Pergamon

0042-6989(93)E0029-7

Vision Res. Vol. 34, No. 12, pp. 1637–1658, 1994

0042-6989/94 $7.00 + 0.00

A Theory of Shape Constancy Based on Perspective Invariants

ZYGMUNT PIZLO*

Received 2 April 1993; in revised form 13 October 1993

Shape constancy refers to the phenomenon in which the percept of the shape of a given object remains constant despite changes in the shape of the object's retinal image. The phenomenon of shape constancy is considered from historical, theoretical and empirical perspectives in this paper. First, four prior theories are discussed; specifically, (1) Helmholtzian theory, which assumes that shape constancy is achieved by taking an object's orientation into account, (2) Gestalt theory, which assumes that shape constancy involves a relationship between the perceived shape and perceived orientation of an object, (3) Gibsonian theory, which assumes that shape constancy is based on projective invariants and (4) multiple view theory, which assumes that shape constancy is achieved by memorizing a large set of different views of the object. It is shown, by an analysis of the prior literature, that none of these theories can actually explain the phenomenon of shape constancy. A new theory, which is based on new perspective invariants of a flat shape, is then proposed. The new Perspective Invariants Theory can account for all prior shape constancy experiments. New experiments, testing predictions of the Perspective Invariants Theory are then described. These experiments showed that: (1) a novel shape can be matched with its single perspective image in the absence of depth cues, (2) perceptual processing of shape is impaired when the range of possible values of tilt is wide, (3) perceptual processing of shape is not affected by the width of the range of possible values of slant. These results support predictions of Perspective Invariants Theory.

Shape constancy Shape perception Perspective invariants

INTRODUCTION

Shape constancy refers to the fact that the percept of the shape of a given object remains constant despite changes in the shape of the object's retinal image. Retinal image shape may change when the orientation of the object relative to the observer changes. Shape is defined conventionally in this paper as the property of the contour of a figure or of the surface of an object that does not change under translation, rotation or changing size. Thus, shape does not change if angles or ratios of distances do not change. This definition seems to be intuitively obvious and is commonly used in research on shape constancy.†

Shape constancy is important because shapes are characteristic properties of objects that can be used to recognize them. Consider the shape of the surface of a table. The shape of the retinal image of the table depends on the viewing direction and viewing distance. It can be a rectangle, a parallelogram, a trapezoid or just an irregular quadrilateral. Yet, in everyday life, shape constancy is observed because our percept of the table's shape is not affected by viewing direction or distance. Several theories of shape constancy have been proposed. None can explain all prior experimental results. Moreover, results of these prior experiments are inconsistent. This paper begins by reviewing prior research and theories of shape constancy. This review led to the formulation of a new theory in which shape constancy is based on perspective invariants. This theory is called, henceforth, the "Perspective Invariants Theory". Perspective invariants are new types of geometrical properties that are invariant only under a perspective transformation and not under an arbitrary projective or affine transformation (Pizlo & Rosenfeld, 1992). The Perspective Invariants Theory is the first to account for all prior experimental results on shape constancy. This theory was then tested in new experiments, first reported in this paper. These new experiments also supported the Perspective Invariants Theory.

Prior Theories

There have been four major theories of shape constancy. According to the first theory, developed in the Helmholtzian tradition, the perceived shape of a figure is reconstructed (computed) from the retinal image after taking into account the figure's orientation relative to

*Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907, U.S.A. [*Email* pizlo @ psych. purdue. edu].

†Other definitions are, and have been, used (Zusne, 1970; Toussaint, 1988). The definition used in the present paper may not be ideal for applications other than shape constancy.

the observer. This theory has received longstanding and widespread attention in the vision community and recently has been formulated in considerable detail (e.g. Marr, 1982). These recent formulations were influenced by some of the early ideas of Gibson (1950, 1979). Gibson pointed out that the visual image of the natural environment (specifically, gradients of texture, shading, hue, binocular disparity and motion calculated from the retinal image) is a rich source of information about objects out there. The Helmholtzian theorists elaborated this idea by recognizing that these gradients are efficient cues to the orientation of the object relative to the observer (i.e. cues about depth) and, furthermore, that the orientation can be taken into account when reconstructing the shape. It has to be pointed out, however, that Gibson himself did not claim that orientation is taken into account in reconstructing shape. Examples of Helmholtzian approaches include Witkin (1981), Kanade and Kender (1983), Ikeuchi (1984), and Aloimonos (1988) all of whom provide mathematical solutions of the problem of reconstructing shape from texture; Ikeuchi and Horn (1981), Pentland (1986) and Lehky and Sejnowski (1990), who discuss reconstructing shape from shading; Johansson (1977), Longuet-Higgins and Prazdny (1980), Ullman (1983), Sperling, Landy, Dosher and Perkins (1989) and Dosher, Landy and Sperling (1989), who analyze reconstructing shape from motion; Ullman (1979), Longuet-Higgins (1981) and Koenderink and van Doorn (1991), who discuss reconstrucing shape from multiple views. Note that in the Helmholtzian theory, the perception of shape reflects the operation of two variables, shape and depth (where by "depth" I mean the orientation of the shape relative to the observer). Hence, this theory predicts that in the absence of depth cues, shape constancy must fail.

According to the second theory, developed by the Gestalt psychologists (Koffka, 1935, pp. 211–264), the perceived shape of a figure and the perceived orientation of the figure relative to the observer are related to each other. Given a fixed retinal image of a shape, if the perceived orientation changes, perceived shape must change accordingly, and vice versa. For example, a retinal image having the shape of an ellipse can be perceived either as a circle slanted in depth or as an ellipse in the frontoparallel plane (slant is the angle formed by a planar shape and a subject's frontoparallel plane). Whether the subject perceives a slanted circle or a non-slanted ellipse depends on the context which is shared by all shapes in the visual field. Assume, for example, that there are two shapes in the retinal image, an ellipse and a trapezoid, produced by a circle and a square slanted by the same angle. If the trapezoid is perceived as a slanted square, the ellipse must be perceived as a slanted circle. If, on the other hand the trapezoid is perceived as an upright trapezoid, the ellipse must be perceived as an upright ellipse (see Koffka, 1935, p. 223). In other words, the percept produced by the trapezoid (or by the ellipse) on the retina gives rise to the percept of its orientation which constrains the percept produced by the retinal image of the other shape.

How would this perceived shape-perceived orientation relationship, which holds for a fixed retinal image, account for shape constancy? According to this relationship, if orientation is perceived accurately, shape is perceived accurately as well (and vice versa). Therefore, if the observer is presented with the same object (and its context) under different orientations and, hence, with different retinal images, and the observer perceives the orientation accurately, shape constancy will be achieved. Note that in Gestalt theory, shape constancy is assumed to require a context which provides information about the orientation of the object. If such a context is not provided, or if it leads to inaccurate perception of orientation, shape constancy must fail. This makes Gestalt theory similar to Helmholtzian; in both, cues for orientation are important. The difference is that Gestalt theory uses the relation between the perceived shape and perceived orientation, whereas Helmholtzian theory uses the relation between the retinal shape and physical orientation with the orientation assumed to be taken into account unconsciously.

The third theory of shape constancy, which will be called here "Gibsonian", is derived from suggestions of Cassirer (1944) and Gibson (1950). It assumes that perceived shape is based on only one variable, shape, and more specifically, on the projective invariants of shape (see Klein, 1939, for a treatment of projective invariants, and Cutting, 1986, for a review of Gibsonian theory). Projective invariants of shape are geometrical properties that are preserved in any perspective image of the shape (perspective is a geometrical transformation that describes the relation between an object in the physical world and the object's retinal image). Figure 1 shows an example of a projective invariant called a "cross ratio of four collinear points". Let O′ be the first viewpoint of an observer and O″ be the second viewpoint of the observer, Let L′ and L″ represent the observer's retina in two spatial positions corresponding to the two different viewpoints. Let L represent an object out there. Given

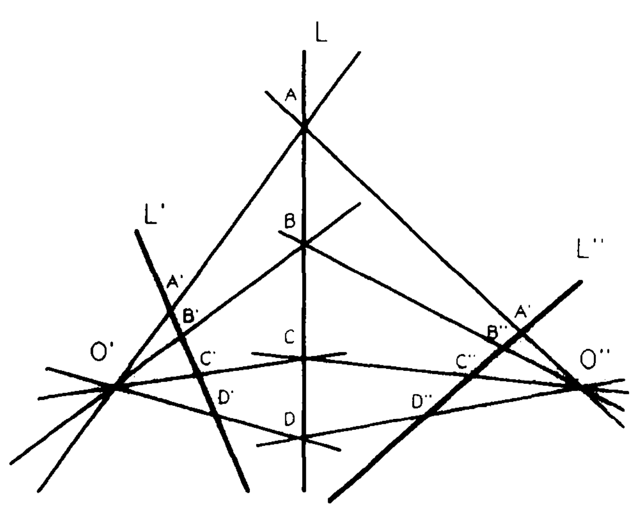

FIGURE 1. Perspective relationships between L and L′ and between L and L″. Cross ratio of four collinear points, which is defined in the text, is a perspective (and projective) invariant.

an image of 4 collinear points one can calculate their cross ratio, defined as: $[(A-B)(C-D)]/[(A-D)(C-B)]$ and this cross ratio has the same value in any perspective image of these points. This example illustrates the idea of how projective invariants can be used to achieve shape constancy.

There have been a number of mathematical solutions of the shape constancy problem formulated within the framework of the Gibsonian theory. Some of the solutions were based on projective invariants. These solutions apply to the perspective transformation (Duda & Hart, 1973; Cutting, 1986; Weiss, 1988). Other solutions were based on affine invariants. These solutions apply only to the affine transformation, which is an approximation to perspective (Hoffman, 1966; Lamdan, Schwartz & Wolfson, 1988; Koenderink & van Doorn, 1991).*

According to the fourth theory of shape constancy, called the multiple view theory, shape constancy is based on the storage of a large set of 2D views of a given object with the object's shape recognized from a novel view by checking whether this novel 2D view can be matched with one of the stored views. If this novel view is not identical with (or very similar to) any of the stored views, shape constancy must fail. In this theory, shape constancy can be achieved when the observer stores about 100 views of the shape (Edelman & Bulthoff, 1992; Poggio & Edelman, 1990).

It has to be pointed out that in these two latter theories (Gibsonian and multiple view), unlike in Helmholtzian and Gestalt theories, shape constancy does not involve depth cues. Therefore, both Gibsonian and multiple view theories predict that depth cues will have little or no effect on shape constancy.

Prior Experiments

The role of depth cues

Thouless conducted one of the most often cited studies on shape constancy (Thouless, 1931b). He tested the effect of the absence of depth cues on the accuracy of judgments about the shapes of ellipses having different orientations relative to the subject. The stimuli he used, ellipses, belong to a special family of shapes that make them inappropriate for research on shape constancy, namely, all ellipses are equivalent under perspective transformation. Specifically, the retinal image of an ellipse is always an ellipse and any ellipse can be a retinal image of any other ellipse.† Thus, in Thouless's experiment each retinal shape could have been produced by any of the stimuli so retinal shape itself was not sufficient to discriminate among the stimuli. This is a very special case, which may lead to what I call "shape ambiguity". The shape ambiguity problem is complementary to, but different from, the shape constancy problem. Shape ambiguity refers to the failure to discriminate the shapes of objects, which when seen from different orientations relative to the subject, give rise to identically-shaped retinal images. For this reason Thouless's experiment is not relevant to shape constancy, only to shape ambiguity. Note that shape ambiguity, unlike shape constancy, is probably very rare in everyday life because the shapes of most objects are quite different from each other and it is rather unlikely that two (or more) of them will give rise to identically-shaped retinal images. The distinction between shape constancy and shape ambiguity seems to have been overlooked in the past and, as a result, shape ambiguity, despite its relative unimportance, has figured prominently in laboratory studies of shape perception where individuals mistakenly thought that they were studying shape constancy.‡

*The affine transformation is an accurate approximation to perspective when the extent of the figure in depth is small relative to the distance of the figure from an observer.

†The proof is as follows: It is well known that a perspective image of a conic section (i.e. of a circle, ellipse, parabola or hyperbola) is also a conic section (Klein, 1939). If one assumes that the entire object is visible, as was the case in Thouless's experiment, a retinal image of a closed curve is a closed curve. Note further that under the assumption that the entire object is visible, an ellipse is the only closed conic section (a circle is simply a special case of an ellipse). Therefore, the perspective image of an ellipse must be an ellipse. Moreover, changing the orientation of an ellipse relative to the subject continuously changes the ratio of the axes of the ellipse's image. The ratio can have an arbitrary value between zero and infinity. This completes the proof because the shape of an ellipse is uniquely defined by the ratio of its axes. It is worth pointing out that Thouless (1931a), in his preliminary experiment, also used quadrilaterals that are adequate stimuli for testing shape constancy. However, in this experiment, depth cues were always available and, therefore, this experiment cannot be used to study the role of depth cues in shape constancy. It must be pointed out that these, and all other claims about the shapes of retinal images, presented in this paper, assume that the human retina is planar, whereas in fact it is spherical. This assumption seems to be justified for the following reason. The vast majority of experimental evidence and theories, related to shape constancy, considered images constrained to the central part of the retina (i.e. relatively small objects viewed with unrestricted eye fixation), where eccentricities are less than about 20 deg. In this part of the retina, the retina can be well approximated by a plane and, therefore, the planarity assumption is satisfied. Note that studying shape constancy only in the central part of the retina is reasonable because visual acuity is relatively high only in this part. As a result, shape perception in the periphery is quite poor (although possible, Menzer & Thurmond, 1970) and, shape constancy is difficult to achieve in the periphery even when the shape of the retinal image does not change.

