (e.g., Figure 9), as well as the remaining effect of slant when RTs are plotted as a function of 3-D angle (Figures B2 and B4), is compatible with Niall's results.

Other recent work has added to the evidence that mental rotation has restricted usefulness as a procedure in judging general shape equivalence or in object identification from different viewpoints. For example, Parsons (1995) has indicated that individuals of high spatial ability are generally unable to imagine even 3-D minicube objects rotated about an axis and angle so as to accurately represent a new orientation or to find an axis and angle of shortest path along which such objects would have to be rotated to make them congruent (see also Pani, 1993; Pani et al., 1995). Moreover, in a recent review of shape constancy and object identification across rotations in the plane and in depth, Lawson (1999) has argued that all of the attempts to demonstrate the

use of mental rotation to undo these orientation differences have failed.

At the same time, we do not want to argue that we have now proved that affine shape matching is based on affine invariants. Indeed, it is difficult to prove the use of invariants unless one directly manipulates them (see also Wagemans & Tibau, 1999). Even so, it is evident that an invariants-based approach can be useful, and its geometric foundations can lead to predictions that are testable and relevant. Thus, the idea of minimal information is based on a hierarchy of transformation geometries (which has found application in other contexts; e.g., Tittle, Todd, Perotti, & Norman, 1995), and this framework allows one to derive predictions about shape equivalence under different sets of transformations (e.g., perspective versus projective; Wagemans, Lamote, & Van Gool, 1997; also Foster, 1975).

Reconciliation Attempts

In addition to the above qualifications about mental transformations and other normalization procedures, it is important to avoid other misunderstandings about our position. We have argued only that mental rotation along the shortest angle in space is unlikely to be the basis of performance in our experiments in which participants had to determine affine shape equivalence between four-point configurations with no cue about orientation in depth other than the possible affine transformation to be recovered from the difference between the two test patterns. This does not mean that mental transformations would never be useful in tasks similar to the present one. In fact, in subsequent experiments (Van Campenhout, Wagemans, Kyllingsbaek, Bundesen, & Larsen, 1998), we have reduced the differences between the present paradigm, and the one that is used in the mental-transformation literature (e.g., closed shapes to enhance depth perception, mirror-reflected distractors to make invariants uninformative, 3-D rotation angles and axes manipulated directly instead of derived post hoc). In that case, we did obtain strong linear effects of 3-D rotation angle on RTs and error rates.

In another recent study (Vanrie, Béatse, Wagemans, Sunaert, & Van Hecke, 1999), we have obtained additional evidence that both procedures, mental rotation and the use of invariant features, can occur in similar experimental circumstances with fully rendered 3-D objects instead of planar point patterns or polygons. Using 3-D block figures similar to the ones used by Tarr (1995) as standard objects, we created derived versions that were either mirrored or skewed. In the skewed versions, the angles between the different parts of the block figures were not 90° but 80° or 100°. Although a single 3-D angle is not invariant across viewpoint changes, the property of all angles being orthogonal or skewed is a viewpoint-invariant characteristic. When standard and mirrored versions had to be discriminated from different viewpoints, RTs and error rates were linearly increasing functions of angular disparity. When orthogonal and skewed block figures had to be matched, there was no effect of angular disparity on RT or error rate. When

participants performed the same experiment in an fMRI scanner, the patterns of brain activation associated with the two tasks could also be dissociated clearly.

The foregoing is one way to reconcile the mental-transformation account with the invariant-features account of shape equivalence: In short, their applicability depends on the experimental context. Another reconciliation is offered by Corballis (1988) and Jolicoeur (1990). They argued that mental transformations are only used to check the possible correspondence between shapes as established preattentively on the basis of invariant features. This hypothesis makes much sense for it helps to explain how participants can know the direction in which to rotate mentally (i.e., RTs are an inverted V-shaped function of angular disparity with the peak at 180°). The evidence in support of this hypothesis is, however, equivocal: Whereas some studies have produced results consistent with this view (e.g., Hamm & McMullen, 1998; Stankiewicz, Hummel, & Cooper, 1998), other studies have shown that normalization does not always follow after some form of match or identification has occurred (e.g., Gibson & Peterson, 1994).

Finally, the major finding that qualitative or quasi-invariant properties such as concavity/convexity, parallelism, collinearity, and proximity are the most useful source of information, even with sparse displays such as four-point patterns, is informative with regard to everyday object recognition under different viewpoint changes. These properties are more common in nonrandom objects, and they form the basis of Biederman's (1987) recognition-by-components theory of object recognition (see Norman, Todd, Perotti, & Tittle, 1996, for a similar argument). The controversy that has surrounded this theory is to a large extent based on arguments about the relevance of the particular stimulus sets used to test viewpoint-independent versus viewpoint-dependent models of object recognition (e.g., Biederman & Gerhardstein, 1993, 1995; Hayward & Tarr, 1997; Tarr & Bülthoff, 1995).