‡Thouless (1934) seemed to be aware of this problem. He pointed out that perceptual constancy is demonstrated only in experiments in which the object's property (size, color, shape) is constant, its retinal image changes and the percept of the object's property is constant. All experiments in which the object changes its properties, its retinal image is constant and the percept changes are not relevant to the phenomenon of perceptual constancy. Instead, they are relevant to what Thouless called "perceptual inconstancy", and to what I have called the problem of perceptual ambiguity. However, Thouless did not realize that this problem also applied to his own experiments on shape because he did not understand the rules of perspective transformation. He thought, erroneously, that the retinal image of an ellipse is not exactly an ellipse (Thouless, 1931a). If this claim were true (and it is not), then his experiment might have been relevant to shape constancy problem. In fact, his experiment was relevant only to the problem of shape ambiguity. It follows from the definition of shape ambiguity that information about the orientation of the object relative to the subject is

—*continued overleaf*

Many subsequent experiments on shape constancy contained the same methodological flaw as Thouless's, testing shape ambiguity, rather than shape constancy. These studies used as stimuli either ellipses (Leibowitz & Bourne, 1956, Meneghini & Leibowitz, 1967; Leibowitz, Wilcox & Post, 1978), or triangles, which, like ellipses, are all equivalent under perspective transformation (Gottheil & Bitterman, 1951; Beck & Gibson, 1955; Epstein, Bontrager & Park, 1962; Wallach & Moore, 1962), or trapezoids chosen in such a way that they were perspectively equivalent (Beck & Gibson, 1955; Kaiser, 1967). All these experiments confirmed and generalized Thouless's result, showing that cues for depth are important for solving the shape ambiguity problem not only for ellipses that are smooth, but also for triangles and trapezoids that have distinctive points. However, these authors, like Thouless, thought, erroneously, that their results were relevant to the phenomenon of shape constancy. They are not.

Stavrianos (1945), unlike all others, actually tested the effect of the absence of depth cues on shape constancy, rather than on shape ambiguity, in her first experiment. She used rectangles as stimuli whose sizes were about 8×15 cm, the distance of the subject from the stimulus was 90 cm and the slant was in the range from 15 to 55 deg. Thus, in her experiment, the extent of the rectangle in depth was not very small as compared to its distance from the observer. Under such conditions, retinal shapes were perspective, not affine, transformations of the rectangles (see footnote on p. 1639). It is known that different rectangles are not equivalent under perspective transformation (although under affine transformation they are equivalent). In other words, different rectangles give rise to differently-shaped retinal images. As a result, Stavrianos's experiment allowed shape constancy to be tested, unconfounded with shape ambiguity. Two standard rectangles were used. The rectangles were different from one another with respect to the ratio of sides (aspect ratio). In the experiment, the standard rectangles were shown in a random order and the subject was asked to make judgments about both slant and shape. For the slant judgments, the subject had to adjust the slant of a comparison rectangle, whose shape was constant, to the slant of a standard rectangle. For the shape judgments, the subject had to adjust the shape (i.e. aspect ratio) of a comparison rectangle, seen in the frontoparallel plane, to the shape of a standard rectangle, slanted in depth. Stavrianos showed that judgments about slant and shape were relatively accurate and precise when depth cues were present. When depth cues were absent, the accuracy and precision of slant judgments were impaired. However, the accuracy and precision of shape judgments were not affected by the absence of depth cues.

Thus, Stavrianos showed that shape constancy did not involve depth cues. Yet her results were not interpreted correctly in the past, even by Stavrianos, herself (Hochberg's review, 1971, seems to be the only exception). Thouless's results were commonly considered to have established the fact that shape constancy requires depth cues and Stavrianos's results appeared to contradict established knowledge. But, there was actually no contradiction because Thouless (and others) had studied shape ambiguity, showing that information about depth is important, and Stavrianos had studied shape constancy, showing that information about depth is unimportant.

It is clear that Stavrianos's results are not consistent with the Helmholtzian theory of shape constancy. Namely, eliminating cues for depth, which impairs slant judgments, should also impair shape judgments because these judgments are assumed to take slant (depth) into account. Shape judgments were not impaired, however. One potential, but invalid, argument that can be used to revive the Helmholtzian theory is based on the concept of "unconscious inference": it is possible that the slant judgments were poor because slant information was not available for conscious judgment. This argument would imply that slant judgments would be equally poor across conditions. However, these judgments were more precise and accurate in the presence of depth cues than in their absence.

Stavrianos's results were also inconsistent with Gestalt theory. According to this theory, the perceived shape and perceived orientation are correlated, so errors in perceived slant would be associated by errors in perceived shape. However, no systematic relationship between slant and shape judgments was observed in Stavrianos's Experiment I: reducing depth cues impaired slant judgments but did not affect shape judgments.

Stavrianos's results, showing that shape constancy is not based on depth cues, are consistent with two groups of more recent experiments. First, Shepard and colleagues (Shepard & Metzler, 1971; Shepard & Cooper, 1982) showed that subjects could reliably match two views of a polyhedron in the absence of depth cues. Second, Rock and DiVita (1987), Rock, Wheeler and Tudor (1989) and Edelman and Bulthoff (1992) showed that subjects could not reliably recognize a 3D wire shape from a novel view even when depth cues were available (binocular disparity, accommodation, vergence and shading). These results clearly imply that depth cues are irrelevant to shape constancy because for some shapes (flat shapes and polyhdera) constancy can be reliably achieved without depth cues, while for others (3D wire shapes, or other nonsense shapes with no flat

footnote continued

necessary to disambiguate the shapes of objects. The fact that the human subject is capable of using orientation to solve the shape ambiguity problem was demonstrated empirically by Thouless. He found that in the presence of cues for depth, the subject perceived the shapes of ellipses more accurately than in the absence of cues for depth. Without cues for depth the subject could not discriminate among the stimuli at all and the subject's judgments were based entirely on the shape of the retinal image. In fact, perceived shape agreed with the shape of the retinal image. These results could seem to provide support for Helmholtzian and Gestalt theories of shape constancy. But since Thouless actually tested shape ambiguity rather than shape constancy, his results do not show that depth cues have an important role in shape constancy, nor do they bear on any theory of shape constancy.

surfaces) constancy cannot be reliably achieved even when depth cues are available.

The idea that depth cues are irrelevant to shape constancy has received further support from studies showing that judgments about depth are neither accurate nor precise (see McKee, Levi & Bowne, 1990, for results on the poor precision of depth judgments from stereopsis; Johnston, 1991, on the poor accuracy of depth judgments from stereopsis; Braunstein, 1976, and Perrone, 1980, 1982, on the poor accuracy of slant judgments from texture; Todd & Akerstrom, 1987, and Todd & Reichel, 1989, on the poor precision and accuracy of depth judgments from texture and shading; Todd & Bressan, 1990, Todd & Norman, 1991, on the poor precision and accuracy of depth judgments from motion). All of these studies imply that it is not likely that depth is a reliable source of information that could be used for achieving shape constancy.

Let me summarize this review of the effect of depth cues on shape constancy. The results of many experiments show unequivocally that shape constancy does not involve depth cues. However, there seem to be two different cases. Shape constancy in the absence of depth cues can be achieved for flat shapes or solid shapes that have flat faces, but shape constancy even in the presence of depth cues is hard to achieve for the rare solid shapes that do not have any flat or nearly flat faces. These results can be explained in one of two ways.

First, shape constancy can be based on projective or affine invariants calculated from a single image (Gibsonian theory). It is known that general case non-trivial invariants exist for the transformation between a flat shape and its (flat) image but they do not exist for the transformation between an arbitrary set of points in space and its single image (Burns, Weiss & Riseman, 1990). This fact could explain shape constancy in the case of flat shapes. Note that this could also explain constancy for solid shapes that have flat faces (e.g. polyhedra) because the constancy for such shapes can be based on invariants calculated for individual faces (see also a recent paper by Rothwell, Forsyth, Zisserman & Mundy, 1993, where projective invariants for vertices of polyhedra were formulated). Finally, this fact can explain why it is difficult to achieve constancy in the case of shapes that resemble arbitrary set of points in 3D space, like the wire stimuli in Edelman and Bulthoff's (1992) experiments.*

Second, shape constancy could be explained by the multiple view theory. Namely, one could argue that shape constancy was achieved for rectangles or polyhedra having rectangular faces because human observers are quite familiar with rectangles. Moreover, subjects in Stavrianos's and in Shepard and colleagues' experiments were exposed, in the course of testing, to many views of the standard stimuli and it could have been this experience, rather than invariants, that led to shape constancy. In fact, extensive experience was shown to be crucial for achieving shape constancy in the case of 3D wire shapes (Edelman & Bulthoff, 1992). This theory of shape constancy predicts that a human observer would not be able to recognize a novel flat shape from its single perspective image. Figure 2 seems to refute this prediction. Figure 2(b) is a view of the Fig. 2(a) when the latter was rotated in depth by 60 deg. The prominent impression is that these two images represent one and the same shape in 3D. This observation cannot be explained easily by the multiple view theory because the reader probably had not been familiarized with many perspective views of Fig. 2(a) previously. Instead, this observation implies, as did a number of prior experimental results (see above), that shape constancy is based on invariants. This suggestion would seem to agree with Gibsonian theory in which shape constancy is based on invariants of projective or affine transformation. However, this is not true. It will be shown next that Gibsonian theory cannot account for existing results on shape constancy.

The role of invariants

In order to evaluate the Gibsonian theory of shape constancy, it is first necessary to discuss geometrical properties of projective and affine transformations. Consider projective transformation. This transformation is a more general transformation that affine and perspective; it is an arbitrary combination of perspectives. Figure 1 illustrates perspective transformation between lines L′ and L and between L and L″. The transformation between L′ and L″ is a combination of two perspectives and, hence, is an example of a projective transformation. The importance of projective transformation stems from the fact that all invariants of perspective transformation that were known in mathematics are also invariants of projective transformation. Therefore, Cassirer's (1944) and Gibson's (1950) suggestions about the role of invariants in shape constancy involved projective invariants.

Note, however, that because projective transformation is more general than perspective transformation, there are shapes that are equivalent under projective tranformation and, hence, cannot be discriminated by projective invariants, even though they may not be equivalent under perspective transformation. A perceptual implication of this fact is illustrated in Fig. 2, which shows a perspective [Fig. 2(b)] and projective [Fig. 2(c)] transformation of the pentagon shown in Fig. 2(a). As already pointed out, Fig. 2(a) and Fig. 2(b) appear to represent the same shape, but this is not as compelling in the case of Fig. 2(a) and Fig. 2(c). This example shows that it is not likely that shape constancy is subserved by projective invariants because if it were, both Fig. 2(b) and Fig. 2(c) should appear equally likely to be images of Fig. 2(a).

*It has to be pointed out that there are general case invariants for a transformation between a solid shape and a pair of images (Weiss, 1988; Koenderink & van Doorn, 1991). Thus, even in the case of solid shapes without flat faces, shape constancy could be based on invariants, provided that at least two images are available. However, psychological plausibility of such invariants is questionable because it was shown that shape constancy did not benefit from viewing the object binocularly (Rock & DiVita, 1987; Rock *et al.*, 1989) and the benefit was small when the subject was shown a large number of views (Edelman & Bulthoff, 1992).

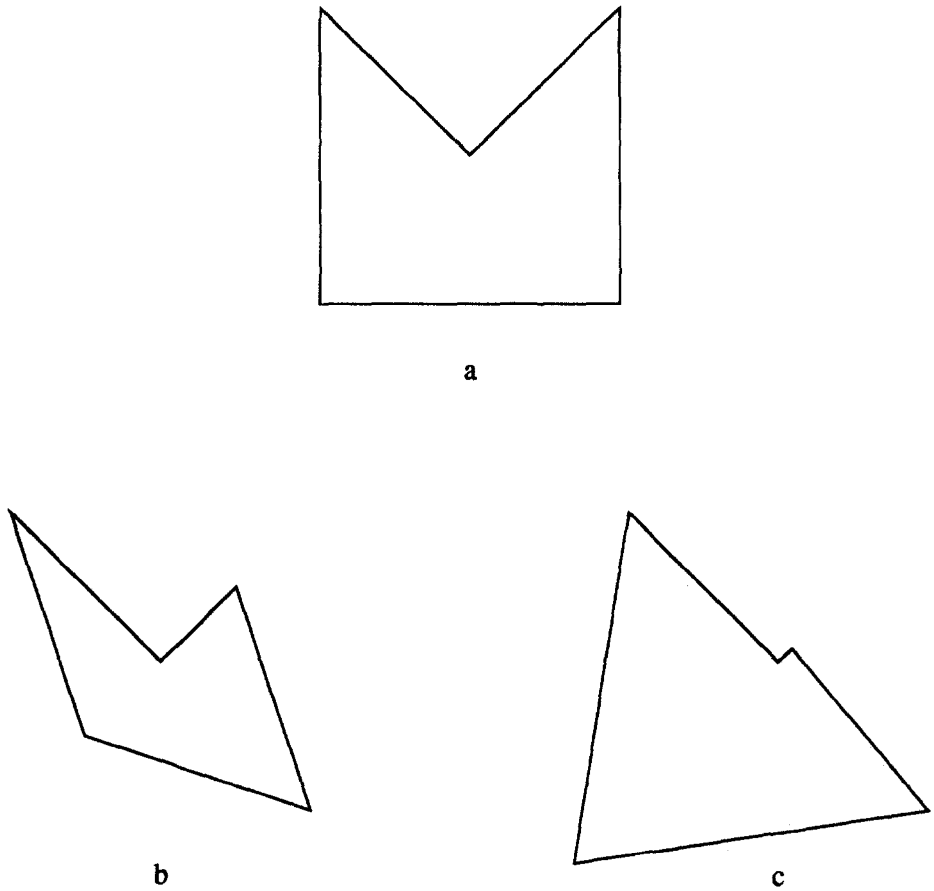

FIGURE 2. (a) Frontal view of a pentagon. (b) If you look at the center of this pentagon, keeping the page in the frontoparallel plane at a large distance from your eye, the pentagon will represent a perspective image of the shape (a) rotated around the oblique axis with slant 60 deg. (c) Projective transformation of the shape (a). Projective transformation corresponds to a sequence of two or more perspective transformations with arbitrary viewpoints. It can be seen that the shape (b), rather than shape (c), is likely to be identified as a perspective transformation of the shape (a).

The above claim that shape constancy is not based on invariants of projective transformation has received support from several studies. First, consider results of Stavrianos's Experiment I described just above. She used rectangles as stimuli, but it is known that all rectangles and their perspective images are equivalent under projective transformation and, therefore, projective invariants cannot be used to discriminate among rectangles. This means that if human shape constancy was based on projective invariants, Stavrianos would have been studying shape ambiguity, rather than shape constancy. In other words, the subject would not be able to discriminate between the two standard rectangles, and Stavrianos's results should have been the same as Thouless's. But they were quite different. Hence, projective invariants cannot account for Stavrianos's results, which means that projective invariants cannot underlie shape constancy.

The results of Stavrianos's Experiment I are also inconsistent with theories of shape constancy based on invariants of the affine transformation. As pointed out in footnote on p. 1639, the affine transformation is an approximation to perspective, which is accurate only for figures whose extent in depth is small relative to the distance of the figure from the observer. In such a case, a perspective image of a given figure is a linear transformation of this figure, i.e., the image can be obtained by changing size, stretching, translating and rotating the figure on the plane. Since each rectangle is a linear transformation of any other rectangle, affine invariants cannot be used to discriminate among rectangles. This means that affine invariants cannot account for Stavrianos's results and, therefore, they, like projective invariants, cannot underlie shape constancy.

A further attempt to test Gibsonian theory was made by Cutting in his analysis of the kinetic-depth-effect (Cutting, 1986). Cutting tested the subject's ability to perceive rigid motion of a set of four parallel lines and the extent to which the subject's performance could be predicted by the cross ratio of the four collinear points, which is a projective invariant. Despite some predictive power of this invariant, Cutting was compelled to conclude that the cross ratio cannot subserve shape constancy for three reasons: (1) it cannot be used with an arbitrary stimulus other than four collinear points, (2) if there are fewer than four collinear points, the cross ratio cannot be calculated, (3) if there are more than four points, many cross ratios can be calculated and it is not clear which points should be used, especially when their correspondence is not known. Therefore, Cutting suggested that some other quantity could account for his results, a quantity, which is not a projective invariant, but has properties similar to the cross ratio and can be easily calculated from an arbitrary set of points. It is worth pointing out that Cutting's argument about the limitations of the cross ratio of four collinear points is quite general and it applies to other projective invariants as well, such as the cross ratio of four areas represented by 5 points on the plane (Klein, 1939).

To summarize, it was shown in the last two sections that shape constancy does not involve depth cues. This contradicts both Helmholtzian and Gestalt theories. Furthermore, shape constancy may be possible for the case of a novel shape and its single perspective image. This contradicts the multiple view theory. These facts imply that shape constancy is based on invariants, but it was shown that projective and affine invariants, which were the only known invariants that could be applied to a perspective transformation, cannot explain shape constancy. This contradicts Gibsonian theory. It was pointed out, however, that projective and affine transformations are not equivalent to perspective transformation. Perspective is a different transformation and, hence, it may have its own invariants. It has to be pointed out, however, that perspective invariants must be different from conventional invariants (Pizlo, 1993). Conventional invariants are formulated for transformations that form a group (Klein, 1939; Cassirer, 1944). However, perspective does not form a group because a combination of two or more perspective transformations is a projective, not perspective transformation (i.e. the closure axiom is not satisfied). As a result, if perspective transformation had a conventional invariant, this invariant would have to have the same value for a perspective of an image, as well as for perspective of this first perspective. But this would be a projective invariant. Therefore, perspective invariants cannot be based on the concept of group (specifically, on the closure axiom). Such perspective invariants, which hold only for a perspective transformation and do not hold for an arbitrary affine or projective transformation, have recently been formulated by Pizlo and Rosenfeld (1992) and the existence of such invariants has been suggested independently by Grimson, Horn and Poggio (1992, p.77) and by Mundy and Zisserman (1992, p.25).

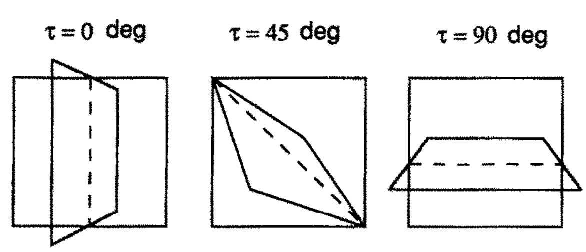

FIGURE 3. Illustration of the definition of tilt. In all three cases, slant is the same but tilt τ is not the same, it varies between 0 deg (the vertical axis of rotation) and 90 deg (the horizontal axis of rotation). Tilt determines the axis of rotation of the plane and slant determines how much the plane is rotated.

These new invariants allowed formulation of a new theory of shape constancy, called here the "Perspective Invariants Theory". In this theory, Pizlo and Rosenfeld's algorithm for shape recognition based on their invariants, serves as a model of the shape constancy phenomenon.

The next section presents the Perspective Invariants Theory. First, details of Pizlo and Rosenfeld's invariants and algorithm are described briefly and, then, it will be shown that the new theory can account for all prior psychophysical results on shape constancy.

PERSPECTIVE INVARIANTS THEORY

Pizlo and Rosenfeld's invariants and algorithm

To introduce the concept of perspective invariants of shape, parameters that characterize the orientation of a flat shape in depth must first be specified. There are two such parameters. They are called "slant" and "tilt". Slant was defined earlier in this paper as the angle of rotation of the shape in depth. Tilt is the orientation of the axis of this rotation (see Fig. 3). More precisely, tilt is the orientation of the orthographic projection of the normal of the shape on the frontoparallel plane.

Visual scientists use two complementary ways to analyze the relationship between a flat shape and its image. One is called perspective projection, i.e. an analysis of the image [such as the shape shown in Fig. 2(b)] as a function of the original shape [Fig. 2(a)]. The other is called an inverse perspective projection, i.e. an analysis of the original shape [Fig. 2(a)] as a function of its image [Fig. 2(b)]. In fact, both of these transformations are called "perspective" in geometry. However, in the case of vision it makes sense to treat these transformations separately because only one of them (perspective) describes the projection of an object out there onto the subject's retina.

Pizlo and Rosenfeld (1992) showed that the shape of the perspective image is insensitive to the distance, f, of the center of projection of the eye (or camera) from the retina (or image plane). This distance, f, is constant and known. The shape of the inverse perspective image, however, is insensitive to the distance, C, of the shape from the subject. The distance, C, is variable and unknown. Thus, the shape of the perspective image depends on 3 variables (slant, tilt and distance C) whereas the shape of the inverse perspective image depends on only 2 variables (slant and tilt) because distance, f, is constant. Therefore, it is simpler, and hence better, to use inverse perspective than perspective to analyze the relationship between a shape and its image. The first step of Pizlo and Rosenfeld's analysis of shape was to show that inverse perspective was more useful than perspective.

Their second step was to analyze the effect of the slant of an inverse perspective transformation on the segment length and angle size for an arbitrary, but constant, tilt. They showed that these effects can be approximated by

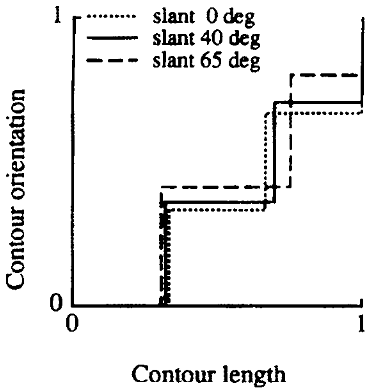

FIGURE 4. Ψ functions for a triangle and its inverse perspective images with different slants and the same tilt.

quadratic functions of the form, constant·(slant)^2. These relationships were used to describe the effect of slant on the Ψ function, which is a contour-based shape descriptor (Ballard & Brown, 1982), closely related to the intrinsic equation of a curve introduced into mathematics by Euler in 18th century (see Coolidge, 1940, p. 343).

Figure 4 shows three such functions for a triangle and its inverse perspective images with slant 40 deg and tilt 0 deg and with slant 65 deg and tilt 0 deg. The Ψ function is constructed in the following way. One starts from a given point on the contour and then moves along the contour plotting a contour orientation as a function of contour length. The total contour length is normalized to unity for each shape. Thus, the abscissa shows the relative contour length q. Similarly, the ordinate shows the relative contour orientation Ψ because it is assumed that the orientation of the contour at the starting point is zero. In the case of a polygon the Ψ function is a step function. Note that the starting point is an important parameter because the Ψ functions are different for different starting points. Tilt is also an important parameter. Different tilts produce different Ψ functions.

Pizlo and Rosenfeld (1992) formulated and proved the existence of two invariant properties of the Ψ function describing an arbitrary shape under an inverse perspective transformation:

First, for any shape, the corresponding points of the Ψ functions for all slants and for a given tilt τ and given starting point are approximately collinear. This is represented by the following set of equations:

$$\frac{q_k(\sigma_1) - q_k(0)}{\Psi[q_k(\sigma_1)] - \Psi[q_k(0)]} = \frac{q_k(\sigma_2) - q_k(0)}{\Psi[q_k(\sigma_2)] - \Psi[q_k(0)]} \qquad (1)$$

where $[q_k(\sigma), \Psi(q_k(\sigma)]$ are coordinates of a kth point on the Ψ function corresponding to the inverse perspective image of the shape with slant σ and a given tilt τ and a given starting point. This collinearity feature is easily seen in Fig. 4 for steps (i.e. discontinuities) of the functions. Successive steps of the Ψ functions are corresponding because they represent corresponding corners of polygons.

Second, for any triplet of the Ψ functions, a ratio of distances of each point on one function to corresponding points on the other two functions is approximately constant. This is represented by the following equation:

$$\frac{\Psi[q_k(\sigma_1)] - \Psi[q_k(0)]}{\Psi[q_k(\sigma_2)] - \Psi[q_k(0)]} = \frac{\sigma_1^2}{\sigma_2^2}. \qquad (2)$$

The main part of Pizlo and Rosenfeld's algorithm for the recognition of flat shapes will now be described. Let the 40 deg slant function in Fig. 4 represent the standard shape and the 0 deg slant function in Fig. 4 represent its perspective image (i.e. retinal shape). In this case the retinal shape is a perspective image of a standard shape with a slant of 40 deg. It is assumed that the tilt and the starting point of the standard shape are known and that the slant of the standard shape (40 deg) is not known. (The case of unknown tilt and unknown starting point will be discussed later.) The unknown slant of the standard shape (here 40 deg) is denoted by σ_1. Then, an inverse perspective transformation of the retinal shape (slant = 0) is calculated for an arbitrary slant σ_2 and for the tilt, which is assumed to be known. The Ψ function for the inverse perspective transformation with the slant $\sigma_2 = 65$ deg is shown in Fig. 4. Then, for some number of points on the Ψ function for the retinal shape with 0 slant, corresponding points on the function for the standard shape with unknown slant σ_1 are found using the collinearity feature [see equation (1)]. Next, ratios ρ of distances between corresponding points of all three functions are calculated [see equation (2)]. These ratios have approximately constant value [equal to $(\sigma_1/65)^2$] for all points *if and only if* the retinal shape is a perspective image of the standard shape (this statement is true for any slant σ_1). Thus, if a histogram of this ratio for a set of points is plotted, a small standard deviation (SD$\ll$1) indicates that the retinal shape is indeed a perspective image of the standard. In addition the mean value ρ_m of this histogram can be used to estimate the unknown slant according to the formula: $\sigma_1 = 65 \cdot \sqrt{\rho_m}$.

This algorithm can recognize a shape from a single perspective image without *a priori* information about the slant. For example, by applying the algorithm, the shape in Fig. 2(b) can be recognized to be a perspective image of the shape in Fig. 2(a) [SD = 0.01] and the slant can be estimated rather accurately ($\sigma_1 = 58$ deg). At the same time the algorithm correctly shows that the shape in Fig. 2(c) is not a perspective image of the shape in Fig. 2(a) [SD > 6]).

This is how the algorithm works when tilt is known. When tilt is unknown, the algorithm has to try several different tilts (usually not more than four are required to span the tilt space) to compare shapes. This requires more computational time. If sufficient time is not allowed, the recognition is likely to be incorrect. Similarly, if the starting point of the Ψ function is not known, several starting points have to be tried (usually not more are needed than a dozen). Again, this requires more computational time, and again, if the additional time is not allowed, recognition is likely to be incorrect.