At the same time, qualitative properties are very important in recent view-based approaches to object recognition (e.g., Edelman, 1995a, 1995b; Gauthier & Tarr, 1997; Hayward, 1998; Hayward & Tarr, 1997; Tarr et al., 1997). These approaches emphasize not complete, 2-D template-like representations but rather sets of qualitatively distinct features (see Wallis & Bülthoff, 1999, for an excellent review of the psychophysical and neurophysiological evidence). For example, Hayward (1998, p. 439) argued that "viewpoint-dependent representations might involve the qualitative encoding of object shape" and he identified as a "crucial issue, still unresolved, [the] explication of the specific features that will be normalized in an image" (p. 438). We propose as a hypothesis for further research that qualitative properties, which we have discovered to play a role in the recognition of simple four-point patterns may also be a useful source of information in more realistic images. In this sense, they may well be the basis of both viewpoint-independency in Biederman's RBC account and viewpoint-dependency in multiple-view approaches.

References

Atkinson, J., & Braddick, O. J. (1989). "Where" and "what" in visual search. *Perception, 18,* 181–189.

Attneave, F., & Frost, R. (1969). The determination of perceived tridimensional orientation by minimum criteria. *Perception & Psychophysics, 6,* 391–396.

Bennett, B. M., Hoffman, D. D., Nicola, J. E., & Prakash, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America, A6,* 1052–1069.

Bennett, B. M., Hoffman, D. D., & Prakash, C. (1993). Recognition polynomials. *Journal of the Optical Society of America, A10,* 759–764.

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: Reply to Tarr and Bülthoff. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1506–1514.

Binford, T. O., & Levitt, T. S. (1993). Quasi-invariants: Theory and explanation. *Proceedings of the DARPA Image Understanding Workshop* (pp. 819–829). Washington, DC: Defense Advanced Research Projects Agency.

Börjesson, E., & Lind, M. (1996). The effect of polar projection on the perception of Euclidean structure from motion. *Perception & Psychophysics, 58,* 871–882.

Braunstein, M. L., Hoffman, D. D., & Pollick, F. E. (1990). Discriminating rigid from nonrigid motion: Minimum points and views. *Perception & Psychophysics, 47,* 205–214.

Braunstein, M. L., Hoffman, D. D., Shapiro, L. R., Andersen, G. J., & Bennett, B. M. (1987). Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 335–343.

Braunstein, M. L., Liter, J. C., & Tittle, J. S. (1994). Recovering 3-D shape from perspective translations and orthographic rotations. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 598–614.

Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 214–220.

Caelli, T. M., & Julesz, B. (1978). On perceptual analyzers underlying visual texture discrimination: Part 1. *Biological Cybernetics, 28,* 165–175.

Caudek, C., & Proffitt, D. R. (1993). Depth perception in motion parallax and stereokinesis. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 32–47.

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology, 7,* 20–43.

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

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

Cooper, L. A., & Shepard, R. N. (1975). Mental transformations in the identification of left and right hands. *Journal of Experimental Psychology: Human Perception and Performance, 104,* 48–54.

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

Cutting, J. E. (1983). Four assumptions about invariance in perception. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 310–317.

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

Cutting, J. E. (1987). On cross-ratios and motion perception: A reply to Niall. *Journal of Mathematical Psychology, 31,* 439–440.

Domini, F., Caudek, C., & Proffitt, D. R. (1997). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1111–1129.

Doner, J., Lappin, J. S., & Perfetto, G. (1984). Detection of three-dimensional structure in moving optical patterns. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 1–11.

Durgin, F. H., Proffitt, D. R., Olson, T. J., & Reinke, K. S. (1995). Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 679–699.

Edelman, S. (1995a). Class similarity and viewpoint invariance in the recognition of 3-D objects. *Biological Cybernetics, 72,* 207–220.

Edelman, S. (1995b). Representation, similarity, and the chorus of prototypes. *Minds and Machines, 5,* 45–68.

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of 3-D objects. *Vision Research, 32,* 2385–2400.

Ferraro, M., & Foster, D. H. (1984). Characterization of discrete and continuous modes of visual pattern discrimination. *Biological Cybernetics, 50,* 9–13.

Ferraro, M., & Foster, D. H. (1994). Elements of a fuzzy geometry for visual space. In A. Toet, D. H. Foster, H. Heijmans, & P. Meer (Eds.), *Shape in picture* (pp. 333–342). New York: Springer-Verlag.

Foster, D. H. (1975). An approach to the analysis of the underlying structure of visual space using a generalized notion of visual pattern recognition. *Biological Cybernetics, 17,* 77–79.

Foster, D. H. (1979). Discrete internal pattern representations and visual detection of small changes in pattern shape. *Perception & Psychophysics, 26,* 459–468.

Foster, D. H. (1980a). A description of discrete internal representation schemes for visual pattern discrimination. *Biological Cybernetics, 38,* 151–157.