This algorithm has been tested on a set of examples (see Pizlo & Rosenfeld, 1992). In all cases, the problem was to match a standard shape with its perspective image using a variety of flat shapes, namely, polygons (including rectangles), smoothly curved contours, shapes that were not exactly perspective images of the standard, partially occluded shapes, as well as polyhedra. This algorithm was tested for the case of known tilt and starting point, as well as for cases when tilt and/or starting point were not known. In all these examples, given enough computational time the algorithm correctly recognized the shapes presented as perspective images of their corresponding standard shapes.

In the next section, it will be shown that the Perspective Invariants Theory, in which Pizlo and Rosenfeld's (1992) algorithm serves as a model of shape constancy, is consistent with prior psychophysical results on shape perception. Then, new psychophysical experiments to test the theory, will be described.

Comparison of the Perspective Invariants Theory with prior psychophysical results

Perspective Invariants Theory assumes that Pizlo and Rosenfeld's (1992) algorithm is a computational model of shape constancy. This means that the algorithm should account for all known psychophysical results on shape constancy and it should also allow formulating testable, new predictions. The new theory is not, however, restricted to computational properties of Pizlo and Rosenfeld's algorithm. It also includes a new treatment of the difference between perspective and projective transformations and between invariants of these transformations (see the section on the role of invariants), as well as a new treatment of the problem of shape ambiguity (see the section on the role of depth cues).

Consider now a comparison between predictions of Perspective Invariants Theory and results of prior psychophysical experiments. In the new theory: (1) The observer can recognize a flat shape on the basis of a single perspective image of the shape, taken from an unknown viewing direction and distance (but not on the basis of an arbitrary projective or affine image). In other words, the observer can solve the shape constancy problem without using depth cues and without seeing many perspective views of the shape before. This agrees with a number of psychophysical studies (e.g. Stavrianos's reported in the section on the role of depth cues as well as with the demonstration experiment in Fig. 2). (2) The failure to discriminate among shapes that give rise to the same retinal image (i.e. shape ambiguity) does not represent a failure of shape constancy (see the section on the role of depth cues). (3) The observer can achieve shape constancy in the case of solid shapes that have flat faces without using depth cues. In such a case, shape constancy can be based on the constancy of individual faces. This agrees with prior experimental results (Shepard & Metzler, 1971; Shepard & Cooper, 1982). (4) Shape constancy cannot be achieved for solid shapes that do not have flat or nearly flat faces. This also is consistent with prior results (Rock & DiVita, 1987; Rock *et al.*, 1989; Edelman & Bulthoff, 1992), all of which show that it is difficult to achieve constancy with 3D wire shapes. (5) The observer can use shape as a cue for depth because, according to Perspective Invariants Theory, the slant of the presented shape is estimated after its shape is recognized. This is consistent with Stavrianos's results, where it was shown that slant could be judged even in the absence of cues for depth (such as binocular disparity, vergence, accommodation), although not as precisely and as accurately as in their presence. Note that this feature makes the Perspective Invariants Theory similar to Gestalt theory. In Gestalt theory, perceived shape and perceived slant can affect one another. In the Perspective Invariants Theory this relation is assumed to hold, but only in one direction, namely, perceived shape determines perceived slant not the reverse as Gestalt theory would hold. This, in the author's view is an important difference. (6) The observer's capacity to recognize a shape does not depend on the slant of the shape, i.e. on the angle of rotation of the shape in depth, because the algorithm is insensitive to slant. However, the algorithm is sensitive to tilt in that tilt must be either known or searched for. The insensitivity of the algorithm to slant agrees with prior psychophysical results of Stavrianos (1945). Sensitivity to tilt had never been tested in the past. This sensitivity, along with other predictions of the Perspective Invariants Theory, will be tested and shown to support the theory in new experiments described in the next section.

TESTING PREDICTIONS OF THE PERSPECTIVE INVARIANTS THEORY IN PSYCHOPHYSICAL EXPERIMENTS ON SHAPE MATCHING

The main element of the Perspective Invariants Theory, which makes it different from the Helmholtzian and Gestalt theories, it that shape constancy is based on invariants and does not involve depth cues. To test this feature it is necessary to test human performance in an experiment in which depth cues indicating the orientation of the shape are severely reduced or even eliminated entirely. To accomplish this, line drawings shown on a flat CRT display, were used. Subjects should be able to recognize that one shape is a perspective image of the other if shape constancy can be achieved without cues to depth with such stimuli.

Perspective Invariants Theory is also different from Gibsonian theory because it involves invariants that are different from traditional geometrical invariants (projective or affine). Namely, in the new theory, the percept of shape is insensitive to rotations in depth only when tilt is known (although tilt can be arbitrary). Therefore, to test this prediction the effects of both slant and tilt on shape constancy should be examined. Finally, the new theory is different from the multiple view theory because it predicts that the subject can recognize a shape from a novel image without having experience with many different views of the standard shape. To test this prediction a single view of a randomly generated polygon was used as the standard stimulus.

The subject's performance was tested in a shape matching task, whose main features were as follows: The subject was briefly shown a *standard* shape. She was then required to indicate whether another shape, the *comparison* shape, seen for the first time, was a perspective image of the standard shape seen shortly before. *Note that there is an infinite number of perspective images of the standard shape, each of them specified by a particular slant and tilt. In other words, the subject was required to match the shape attribute of a comparison polygon to the shape attribute of a previously seen standard in the presence of variations of other attributes, namely, slant and tilt.*

Now, consider predictions related to the sensitivity of the shape judgments to slant and tilt. According to the Perspective Invariants Theory, the human subject is sensitive to tilt because if tilt is not known, the comparison shape and the standard shape have to be compared for a number of tilts. This increases the time required for processing. If processing time is not increased, the response may be incorrect. Therefore, the theory predicts that performance, both speed and accuracy, depends on whether the range of variation of the tilt of the comparison stimuli is wide or narrow. If the range is narrow and both the range and its limits are known to the subject, a correct response can be produced quickly by using any arbitrary tilt selected from within this narrow range. If the range is wide, however, an arbitrary tilt selected from this wide range is much more likely to produce an incorrect response.

By contrast, the Perspective Invariants Theory predicts that the shape judgments will be insensitive to the range of variation of slant because, according to this theory, information about slant is not used in shape constancy. Therefore, the accuracy of the response, as well as the time required for processing, is predicted to be the same regardless of whether the range of slant variation is wide or narrow.

The new theory predicts also an interaction of the range of variation of slant and tilt. Namely, the harmful effects of unknown tilt (slow processing or high error rate) is predicted to be small when slant is small. For slants close to zero, the shape of the perspective image is very similar to the shape of the object for any tilt, and thus for such small slants the ability to indicate that the comparison shape is a perspective image of the standard would not be very sensitive to tilt.* Therefore, an additional prediction of the Perspective Invariants Theory is that a large range of variation of tilt is more likely to be deleterious when slant is large.

Consider now the predictions of prior theories of shape constancy. Helmholtzian theory assumes that shape constancy involves depth cues. Therefore, it predicts that in the absence of depth cues the subject's performance will be poor unless the subject can successfully guess the slant and tilt, which might be possible if the range of variation in slant and tilt was small and known to the subject. So, in all, but the highly constrained case of narrow ranges of slant and tilt, performance is predicted to be poor. Thus, the prediction of the Helmholtzian theory is different from the prediction of the Perspective Invariants Theory.

In Gestalt theory the perception of shape is related to the perception of orientation and for a fixed retinal image both perceived shape and perceived orientation are undetermined, unless contextual information is available that determines both. Note that in the present experiments the only contextual information, which could be used to determine the orientation of the comparison stimulus in a given trial, was provided by the comparison stimuli in other trials. Therefore, Gestalt theory predicts that shape constancy cannot be achieved, unless orientation was the same for all of the comparison stimuli. Thus, Gestalt theory, like Helmholtzian theory, predicts that in all, but the highly constrained case of narrow ranges of slant and tilt, performance will be poor.

Next, consider the prediction of Gibsonian theory. According to this theory, shape constancy holds for any slant and tilt and, therefore, the theory predicts that the range of tilt and slant variation will not affect shape perception at all. Thus, the prediction of Gibsonian theory is different from the prediction of the Perspective Invariants Theory, of Helmholtzian and of Gestalt theories. It is worth remembering that there are two versions of Gibsonian theory and they make different predictions. In one version, shape constancy is based on affine invariants and in the other shape constancy is based on projective invariants (see the section on the role of invariants for details about affine and projective invariants). If affine and projective invariants are applied to affine images of quadrilaterals (i.e. to perspective images of quadrilaterals whose extent in depth is small relative to their distance from the observer), they will lead to very different results. Namely, affine invariants will allow perfect discrimination between affine and non-affine images (regardless of slant and tilt). Projective invariants, on the other hand, will produce performance at chance level (regardless of slant and tilt) because calculation of projective invariants for the 2D shape requires more than 4 points and, hence, quadrilaterals cannot be discriminated on the basis of projective invariants.†

Finally, consider the multiple view theory. According

*The fact that small slants have little effect on the shape of a perspective image is implied by Pizlo and Rosenfeld's (1992) theoretical analysis of the effect of slant on angles and distances. This effect can be approximated by a quadratic function of the form: constant·(slant)^2. Note that this function does not have a linear term. Thus, if slant is close to zero, the square of slant is very small, which means that angles and distances are almost exactly constant and this is true regardless of the value of tilt. This fact was demonstrated earlier in simulation analyses by Ben-Arie (1990) and by Burns *et al.* (1990).

†The fact that 4 non-collinear points are not sufficient to apply projective invariants does not mean that all quadrilaterals are projectively equivalent. It only means that all quadruples of non-collinear points are projectively equivalent. For example, although there is no projective transformation, which can transform a convex quadrilateral into a concave quadrilateral (Isaac Weiss—personal communication), there is always a projective tranformation that can transform vertices of a convex quadrilateral into vertices of a concave one.

to this theory the subject's performance should be at chance level in all conditions because in this theory, shape constancy is based on memorizing many perspective views of each standard. In the present experiment, every standard shape will be a random polygon never seen before by the subject, which means that there will be no opportunity to memorize many perspective views for any shape.

Experiment 1: The Effect of the Ranges of Slant and Tilt

Method

Subjects. Three subjects (including the author) participated in this experiment (ZP, JE and AP). ZP and JE were myopes and they used their normal corrective lenses during the experiment. AP was emmetropic. Each subject received about 4000 practice trials before starting these experiments. These practice trials were similar to those used in the main experiment. This amount of practice proved to be sufficient for all subjects to achieve stable reaction times (RTs).

Stimuli. Stimuli were generated on a Tektronix 604 (P4 phosphor) CRT controlled by an LSI 11/73 computer. Refresh rate was 50 Hz. Luminance intensity per point was 0.09 μcd (Sperling, 1971). The background luminance was 0.11 cd/m^2 and the room luminance was 0.68 cd/m^2. All of the stimuli were clearly visible and free of flicker. The subject viewed the stimuli with her right eye (the left eye was covered) and the subject's head was supported on a dental biteboard. These arrangements minimized cues available for processing the depth of the screen that could interfere with the percept of the shape presented on the display. The distance of the subject's eye from the display was 50 cm.

Randomly-shaped quadrilaterals were used as stimuli (see Fig. 5 for examples). The first stimulus shown on each trial was the standard quadrilateral located at the center of the display. Positions of the vertices of each quadrilateral were generated randomly on the circumference of a circle with a diameter of 1 cm (1.15 deg at the 50 cm viewing distance). Adjacent points were connected with straight lines to form the quadrilateral. The width of the line connecting the points was 1 pixel.

The second stimulus in each trial was the comparison shape. In half of the trials, the comparison shape was a perspective image of the standard quadrilateral (called henceforth a *perspective* quadrilateral). In the rest of the trials, the comparison shape (henceforth called a "*non-perspective* quadrilateral") was a perspective image of a quadrilateral that was different from the standard quadrilateral. This different quadrilateral, which was used to generate the non-perspective quadrilateral, was constructed by displacing one randomly-chosen vertex of the standard quadrilateral clockwise or counter-clockwise by a central angle of 50 or 60 deg (the direction of displacement was also random). The specific angle of the displacement was chosen in such a way that in the easiest condition (which was: the narrow range of tilt, wide range of slant, the longest RT deadline) the subject responded correctly on about 90% of the trials. The angular displacement meeting this requirement was 50 deg for subjects ZP and AP, and 60 deg for subject JE.