Foster, D. H. (1980b). A spatial perturbation technique for the investigation of discrete internal representations of visual patterns. *Biological Cybernetics, 38,* 159–169.

Foster, D. H. (1983). Visual discrimination, categorical identification, and categorical rating in brief displays of curved lines: Implications for discrete encoding processes. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 785–806.

Garmendia, A. B., & Van Bockstaele, K. (1993). *Collinearity, parallelism and proximity as fuzzy regularities used in visual shape matching.* Unpublished master's thesis, University of Leuven, Leuven, Belgium.

Gauthier, I., & Tarr, M. J. (1997). Orientation-priming of novel shapes in the context of viewpoint-dependent recognition. *Perception, 26,* 51–73.

Gibson, B. S., & Peterson, M. A. (1994). Does orientation-

independent object recognition precede orientation-dependent recognition? Evidence from a cuing paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 299–316.

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

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

Green, M. (1992). Visual search: Detection, identification, and localization. *Perception, 21,* 765–777.

Hamm, J. P., & McMullen, P. A. (1998). Effects of orientation on the identification of rotated objects depend on the level of identity. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 413–426.

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.

Helmholtz, H., von (1962). *Treatise on physiological optics.* New York: Dover. (Originally published in German in 1857)

Huang, T. S., & Lee, C. H. (1989). Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-11,* 536–540.

Johansson, G. (1975). Visual motion perception. *Scientific American, 232* (6), 76–88.

Johansson, G., von Hofsten, C., & Jansson, G. (1980). Event perception. *Annual Review of Psychology, 31,* 27–63.

Johnston, J. C., & Pashler, H. (1990). Close binding of identity and location in visual feature perception. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 843–856.

Jolicoeur, P. (1990). Identification of disoriented objects: A dual-systems theory. *Mind and Language, 5,* 387–410.

Kandel, A. (1986). *Fuzzy mathematical techniques with applications.* Tallahassee, FL: Addison-Wesley.

Klein, F. (1939). *Elementary mathematics from an advanced standpoint. Vol. 2. Geometry.* New York: Dover. (Originally published in German in 1908)

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America, A8,* 377–385.

Koffka, K. (1935). *Principles of Gestalt psychology.* New York: Harcourt-Brace.

Kolers, P. A., & von Grünau, M. W. (1976). Shape and color in apparent motion. *Vision Research, 16,* 329–355.

Koriat, A., & Norman, J. (1988). Frames and images: Sequential effects in mental rotation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 93–111.

Kubovy, M., Holcombe, A. O., & Wagemans, J. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology, 35,* 71–98.

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science, 6,* 225–234.

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.

Lamdan, Y., Schwartz, J. T., & Wolfson, H. J. (1988). Object recognition by affine invariant matching. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 335–344). Washington, DC: Institute of Electrical & Electronics Engineers Computer Society Press.

Lappin, J. S., Ahlström, U. B., Craft, W. D., & Tschantz, S. T. (1995). Spatial primitives for seeing 3-D shape from motion. In T. V. Papathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early vision and beyond* (pp. 145–153). Cambridge, MA: MIT Press.

Lappin, J. S., Doner, J. F., & Kottas, B. L. (1980). Minimal conditions for the visual detection of structure and motion in three dimensions. *Science, 209,* 717–719.

Lappin, J. S., & Fuqua, M. A. (1983). Accurate visual measurement of three-dimensional moving patterns. *Science, 221,* 480–482.

Lappin, J. S., & Preble, L. D. (1975). A demonstration of shape constancy. *Perception & Psychophysics, 17,* 439–444.

Larsen, A., & Bundesen, C. (1978). Size scaling in visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 1–20.

Larsen, A., & Bundesen, C. (1998). Effects of spatial separation in visual pattern matching: Evidence on the role of mental translation. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 719–731.

Lawson, R. (1999). Achieving visual object constancy across plane rotation and depth rotation. *Acta Psychologica, 102,* 221–245.

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

Marcus, M., & Minc, H. (1988). *Introduction to linear algebra.* New York: Dover.

McMullen, P. A., Hamm, J., & Jolicoeur, P. (1995). Rotated object identification with and without orientation cues. *Canadian Journal of Experimental Psychology, 49,* 133–149.

Metzler, J., & Shepard, R. N. (1982). Transformational studies of the internal representation of three-dimensional objects. In R. N. Shepard & L. A. Cooper (Eds.), *Mental images and their transformations* (pp. 25–71). Cambridge, MA: MIT Press. (Originally published 1974)

Michaels, C. F., & Carello, C. (1981). *Direct perception.* Englewood Cliffs, NJ: Prentice-Hall.

Murray, J. E., Jolicoeur, P., McMullen, P. A., & Ingleton, M. (1993). Orientation-invariant transfer of training in the identification of rotated natural objects. *Memory & Cognition, 21,* 604–610.

Niall, K. K. (1987). Perspectives yet unseen. *Journal of Mathematical Psychology, 31,* 429–438.