Perspective images of shapes were calculated by using a slant and tilt pairs. Ten equally spaced slants and tilts were used for each block of 100 trials. Each pair was used only once in a block and the order of their presentation was randomized. Two ranges of tilt variation were used, "narrow" (± 11.25 deg) and "wide" (± 45 deg). Both of these ranges were centered around the same value, namely, 0 deg (i.e. the axis of rotation was vertical). This choice of tilt ranges was based on the simulation results of Pizlo and Rosenfeld (1992), who showed that in the case of unknown tilt it is necessary to check, at most, four tilts in any quadrant that are about 20 deg apart. This means that varying tilt randomly over a range of 90 deg (the wide range) should make it harder to compare the comparison and standard shapes than varying tilt over a range of about 20 deg (the narrow range) because in the former case more tilts will

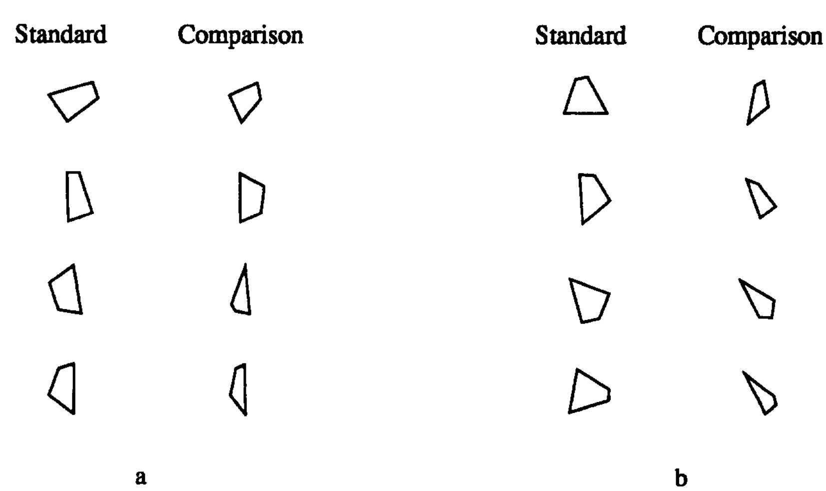

FIGURE 5. Examples of stimuli used in the experiments. Standard shapes are shown in the left column and their corresponding comparison shapes are shown in the right column. (a) The range of tilt variation was -11.25, $+11.25$ deg and the range of slant variation was 30, 60 deg. (b) The range of tilt variation was -45, $+45$ deg and the range of slant variation was 52.25, 60 deg.

have to be checked before a successful match of standard and comparison stimulus can be made.

Now, consider slant. Slants close to zero were not used because small slants do not change shape, and large slants (close to 90 deg) were not used because shape becomes too narrow when slants are large. Thus, the wide-slant range was 30–60 deg. It was four times wider than the narrow range (52.5–60 deg). Large slants (close to 60 deg) were used in the narrow range because the Perspective Invariants Theory predicts the harmful effect of tilt to be noticeable when slants are large. The magnitude of slants which is perceptually large could not be known in advance. For this reason slants were varied at the largest possible level that could be used without causing shapes to become to narrow (see just above).

Procedure. Each block of 100 experimental trials began with 10 practice trials. The subject had the option to repeat the 10 practice trials if she felt it was necessary. Pressing two buttons simultaneously initiated each experimental or practice trial [the subject used her dominant (right) hand]. Buttons remained depressed until the response was made. On each trial, the standard quadrilateral was shown for 1 sec and then it disappeared. The screen then remained blank for 1 sec during which time the subject was asked to fixate at the center of the screen, i.e. the place where the comparison quadrilateral was to be presented. At the end of the one sec blank interval, a comparison quadrilateral was shown for 100 msec. The subject's task was to indicate whether the comparison quadrilateral was a perspective quadrilateral. The subject indicated her choice by releasing one of two buttons. The assignment of buttons (left vs right) to response type (perspective vs non-perspective) was randomized between blocks. The subject was given auditory feedback about the accuracy of her response at the end of each practice or experimental trial. The time required to make each response (RT), as well as the accuracy of each response, was recorded.

In a block of 100 experimental trials, the comparison quadrilateral was a perspective quadrilateral in 50 trials and a non-perspective quadrilateral in the other 50 trials. The presentation order of the perspective and non-perspective comparison quadrilaterals was randomized. The subject was informed as to the range of slant and tilt variation that would be used in a given block of trials. Practice trials that preceded the experimental trials also provided this information.

The subject's performance was summarized by the speed–accuracy tradeoff function (SATF). The SATF is the relation between the accuracy of the subject's response and the mean RT. When the response is very fast (short RT), accuracy is close to the chance level. When the response is slow (long RT), accuracy approaches an asymptotic level. For the intermediate RTs, the SATF can be approximated by straight line and this line can be used as the dependent variable (Luce, 1986). To estimate the SATF, one needs several data points representing different levels of RT. To obtain such data points the subject's speed–accuracy criterion was varied by using an auditory deadline and asking the subject to respond just before this deadline. Six deadlines were used: 400, 450, 500, 550, 600 and 650 msec. The deadline was constant within a block of practice and experimental trials and was varied between blocks. After the 10 practice trials, which started each block, the subject was informed about the mean RT exhibited on these trials.

Design. There were a total of 24 different kinds of experimental blocks: 2 ranges of slant variation × 2 ranges of tilt variation × 6 deadlines. Each set of 24 blocks (one replication) was repeated 5 times for subjects JE and AP and 4 times for subject ZP. The order of conditions was determined by a randomized Latin square design, with each subject serving in a different order of test conditions.

Analysis. The SATF was estimated by using logit analysis (Ashton, 1972). The proportion of correct responses (p) was transformed to the logit by $\ln[p/(1-p)]$ and this was plotted as a function of the mean RT. The SATF was estimated separately for each subject, each experimental condition and each replication by fitting a straight line to the experimental points. The equation of the best fitting line can be written as logit $= m(\text{RT}-c)$, where m is the slope and c is the RT intercept. The slopes and intercepts of the individual lines were then analyzed by a repeated measures ANOVA. This analysis was performed separately for each subject.

Results

Preliminary analysis. The order of the mean RTs was the same as the order of the deadlines for each of the subjects, which means that the subjects varied their RTs as requested. Then, it was verified that each subject's speed-accurcy criterion varied reliably between blocks as the deadlines were varied. This was determined by analyzing the main effect of the various deadlines on the mean RTs and on error rates. This main effect was significant for all three subjects with $P < 0.001$. With respect to RT, the F ratios were as follows: subject ZP, $F(5,75) = 591$; subject JE, $F(5,95) = 383$, subject AP, $F(5,95) = 283$. With respect to error rate the F ratios were as follows: subject ZP, $F(5,75) = 22.6$; subject JE, $F(5,95) = 59$, subject AP, $F(5,95) = 165$. These results show that the subjects reliably varied their speed–accuracy criteria as requested.

Estimated SATFs and data points for each subject, each condition and each replication are shown in Fig. 6. It is seen that most data points for all subjects and experimental conditions represent accuracy better than $p = 0.61$ (logit $= 0.45$), which is significantly ($P < 0.01$) greater than the chance level. Thus, the subjects were able to match shapes with their perspective images in the absence of cues to depth and without having experience with many different views of the standard shape. This experimental fact is consistent with the main prediction of the Perspective Invariants Theory.

The goodness of fit of an approximating linear SATF to the data points was determined by a Chi-square (χ^2) test (Ashton, 1972). If $\chi^2_{\text{observed}} \geq \chi^2_{\text{critical}}$ at a given significance level, it means that there is a significant heterogeneity in the departures of the data points from the line fitted. There are two possible sources of such heterogeneity. First, the function used in the approximation may not be appropriate, producing systematic departures of data points from the line fitted. Second, the observed random variability of the data points may be larger than assumed in the analysis, which means that there are some additional factors that contribute to the random variability of individual data points.

Assuming $P = 0.05$ as the significance level, significant heterogeneity was found in two cases for subject ZP, in one case for subject JE and in two cases for subject AP. The estimation of the SATF was repeated with a quadratic function, rather than the linear function, to determine whether the heterogeneities observed were produced by using an inappropriate approximating function. The significance of the coefficient of the quadratic component was determined by means of the F-test (Bevington, 1969). If the SATF grows faster (or slower) than a linear function would grow, the coefficient of the quadratic component is expected to be significantly different from zero.

The coefficient of the quadratic component was significantly different from zero in five cases for subject ZP (in three cases the coefficient was negative and in two cases it was positive), and in two cases for subject JE (in both cases the coefficient was negative). The coefficient of the quadratic component was not significant in any case for subject AP. Thus, it is not likely that the significant heterogeneity in the departures of the data points from the fitted line was produced by any systematic departures of the data points from the linear function used to estimate the SATFs (i.e. it is not likely that the SATF was either always positively or always negatively accelerated). It is more likely, then, that the heterogeneities in the departures of the data points from the fitted line were produced by some additional factors that merely increased the random variability of individual data points. The instability of the speed–accuracy