Niall, K. K. (1992). Projective invariance and the kinetic depth effect. *Acta Psychologica, 81,* 127–168.

Niall, K. K. (1997). "Mental rotation," pictured rotation, and tandem rotation in depth. *Acta Psychologica, 95,* 31–83.

Niall, K. K., & Macnamara, J. (1989). Projective invariance and visual shape constancy. *Acta Psychologica, 72,* 65–79.

Niall, K. K., & Macnamara, J. (1990). Projective invariance and picture perception. *Perception, 19,* 637–660.

Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 173–186.

Pani, J. R. (1993). Limits on the comprehension of rotational motion: Mental imagery of rotations with oblique components. *Perception, 22,* 785–808.

Pani, J. R., William, C. T., & Shippey, G. T. (1995). Determinants of the perception of rotational motion: Orientation of the motion to the object and to the environment. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1441–1456.

Parsons, L. M. (1987). Visual discrimination of abstract mirror-

reflected 3-D objects at many orientations. *Perception & Psychophysics, 42,* 49–59.

Parsons, L. M. (1995). Inability to reason about an object's orientation using an axis and angle of rotation. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1259–1277.

Perkins, D. N. (1972). Visual discrimination between rectangular and nonrectangular parallelopipeds. *Perception & Psychophysics, 12,* 396–400.

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

Pizlo, Z., & Rosenfeld, A. (1992). Recognition of planar shapes from perspective images using contour-based invariants. *Computer Vision, Graphics, and Image Processing: Image Understanding, 56,* 330–350.

Pizlo, Z., Rosenfeld, A., & Weiss, I. (1997). The geometry of visual space: About the incompatibility between science and mathematics. *Computer Vision and Image Understanding, 65,* 425–433.

Pizlo, Z., & Salach-Golyska, M. (1995). 3-D shape perception. *Perception & Psychophysics, 57,* 692–714.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 343,* 263–266.

Proffitt, D. R., Rock, I., Hecht, H., & Schubert, J. (1992). Stereokinetic effect and its relation to the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 3–21.

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

Sagi, D., & Julesz, B. (1985). "What" and "where" in vision. *Science, 228,* 1217–1219.

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

Simpson, W. A. (1986). The cross-ratio and the perception of motion and structure. In N. I. Badler & J. K. Tsotsos (Eds.), *Motion: Representation and perception* (pp. 248–252). New York: North-Holland.

Stankiewicz, B. J., Hummel, J. E., & Cooper, E. E. (1998). The role of attention in priming for left-right reflections of object images: Evidence for a dual representation of object shape. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 732–744.

Stavrianos, B. K. (1945). The relation of shape perception to explicit judgments of inclination. *Archives of Psychology, 296,* 1–94.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science: Vol. 4: Methods, models, and conceptual issues* (2nd ed., pp. 703–863). Cambridge, MA: Bradford Books/MIT Press.

Stevens, K. A. (1983). Surface tilt (the direction of slant): A neglected psychophysical variable. *Perception & Psychophysics, 33,* 241–250.

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 & 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? Comment on Biederman and Gerhardstein (1993). *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1494–1505.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science, 8,* 282–289.

Thouless, R. H. (1931). Phenomenal regression to the real object II. *British Journal of Psychology, 22,* 1–30.

Thouless, R. H. (1934). The general principle underlying effects attributed to the so-called phenomenal constancy tendency. *Psychologische Forschung, 19,* 300–310.

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 663–678.

Todd, J. T. (1982). Visual information about rigid and nonrigid motion: A geometric analysis. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 238–252.

Todd, J. T. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. *Perception & Psychophysics, 36,* 97–103.

Todd, J. T. (1985). The perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception and Performance, 11,* 689–710.

Todd, J. T., Akerstrom, R. A., Reichel, F. D., & Hayes, W. (1988). Apparent rotation in three-dimensional space: Effects of temporal, spatial, and structural factors. *Perception & Psychophysics, 43,* 179–188.

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 48,* 419–430.

Tsal, Y., & Lavie, N. (1993). Location dominance in attending to color and shape. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 131–139.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: MIT Press.

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

Van Campenhout, S., Wagemans, J., Kyllingsbaek, S., Bundesen, C., & Larsen, A. (1998, August). *Rotating simple two-dimensional figures through space along the shortest path.* Poster session presented at the 21st European Conference on Visual Perception, Oxford, England. Abstract published in *Perception, 27* (Suppl.), 125a.

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

Vanrie, J., Béatse, E., Wagemans, J., Sunaert, S., & Van Hecke, P. (1999, August). *Mental rotation versus invariant features in object perception from different viewpoints: An fMRI study.* Paper presented at the 22nd European Conference on Visual Perception, Trieste, Italy. Abstract published in *Perception, 28* (Suppl.), 10b.

Wagemans, J. (1992). Perceptual use of nonaccidental properties. *Canadian Journal of Psychology, 46,* 236–279.