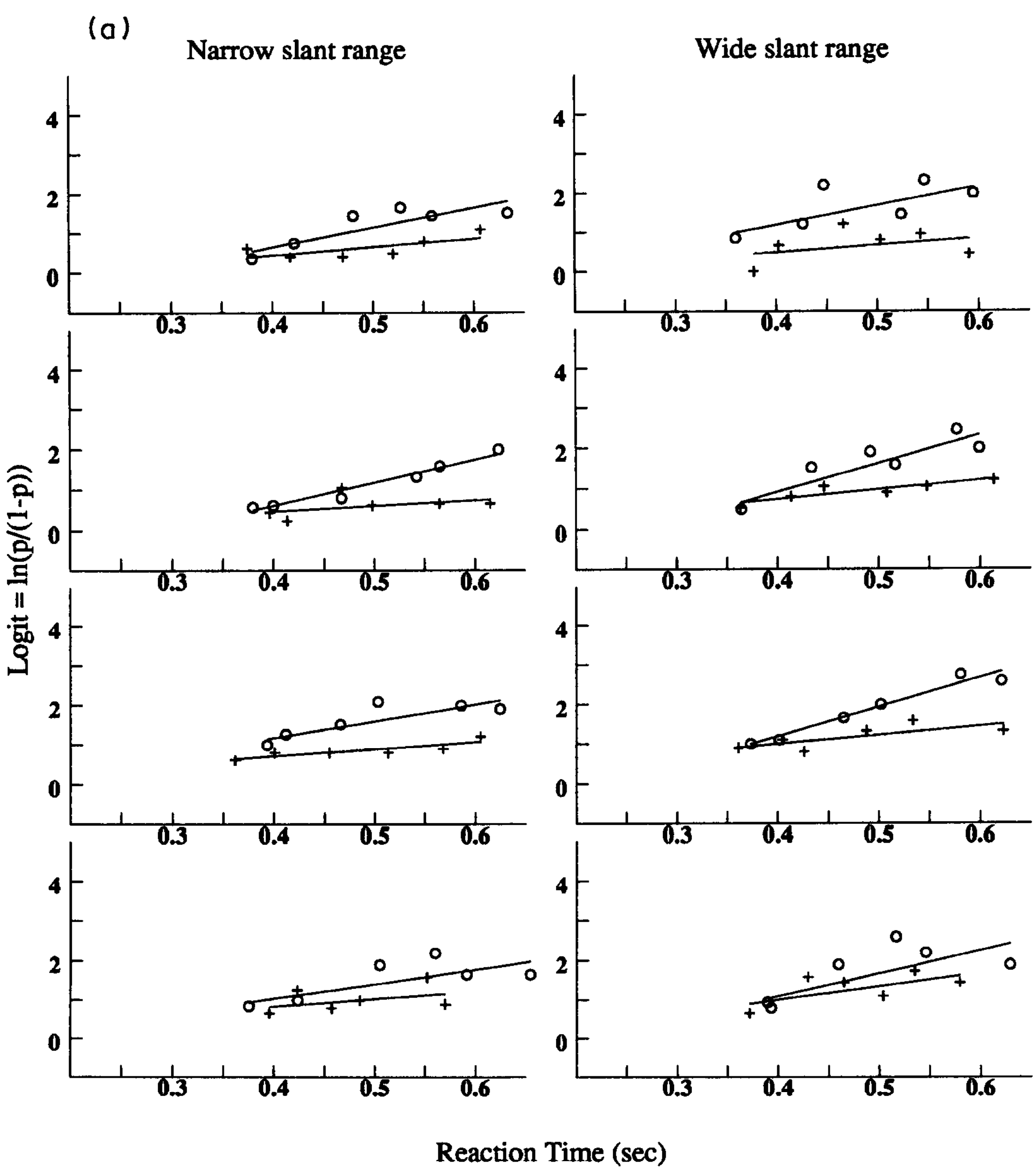

FIGURE 6(a). *Caption on p. 1651.*

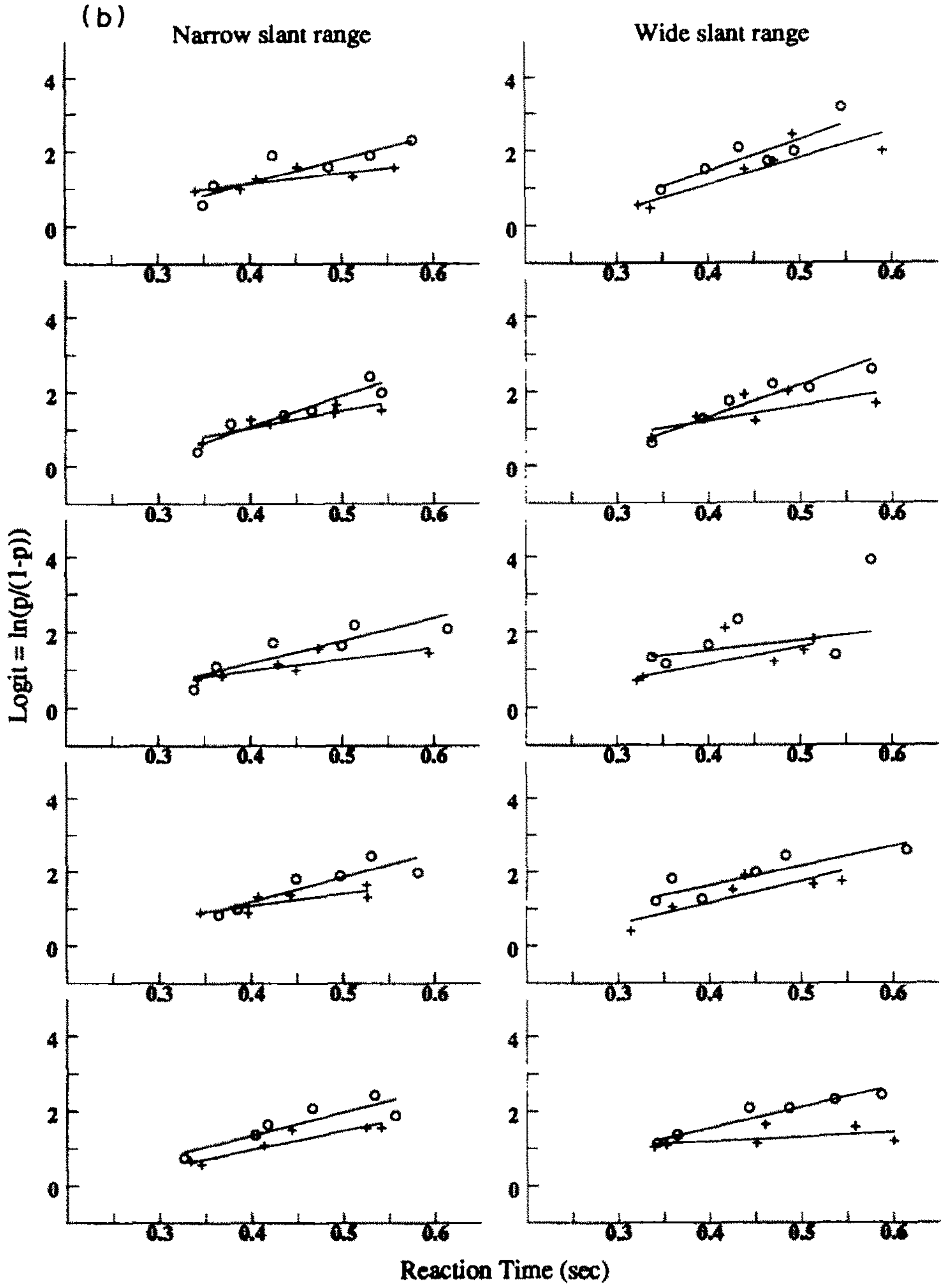

FIGURE 6(b). *Caption on facing page.*

criterion within a block of trials could be one such factor. Heterogeneity of the stimulus set due to differences in the difficulty of processing different quadrilaterals could be another factor. Note, however, that this increased variability of the data points did not overshadow the speed–accuracy tradeoff and, hence, was not a problem in the analysis.

Analysis of the slope of the SATF

The effect of tilt. The effect of the range of tilt variation can be seen in the greater slope for the narrow (circles) than for the wide (crosses) range of tilts regardless of the range of slant variation (see Fig. 6). The mean values of the slopes of the SATFs are given in Table 1 and the mean values of RT intercepts are given in Table 2. The main effect of the range of tilt variation on slope was significant for all subjects [subject ZP, $F(1,3) = 40.8$, $P < 0.01$; subject JE, $F(1,4) = 12.5$, $P < 0.025$; subject AP, $F(1,4) = 13.4$, $P < 0.025$]. These results mean that shape processing is more efficient when the range within which tilt is varied is narrow, i.e. accuracy is higher for a given RT. This confirms one of the predictions of the Perspective Invariants Theory.

The effect of slant. The range of slant variation did not have a large effect on the SATF as can be seen in Fig. 6, by comparing the left and right columns of graphs. The wide range of slants produced a slightly steeper slope than the narrow range of slants as can be seen by an examination of the mean slopes of the SATFs in Table 1. Thus, a wide range of slant variation did not impair perceptual processing. Rather, it was the narrow range, which contained only large slants, that tended to impair

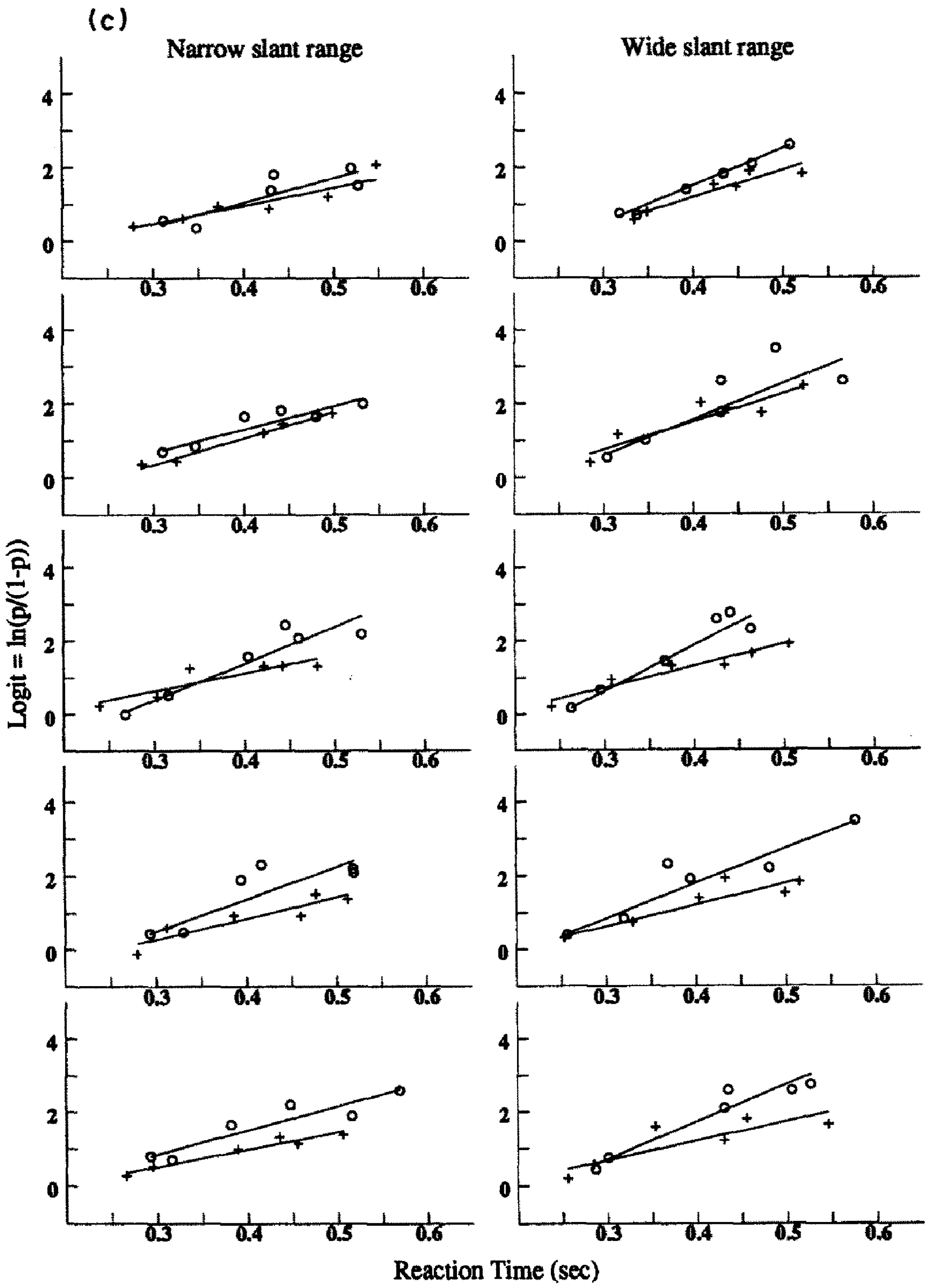

FIGURE 6. Results from the Experiment 1. A narrow range of slant variation is shown in the left column and a wide range of slant variation is shown in the right column. Circles show a narrow range of tilt variation and crosses show a wide range of tilt variation. Successive rows of graphs represent individual replications of the experiment. (a) Results of ZP. (b) Results of JE. (c) Results of AP.

processing. This effect was only statistically significant in subject AP [$F(1,4) = 18.8$, $P < 0.025$], but this tendency was also observed in JE and ZP, although it was not significant ($P > 0.05$). Thus, the effect of slant on the SATF agrees with the prediction of the Perspective Invariants Theory; namely, it is not the width of the range of slant variation, but rather, large slants in the presence of tilt variation that impair the perceptual processing of shape.

Slant–tilt interaction. Table 1 shows that the beneficial effect of the narrow range of tilts on the slope of the SATF is weaker when the range of slants is narrow. This means that in the case of narrow range of slants the subject's percept was sensitive to tilt changes even over the relatively narrow range (± 11.25 deg). It is likely that narrow range of slants produced more sensitivity to tilt not because the range was narrow, but because the slants used in the narrow range were large. This interaction between slant and tilt was seen in all three subjects, although it was significant only for subject AP

TABLE 1. Mean slopes of SATFs (1/sec) for the Experiment 1

Subject	Slant: narrow range, Tilt: narrow range	Slant: narrow range, Tilt: wide range	Slant: wide range, Tilt: narrow range	Slant: wide range, Tilt: wide range
ZP	4.63	1.75	6.29	2.46
JE	6.70	3.78	6.13	4.54
AP	7.59	5.46	10.4	6.36

TABLE 2. Mean RT intercepts of SATFs (sec) for the Experiment 1

Subject	Slant: narrow range, Tilt: narrow range	Slant: narrow range, Tilt: wide range	Slant: wide range, Tilt: narrow range	Slant: wide range, Tilt: wide range
ZP	0.199	0.041	0.221	0.071
JE	0.220	0.104	0.104	0.015
AP	0.220	0.213	0.236	0.196

[$F(1,4) = 13.0$, $P < .025$].* This interaction was probably produced by the same factor that produced the effect of the range of slants on the slope of the SATF, namely, changes of tilt give rise to large changes of the retinal shape only when slant is large.

Analysis of the RT intercept of the SATF

Only subject ZP showed a significant effect of the range of tilt on the RT intercept [$F(1,3) = 16.7$, $P < 0.05$]. For him, the wider tilt range produced the lower RT intercept. (The same, but non-significant, tendency was observed in the two other subjects.) The other effects were not significant ($P > 0.1$). It is likley that the intercept is not independent of the slope (see Jennings, Wod & Lawrence, 1975) and, therefore, a change in the RT intercept, observed in this experiment, merely reflects the change in the slope of the SATF. This explanation seems to be reasonable because the RT intercept was estimated, in many cases, from an extrapolation of the data points and, therefore, this estimate might not be accurate. This means that the slope of the SATF, rather than the intercept, is the better measure of the subject's performance.

Discussion

The results of the experiment agree with predictions of the Perspective Invariants Theory. Specifically, (1) the subject can match a shape with its perspective image without depth cues and without seeing many different views of the standard shape, (2) the range of slant variation has no effect on the efficiency with which the shape is processed and (3) the range of tilt variation does have an effect on processing, namely, the wider range impairs perceptual processing. These results are not consistent with either the Helmholtzian, Gestalt or the multiple view theory. They also contradict Gibsonian theory. As was pointed out earlier in this paper, if projective invariants underlay shape constancy, the subjects should perform at a chance level in all experimental conditions because projective invariants require more than 4 points and, hence, quadrilaterals cannot be discriminated on the basis of projective invariants. Performance was also not likely to be based on affine invariants. Affine invariants are insensitive to both slant and tilt and, hence, if the shape matching were based on affine invariants the subject would perform at the same, high level in all conditions. Thus, all four prior theories are contradicted by the present results while Perspective Invariants Theory is supported by all the results. This means that Perspective Invariants Theory is the only theory that can account for the subject's ability to discriminate perspective from non-perspective images. Perspective Invariants Theory also provides a quantitative measure of the similarity of 2D shapes, regardless of perspective changes. Such a similarity can underlie shape constancy in a human observer.

It is worth pointing out, however, that the results of this experiment do not allow rejecting the suggestion that depth cues *cannot* facilitate shape constancy. Note that the narrow range of tilts did give rise to better performance. Therefore, if depth cues were present and if tilt was estimated from these depth cues, the improvement in performance could have been based on the presence of these depth cues. This, however, was recently shown not to be the case in a subsequent experiment in which a perspective image of a square, textured background was shown along with the comparison shape. This background was rotated in 3D with the same slant and tilt as the comparison shape, and thus, provided cues to slant and tilt. However, this background did not produce a significant, or even systematic effect on the subject's performance (Salach-Golyska, Pizlo, Vyain & Standish, 1993).

The next experiment tested the suggestion, apparent in the results of Experiment 1, that the amount of slant in the presence of a variation of tilt is an important factor, affecting the slope of the SATF. Specifically, in the next experiment the SATF was measured for the case of a narrow range of slant at moderate slants, rather than at large slants, as was the case in the first experiment. This was done for both wide and narrow ranges of tilt. This experiment was performed for two reasons. First, it extended the results of the first experiment to new experimental conditions. Second, it tested another prediction of the theory. Namely, it was pointed out that the harmful effect of tilt variation should be smaller when only moderate slants are used (the results of the first experiment supported this prediction). It is even possible that the effect of the range of tilt might be eliminated

*This interaction can be seen in the data of ZP and AP but not in JE. Note that for JE the effect of the range of tilts in the case of the wide range of slants was not very strong. This was produced by the fact that in the narrow range of tilts, the mean slope (6.13) is relatively small. It has to be pointed out, however, that this small mean slope was caused by only one very small slope in the third replication for this subject [see the graph in the third row and in the right column in Fig. 6(b)]. In this case, the small slope was produced by one data point, which is clearly an outlier. Without this point the mean slope for this condition would be 7.81 instead of 6.13, which agrees with the tendency described in the other subjects.

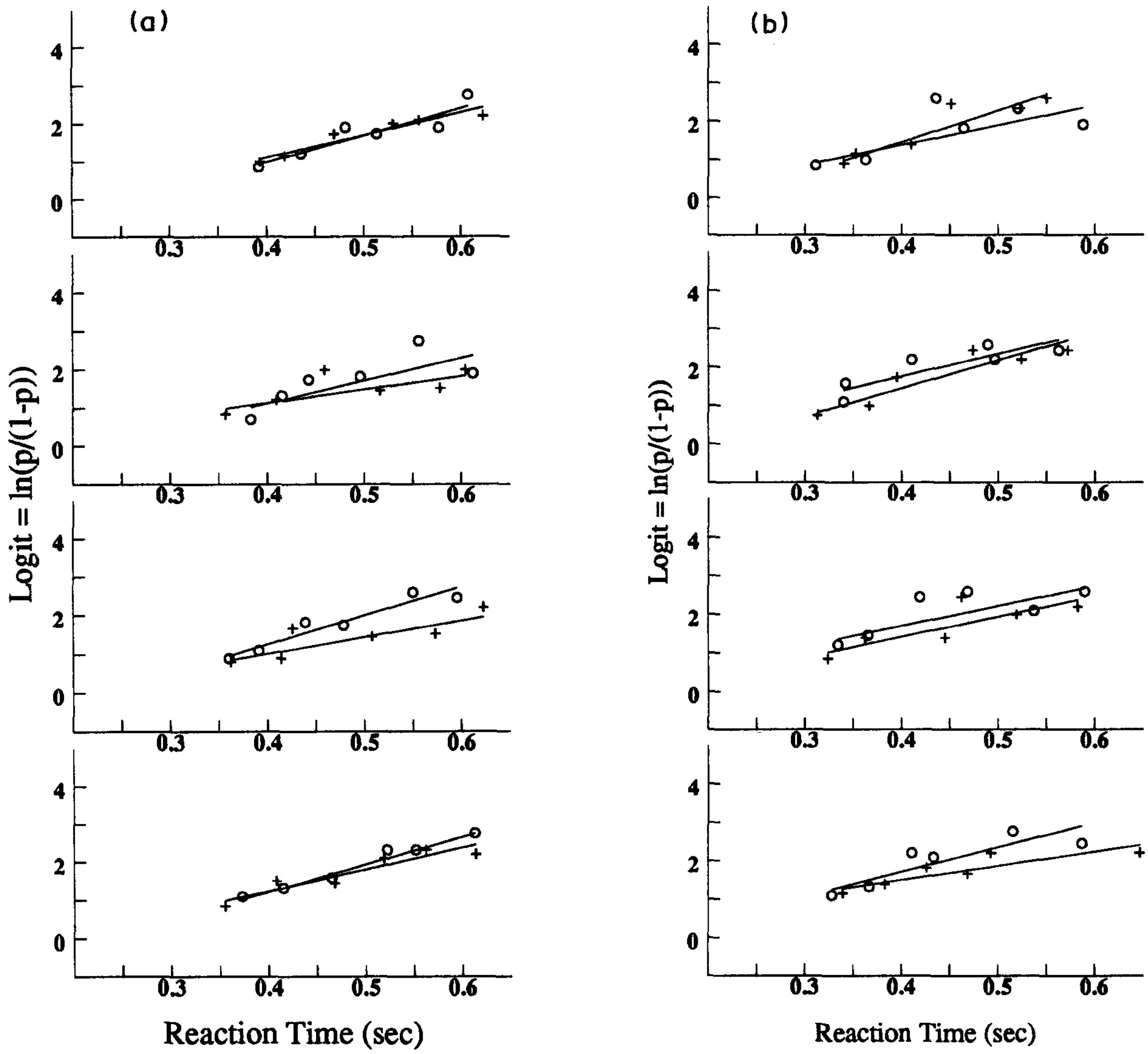

FIGURE 7. Results from the Experiment 2. Circles show a narrow range of tilt variation and crosses show a wide range of tilt variation. Successive rows of graphs represent individual replications of the experiment. (a) Results of ZP. (b) Results of JE.

completely once moderate, rather than large slants, are used.

Experiment 2: The Effect of the Range of Tilts for Moderate Slants

Method

Subjects. Two of the subjects (ZP and JE), who served in the first experiment, participated in this experiment.

Stimuli. The stimuli used in this experiment were the same as those in the first experiment except that only moderate, rather than large, slants were used, namely, from 41.25 to 48.75 deg. The width of this range (7.5 deg) was the same as the width of the narrow range in the prior experiment. The difference was that in this experiment this range was centered around 45 deg, whereas in the previous experiment the range was centered around 56.25 deg. Although the difference between 45 deg and 56.25 deg does not seem to be large, the difference in the magnitude of image change is considerable because the magnitude of the change of the area of the image varies with the cosine of the slant. The same two ranges of tilt variation were used as in the first experiment, viz. from -11.25 to $+11.25$ deg and from -45 to $+45$ deg.

Procedure. The procedure was identical to that used in the first experiment.

Design. There were a total of 12 different kinds of experimental blocks: 2 ranges of tilt $\times$ 6 deadlines. Each single replication of the set of 12 blocks was repeated 4 times for each subject.

Results and discussion

Preliminary analysis. The order of mean RTs was the same as the order of deadlines for each of the subjects. The main effect of the deadline on mean RT and on error rate was significant for both subjects with $P < 0.001$. With respect to RT, the F ratios were as follows: subject ZP, $F(5,35) = 511$; subject JE, $F(5,35) = 245$. With respect to error rate the F ratios were as follows: subject ZP, $F(5.35) = 55.5$; subject JE, $F(5,35) = 46.7$. This

TABLE 3. Mean slopes of SATFs (1/sec) for the Experiment 2

Subject	Tilt: narrow range	Tilt: wide range
ZP	6.87	4.79
JE	5.61	6.06

means that both subjects reliably varied their speed–accuracy criteria as requested.

Estimated SATFs and data points are shown in Fig. 7. As in the Experiment 1, most of the data points show significantly ($P < 0.01$) better than chance performance (logit > 0.45).

A significant heterogeneity in the departures of the data points from the best-fitting line was found in one case for subject ZP and in one case for subject JE. The estimation of the SATF was repeated with a quadratic function. The coefficient of the quadratic component was significantly different from zero in one case for subject ZP (the coefficient was negative). The coefficient of the quadratic component was not significant in any case for subject JE. Thus, it is likely that the heterogeneities in the departures of the data points from the fitted line were produced by factors that merely increased the random variability of individual data points. Similarly, as in the first experiment, this increased variability of the data points did not overshadow the speed–accuracy tradeoff and, hence, was not a problem in the analysis.

Analysis of the slope and RT intercept of the SATF. The mean values of the slopes of the SATFs are given in Table 3 and the mean values of RT intercepts are given in Table 4. There was no significant effect of tilt range on the slope of the SATF for subject JE [$F(1,3) = 0.14$, $P > 0.7$] but this effect was significant for subject ZP [$F(1,3) = 21.3$, $P < 0.02$]. The larger slope was obtained with the narrower tilt range. There was no significant effect of tilt range on the RT intercept for subject JE [$F(1,3) = 0.24$, $P > 0.6$] but the effect was significant for subject ZP [$F(1,3) = 11.4$, $P < 0.05$]. A larger RT intercept was obtained with the narrower tilt range. As was the case in the first experiment, it seems likely that the intercept is not independent of the slope of the SATF and, therefore, a change in the RT intercept merely reflects the change of slope.

The results of subject JE suggest that at moderate levels of slant, the tilt range does not affect the efficiency with which the shape is processed. This result tends to agree with the predictions of Perspective Invariants Theory. For sufficiently small slants the perspective image of a shape is not very different from the original shape regardless of the value of tilt (see footnote on p. 1646).

TABLE 4. Mean RT intercepts of SATFs (sec) for the Experiment 2

Subject	Tilt: narrow range	Tilt: wide range
ZP	0.233	0.154
JE	0.108	0.137

Consider now the results of subject ZP. For this subject the tilt range was a significant factor and, as in the first shape experiment, the wider tilt range produced a smaller slope, which means a lower efficiency of processing. Note, however, that the results of this experiment for subject ZP were not simply a replication of the results of the previous experiment. The mean slope for the case of the wide tilt range in this experiment (4.79) was quite large as compared to the mean slope (1.75) in the corresponding condition from the first experiment, namely, a narrow slant range and a wide tilt range. This difference in mean slopes was significant [$F(1,6) = 25.5$, $P < 0.0025$]. The difference in RT intercepts between these two conditions was not significant ($P > 0.1$). This means that for the moderate slants, the wide tilt range was less harmful, which makes ZP's result analogous to the result of subject JE. Thus, results of subject ZP also confirm the predictions of Perspective Invariants Theory.

To summarize, it was shown in two experiments that the results of human subjects matching shapes from perspective images agree qualitatively with predictions of the Perspective Invariants Theory. It is possible, and desirable, however, to make a direct, quantitative comparison between the theory and these experiments by performing simulations and comparing the accuracy predicted by the theory with the accuracy observed in the experiments. Results of such simulations are described below.

Results of Simulation

In the simulation, Pizlo and Rosenfeld's (1992) algorithm was used to match the same quadrilaterals as were used in the psychophysical experiments. On each trial the algorithm compared the comparison shape with the standard shape and the standard deviation, SD, of ρ was calculated. A small SD meant that the algorithm "matched" the comparison and standard shapes. A large SD meant the opposite. The criterion for a "large" SD was chosen for each condition to maximize the proportion of correct matches. The same conditions were used as in the psychophysical experiments (400 quadrilaterals were used per condition). In all conditions, it was assumed that recognition is based on a single comparison of the standard and comparison shapes with tilt equal to zero (this was the mean value of tilts in each condition in the human experiments), It was also assumed in the simulation, that the starting point was known. In the psychophysical experiment, this assumption was likely to be correct, i.e. the subjects knew the correspondence between the vertices of the standard and comparison shapes because the orientation of the comparison shape on the screen with respect to the orientation of the standard shape did not change.

Figure 8 shows the proportions of correct responses obtained in the simulation in the first experiment. For comparison, this figure also shows proportions of

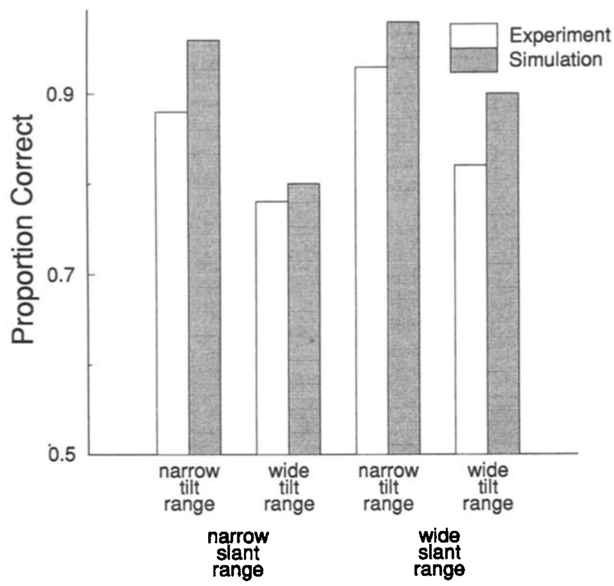

FIGURE 8. Comparison of the simulation and experimental accuracy in the Experiment 1 (see text for more details).

correct responses averaged from all subjects in the block of trials corresponding to the longest RT deadline (the longest RT deadline usually produced the best performance). It can be seen that the simulation produced the same pattern of results as was observed in the psychophysical experiment. Moreover, the level of performance was similar in both, although the algorithm did better than the human subjects in all conditions. This, however, is not surprising. The algorithm's performance was calculated for the *optimal* criterion and for a criterion that was perfectly stable, whereas in the human subject, the criterion was not likely to be either optimal or stable.

Consider now an alternative way of comparing the simulation and experimental performances. Figure 9 shows the relationship between the slope of the SATF,

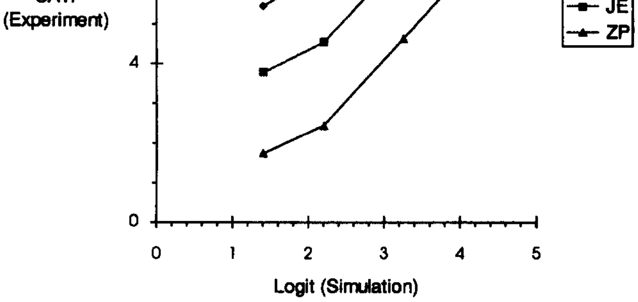

FIGURE 9. The relation between slopes of the SATFs estimated in the psychophysical experiment for individual subjects and the accuracy of the simulation performance.