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., De Troy, A., Van Gool, L., Wood, J. R., & Foster, D. H. (1994). Affine shape equivalence. In F. Dillen, & L. Verstraelen (Eds.), *Geometry and topology of submanifolds VI* (pp. 307–314). Singapore: World Scientific.

Wagemans, J., Lamote, C., & Van Gool, L. (1997). Shape equivalence under perspective and projective transformations. *Psychonomic Bulletin & Review, 4,* 248–253.

Wagemans, J., & Tibau, S. (1999). Visual measurement of relative distances between three collinear dots rotating in a slanted plane. *Perception, 28,* 267–282.

Wagemans, J., Vanden Bossche, P., Segers, N., & d'Ydewalle, G.

(1994). An affine group model and the perception of orthographically projected planar random polygons. *Journal of Mathematical Psychology, 38,* 59–72.

Wagemans, J., Van Gool, L., & Lamote, C. (1996). The visual system's measurement of invariants need not itself be invariant. *Psychological Science, 7,* 232–236.

Wallis, G., & Bülthoff, H. (1999). Learning to recognize objects. *Trends in Cognitive Sciences, 3,* 22–31.

Willems, B., & Wagemans, J. (1998, August). *Viewpoint-dependent recognition of 3-D objects need not imply mental rotation in space.* Poster session presented at the 21st European Conference on Visual Perception, Oxford, England. Abstract published in *Perception, 27* (Suppl.), 126a.

Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 397–412.

Appendix A

Parameterization of the Affine Transformation

A 2-D affine transformation can be interpreted as describing the effect on a (pseudo) orthographically projected planar image of a planar shape oriented arbitrarily in depth as the position and orientation of that planar shape are altered. An orthographic or parallel projection is one with the center of projection at infinity or, in other words, without perspective distortion; a pseudo-orthographic projection additionally takes account of the influence of the distance between the object and the image (see Van Gool et al., 1994, for a more mathematical treatment).

In terms of this projection, a general affine transformation is generated from 3-D rotation and 3-D translation, followed by pseudo-orthographic projection. By combining these basic transformations and choosing appropriate parameters, one can obtain every possible affine transformation with a clear 3-D interpretation. The image projection of the 3-D rotation can be decomposed further as a rotation in the image plane,

$$R_\phi = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

or rotation about the optical axis; a rotation in depth or slant about the horizontal axis,

$$R_\sigma = \begin{pmatrix} 1 & 0 \\ 0 & \cos\sigma \end{pmatrix};$$

and another rotation or tilt about the optical axis,

$$R_\tau = \begin{pmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{pmatrix}.$$

Likewise, the translation has two components, translation in the image plane,

$$T = \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

and translation in depth, corresponding to a scaling of the image,

$$S = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}.$$

The combination of these components yields, in vector notation,

$$S\,(R_\tau \quad R_\sigma \quad R_\phi \overline{X} + T),$$

and, in matrix notation,

$$\begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}$$

$$\left[\begin{pmatrix} \cos\tau & -\sin\tau \\ \sin\tau & \cos\tau \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \cos\sigma \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \right].$$

This is the parameterization of the general affine transformation used in this study.

In the experiments, the *translation component* was omitted, and the two patterns being compared were always presented side by side on the screen. Translation in depth was omitted because its reformulation as a scaling is somewhat arbitrary, and it involves irrelevant issues of visual resolution. Neither translation nor scaling constitutes a typically affine deformation. Hence, the two parameters (t_x, t_y) were set to zero and the third (k) to unity. The remaining three parameters, the *rotation components*, had the most salient effect on the appearance of the projected shape and were the ones that were varied in the experiments. In Experiment 1, only two of them were varied: All patterns were first rotated about the optical axis and then slanted about the horizontal axis. In Experiments 2 and 3, a tilt transformation was also included.

The three angles, ϕ, σ, and τ, uniquely describe the 3-D orientation of the transformed pattern, except when the slant vanishes. In that case, the whole operation amounts to a single rotation about the optical axis with angle $\phi + \tau$. Referring to Figure 3, it is clear that with $\sigma = 0°$, τ is no longer defined as it is supposed to specify the orientation of the intersection of the object plane (containing the shape) and the image plane. Without slant, these planes coincide, and there is no unique intersection line. In the formulas, we can still plug in τ with $\sigma = 0$; tilt then works as a simple rotation in the image plane. So, with our implementation, we can manipulate the parameters continuously.

Appendix B

Mental Rotation Along the Shortest Path in 3-D Space: Computational Formulas and Reanalysis

Computational Formulas

As outlined in Appendix A, the affine transformation used in the experiments here was restricted to the rotation components, consisting of rotation, slant, and tilt. In matrix notation,

$$A = \begin{pmatrix} \cos \tau & -\sin \tau \\ \sin \tau & \cos \tau \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \cos \sigma \end{pmatrix} \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix}. \quad (B1)$$

This matrix A is the result of an orthographic projection of the following 3-D rotation:

$$T = \begin{pmatrix} \cos \tau & -\sin \tau & 0 \\ \sin \tau & \cos \tau & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \sigma & -\sin \sigma \\ 0 & \sin \sigma & \cos \sigma \end{pmatrix}$$

$$\begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (B2)$$

On the basis of this matrix, one can compute the shortest path in 3-D space through which the untransformed pattern must be moved to obtain the transformed pattern (or vice versa). In other words, it is possible to derive the unique 3-D angle and orientation of the rotation axis in space that yields the shortest path.

Any composition of orthogonal matrices yields another orthogonal matrix, which amounts to a pure rotation in this case as the determinant is +1. Here, we specify the relation between the parameters σ, τ, and ϕ, and the angle and axis of this overall rotation. Consider T as an orthogonal matrix with determinant equal to +1. Then the trace of T equals $1 + 2 \cos \theta$ (i.e., the sum of the characteristic roots of T), and there is a unique angle θ for which $0 \le \theta \le \pi$ (for a mathematical proof, see Marcus & Minc, 1988). Thus, $\cos \theta = (\text{trace } T - 1)/2$, or

$$\cos \theta = |(\cos \sigma + 1) [\cos (\tau + \phi) + 1]/2| - 1. \quad (B3)$$

One can then obtain the 3-D rotation angle θ from the separate affine parameters, rotation ϕ, slant σ, and tilt τ.

The orientation of the rotation axis is found by computing one of the characteristic vectors associated with its respective characteristic root. By determining this vector for matrix T and its root 1, one obtains the following coordinates of the rotation axis:

$$n = \begin{pmatrix} \dfrac{(\cos (\tau + \theta))}{2} \\[2ex] \dfrac{(\sin (\tau + \theta))}{2} \\[2ex] \cot \dfrac{\sigma}{2} \times \dfrac{(\sin (\tau - \theta))}{2} \end{pmatrix}. \quad (B4)$$

Without the tilt component, B3 is reduced to

$$\cos \theta = |(\cos \sigma + 1) [\cos \phi + 1]/2| - 1, \quad (B5)$$

and B4 is reduced to

$$n = \begin{pmatrix} \dfrac{(\cos \theta)}{2} \\[2ex] \dfrac{(\sin \theta)}{2} \\[2ex] \cot \dfrac{\sigma}{2} \times \dfrac{(\sin (- \theta))}{2} \end{pmatrix}. \quad (B6)$$

Results for Experiment 1c

Analyses were restricted to RTs for correct *same* responses. Extreme outliers (≥ 3 *SD* per participant) were removed to maximize the chance of finding statistically reliable effects (i.e., 5% of all correct *same* responses). Using equation B5, 40 different 3-D angles were derived. This variable had a highly significant effect on RT, $F(39, 411) = 3.18$, $p < .0001$, but its linear component was only marginally significant, $F(1, 1715) = 3.68$, $p < .10$ (see Figure B1). The effect of 3-D angle was also examined separately for all participants, all slant levels, and all axis orientations (found from equation B6). In summary, these analyses show that the effect of 3-D angle varied somewhat across these data sets; for example, it yielded a significant effect for only half the participants and for all slant levels (except for 30°), although it was variable across slant levels (see Figure B2). In fact, the only trend in these effects pointed to the significant nonlinearities in the performance (in contrast to, e.g., Parsons, 1987; Shepard & Cooper, 1982).
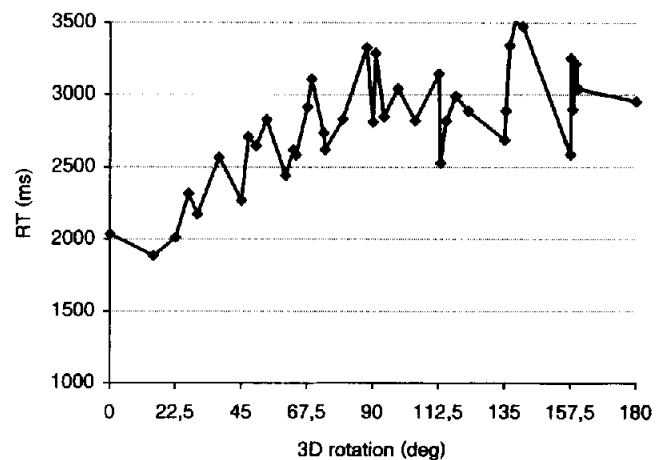

Figure B1. The effect of 3-D rotation angle on response times (RTs in ms) for correct *same* trials in Experiment 1c. Commas in numbers along *x*-axis represent decimal points.
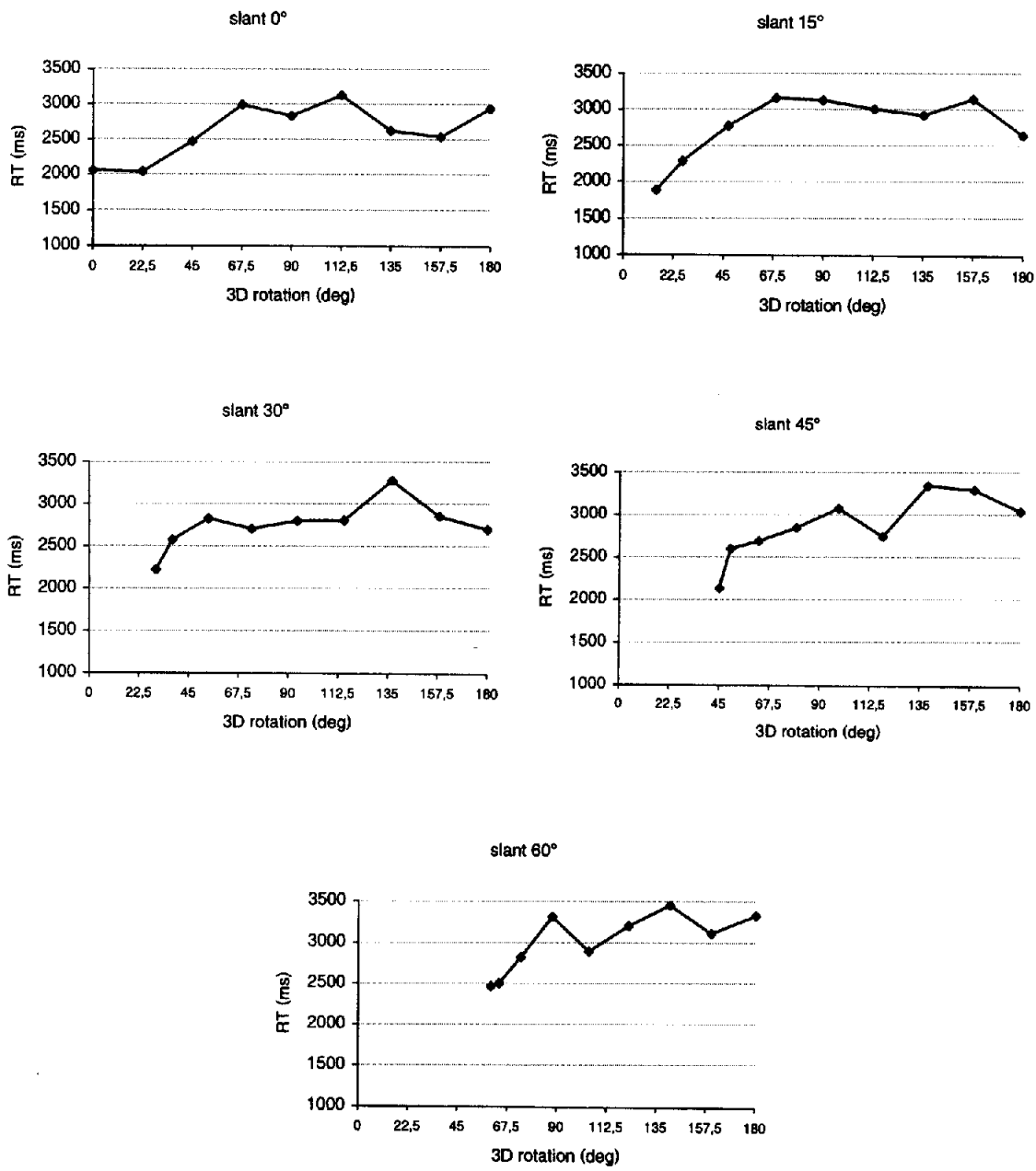
Figure B2. The effect of 3-D rotation angle on response times (RTs in ms) for correct *same* trials in Experiment 1c, at separate levels of slant. Commas in numbers along *x*-axis represent decimal points.
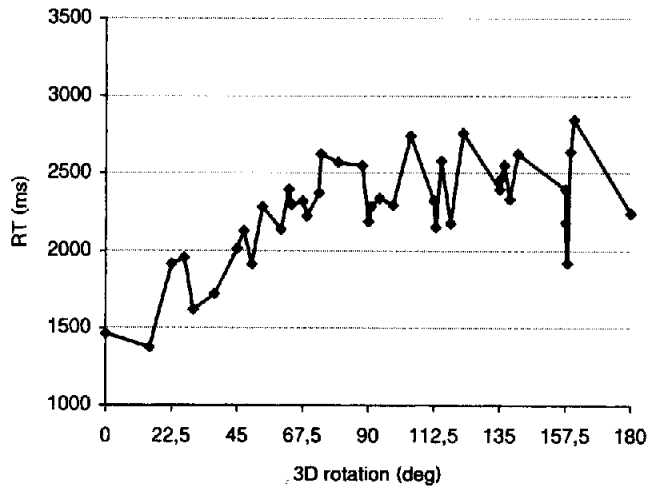

Figure B3. The effect of 3-D rotation angle on response times (RTs in ms) for correct *same* trials in Experiment 2a. Commas in numbers along *x*-axis represent decimal points.

slant 0°

slant 15°
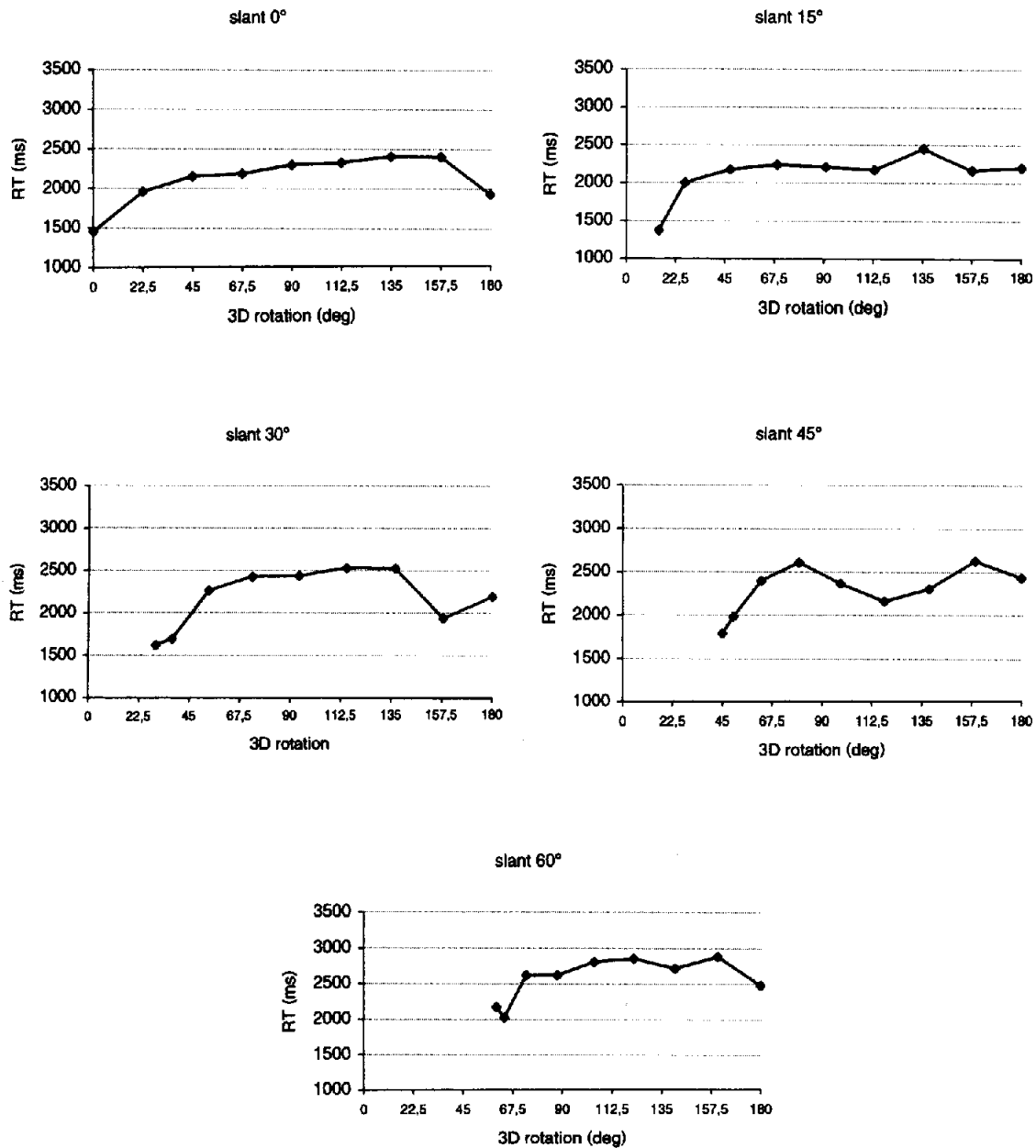

slant 30°

slant 45°

slant 60°

Figure B4. The effect of 3-D rotation angle on response times (RTs in ms) for correct *same* trials in Experiment 2a, at separate levels of slant. Commas in numbers along *x*-axis represent decimal points.

Results for Experiment 2a

The same analyses were used here. The addition of tilt in this experiment yielded only one extra 3-D angle (equation B1). To facilitate comparison with Experiment 1c, this condition (constituting 0.6% of the data) was eliminated from the data. As in Experiment 1c, 3-D angle had a highly significant effect on RT, $F(39, 411) = 3.58, p < .0001$, but now its linear component was also very reliable, $F(1, 1715) = 25.52, p < .0001$ (see Figure B3). Linear regressions for individual participants' data yielded reliable but small effects for 6 out of 10 observers. As in Experiment 1c, the

effect of 3-D angle was statistically reliable with significant nonlinearities at all levels of slant (see Figure B4). When the combined tilt-minus-rotation was incorporated as a variable, an ANOVA on correct *same* responses yielded a statistically significant effect only for slant, not for tilt minus rotation.