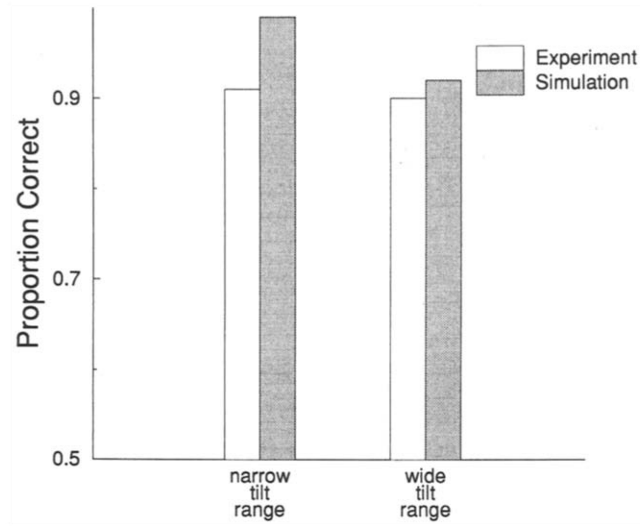

FIGURE 10. Comparison of the simulation and experimental accuracy in the Experiment 2 (see text for more details).

which was found to be an adequate measure of the subject's overall performance and the logit which represents accuracy in the simulation. The three sets of points on the graph show results for the three subjects in four experimental conditions [in the case of JE the experimental point identified as an outlier in the third replication of the first experiment was not used (see footnote on p. 1652)]. This graph shows that there is a clear relationship between simulation and experimental performance for all three subjects, namely, experimental conditions, which produced better simulation performance (higher logit), produced proportionally better psychophysical performance (the slope of the SATF is proportionally steeper).

Consider now the simulation of the second experiment. Figure 10 shows the proportions of correct responses from both this simulation and the psychophysical experiment. It can be seen that the psychophysical and simulation performance are similar, although the agreement is not as close as for the first experiment because the subjects did not obtain the very high level of performance predicted by the model. This relatively worse subjects' performance could be produced by the fact that for small slants using perspective cues might seem to be less important because the perspective image of a given shape was not very different from this shape. As a result, the narrow tilt range was not very beneficial and did not give rise to very high performance.

SUMMARY AND CONCLUSIONS

The phenomenon of shape constancy was considered from historical, theoretical and empirical perspective in this paper. First, it was shown, by analysis of the prior literature that none of the three traditional theories (Helmholtzian, Gestalt and Gibsonian) could explain shape constancy and that existing experimental results were consistent only with the multiple view theory. It was pointed out, however, that the multiple view theory cannot explain a simple observation in which a novel shape can be recognized from its single image. A new theory was then proposed, which is based on new perspective invariants of a flat shape, to account for the prior results and also for this observation. This new theory is different from Helmholtzian and Gestalt theories in that it is based only on the internal representation of shape; it does not involve depth cues. This new theory is also different from Gibsonian theory in that it is based on new types of invariants, namely, perspective invariants of shape. These invariant shape properties, along with a computer algorithm for shape recognition based on these properties, have been formulated recently by Pizlo and Rosenfeld (1992). Finally, this new theory is different from the multiple view theory in that it predicts that shape constancy does not require memorizing large number of perspective views.

The Perspective Invariants Theory is the first that can account for both the prior and my new results on shape constancy, where: (1) the percept of a flat shape was not affected by the absence of depth cues, (2) shapes that were equivalent under projective and affine transformations were not equivalent perceptually, (3) a novel shape can be recognized from its single image, (4) the perceptual processing of shape is impaired when the range of possible tilt values is wider, (5) the perceptual processing of shape is not affected by the width of the range of possible slant values, and (6) shape processing is affected by the value of slant in the presence of variations of tilt in that the perceptual processing of shape is less efficient when slants are large.

It was pointed out in this paper that perspective invariants calculated from a single image can only be used in the case of flat shapes or solid shapes that have flat or nearly flat faces. They cannot be used in the case of arbitrary objects like 3D wire shapes because there are no general case invariants for the transformation from 3D to 2D. This, in fact, is consistent with prior experimental results because it has been shown that it is difficult to achieve shape constancy in the case of 3D wire shapes (Rock & DiVita, 1987; Rock *et al.*, 1989; Edelman & Bulthoff, 1992). However, shape constancy can be achieved if extensive training is provided, sufficient for the subject to memorize a large number of 2D views of a given object. This means that Perspective Invariants Theory is not a completely general theory of shape constancy. In the case of flat shapes or solid shapes that have flat faces, performance can be explained satisfactorily by invariants as developed in the Perspective Invariants Theory, and it seems that other mechanisms need not be invoked in this case. It is possible, however, that other mechanisms can contribute, to some extent to achieving shape constancy, but their nature and the degree of their contribution is not known at this time. In the case of those rare shapes that do not have any flat or nearly flat faces, however, performance can be explained *only* by storing many different views of the shape (the multiple view theory); invariants cannot explain performance. It seems, however, that although a human observer has the capacity to memorize many views of a single object (as demonstrated in the studies cited just above), this mechanism is not likley to be used in everyday life because of its limited efficiency (i.e. a large number of views are required for each object) and its limited utility (objects without any flat or nearly flat faces are exceedingly rare). Verification of this claim requires performing psychophysical experiments testing human shape perception for objects selected as representative of objects contained in our everyday life environment. It can be expected that our everyday life environment imposes some constraints that can be crucial in deciding which mechanism of shape perception is more useful.* Such psychophysical research to determine the ecological value of natural shapes should be supplemented by simulation studies that will establish whether invariants of flat shapes can, in fact, lead to the successful recognition of such objects. This kind of future research on human beings may help the computer scientist to formulate an efficient algorithm for machine and robot vision applications as well as advance our understanding of human perception.

*This kind of ecologically valid test for the case of size constancy was performed many years ago by Brunswik (1955). At the time it was commonly accepted that size constancy involved depth perception (Holway & Boring, 1941), but Brunswik showed that the environmental constraints, which are present in our everyday life, can account for as much as 50% of the variability of the subject's responses. This result implies that environmental constraints significantly contribute to size constancy.

REFERENCES

Aloimonos, J. (Y.) (1988). Shape from texture. *Biological Cybernetics, 58*, 345–360.

Ashton, W. D. (1972). *The logit transformation*. New York: Hafner.

Ballard, D. H. & Brown, C. M. (1982). *Computer vision*. Englewood Cliffs, N.J.: Prentice-Hall.

Beck, J. & Gibson, J. J. (1955). The relation of apparent shape to apparent slant in the perception of objects. *Journal of Experimental Psychology, 50*, 125–133.

Ben-Arie, J. (1990). The probabilistic peaking effect of viewed angles and distances with application to 3-D object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12*, 760–774.

Bevington, P. R. (1969). *Data reduction and error analysis for the physical sciences*. New York: McGraw-Hill.

Braunstein, M. L. (1976). *Depth perception through motion*. New York: Academic Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62*, 193–217.

Burns, J. B., Weiss, R. & Riseman, E. M. (1990). View variation of point set and line segment features. *Proceedings of DARPA Image Understanding Workshop*, (pp. 650–659).

Cassirer, E. (1944). The concept of group and the theory of perception. *Philosophy and Phenomenological Research, 5*, 1–35.

Coolidge, J. L. (1940). *A history of geometrical methods*. Oxford University Press.

Cutting, J. E. (1986). *Perception with an eye for motion*. Cambridge, Mass.: MIT Press.

Dosher, B. A., Landy, M. S. & Sperling, G. (1989). Kinetic depth effect and optic flow—I. 3D shape from Fourier motion. *Vision Research, 29*, 1789–1813.

Duda, R. O. & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: Wiley.

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

Epstein, W., Bontrager, H. & Park, J. (1962). The induction of nonveridical slant and the perception of shape. *Journal of Experimental Psychology, 63,* 472–479.

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

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

Gottheil, E. & Bitterman, M. E. (1951). The measurement of shape-constancy. *American Journal of Psychology, 64,* 406–408.

Grimson, W. E. L., Horn, B. K. P., Poggio, T. & staff (1992). Progress in image understanding at M.I.T. *Proceedings of the Image Understanding Workshop* (pp. 69–82).

Hochberg, J. (1971). Perception. In Kling, J. W. & Riggs, L. A. (Eds), *Experimental psychology* (pp. 395–550). New York: Holt, Rinehart & Winston.

Hoffman, W. C. (1966). The Lie algebra of visual perception. *Journal of Mathematical Psychology, 3,* 65–98.

Holway, A. H. & Boring, E. G. (1941). Determinants of apparent visual size with distance variant. *American Journal of Psychology, 54,* 21–37.

Ikeuchi, K. (1984). Shape from regular patterns. *Artificial Intelligence, 22,* 49–75.

Ikeuchi, K. & Horn, B. K. P. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence, 17,* 141–184.

Jennings, J. R., Wood, C. & Lawrence, B. E. (1975). Effects of graded doses of alcohol on speed-accuracy tradeoff in choice reaction time. *Perception and Psychophysics, 19,* 85–91.

Johansson, G. (1977). Spatial constancy and motion in visual perception. In Epstein, W. (Ed.), *Stability and constancy in visual perception: Mechanisms and processes,* (pp. 375–419). New York: Wiley.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research, 31,* 1351–1360.

Kaiser, P. K. (1967). Perceived shape and its dependency on perceived slant. *Journal of Experimental Psychology, 75,* 345–353.

Kanade, T. & Kender, J. R. (1983). Mapping image properties into shape constraints: Skewed symmetry, affine-transformable patterns, and the shape-from-texture paradigm. In Beck, J., Hope, B. & Rosenfeld, A. (Eds), *Human and machine vision* (pp. 237–257). New York: Academic Press.

Klein, F. (1939). *Elementary mathematics from an advanced standpoint. Geometry.* New York: Dover.

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

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

Lamdan, Y., Schwartz, J. T. & Wolfson, H. J. (1988). Object recognition by affine invariant matching. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 335–344).

Lehky, S. R. & Sejnowski, T. J. (1990). Neural network model of visual cortex for determining surface curvature from images of shaded surfaces. *Proceedings of the Royal Society of London B, 240,* 251–278.

Leibowitz, H. W. & Bourne, L. E. Jr (1956). Time and intensity as determiners of perceived shape. *Journal of Experimental Psychology, 51,* 277–281.

Leibowitz, H. W., Wilcox, S. B. & Post, R. B. (1978). The effect of refractive error on size constancy and shape constancy. *Perception, 7,* 557–562.

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature, London, 293,* 133–135.

Longuet-Higgins, H. C. & Prazdny, K. (1980). The interpretation of a moving retinal image. *Proceedings of the Royal Society of London B, 208,* 385–397.

Luce, R. D. (1986). *Response times.* Oxford University Press.

Marr, D. (1982). *Vision.* New York: Freeman.

McKee, S. P., Levi, D. M. & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research, 30,* 1763–1779.

Meneghini, K. A. & Leibowitz, H. W. (1967). Effect of stimulus distance and age on shape constancy. *Journal of Experimental Psychology, 74,* 241–248.

Menzer, G. W. & Thurmond, J. B. (1970). Form identification in peripheral vision. *Perception and Psychophysics, 8,* 205–209.

Mundy, J. L. & Zisserman, A. (1992). Introduction—towards a new framework for vision. In Mundy, J. L. & Zisserman, A. (Eds), *Geometric invariance in computer vision* (pp. 1–39). Cambridge, Mass.: MIT Press.

Pentland, A. P. (1986). Local shading analysis. In Pentland, A. P. (Ed.), *From pixels to predicates* (pp. 40–77). N.J.: Ablex.

Perrone, J. A. (1980). Slant underestimation: A model based on the size of the viewing aperture. *Perception, 9,* 285–302.

Perrone, J. A. (1982). Visual slant underestimation: A general model. *Perception, 11,* 641–654.

Pizlo, Z. (1993). Geometry of visual space. *Investigative Ophthalmology and Visual Science, 34,* 708 (Abstract).

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.

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

Rock, I. & DiVita, J. (1987) A case of viewer-centered object perception. *Cognitive Psychology, 19,* 280–293.

Rock, I., Wheeler, D. & Tudor, L. (1989). Can we imagine how objects look from other viewpoints? *Cognitive Psychology, 21,* 185–210.

Rothwell, C. A., Forsyth, D. A., Zisserman, A. & Mundy, J. L. (1993). Extracting projective structure from single perspective views of 3D point sets. *Proceedings of the IEEE International Conference on Computer Vision,* (pp. 573–582).

Salach-Golyska, M., Pizlo, Z., Vyain, S. & Standish, J. (1993). 3D shape perception. *Investigative Ophthalmology and Visual Science, 34,* 1082 (Abstract).

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

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

Sperling, G. (1971). The description and luminous calibration of cathode ray oscilloscope visual displays. *Behavior Research Methods and Instruments, 3,* 148–151.

Sperling, G., Landy, M. S., Dosher, B. A. & Perkins, M. E. (1989). Kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 826–840.

Stavrianos, B. K. (1945). The relation of shape perception to explicit judgments of inclination. *Archives of Psychology, No.296,* 1–94.

Thouless, R. H. (1931a). Phenomenal regression to the real object. I. *British Journal of Psychology, 21,* 339–359.

Thouless, R. H. (1931b). 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.

Toussaint, G. T. (1988). Preface. In Toussaint, G. T. (Ed.), *Computational morphology,* (pp. vii–xii). Amsterdam: North Holland.

Todd, J. T. & Akerstrom, R. A. (1987). Perception of three-dimensional form from patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 242–255.

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

Todd, J. T. & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception and Psychophysics, 50,* 509–523.

Todd, J. T. & Reichel, F. D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved shapes. *Psychological Review, 96,* 643–657.

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

Ullman, S. (1983). Recent computational studies in the interpretation of structure from motion. In Beck, J., Hope, B. & Rosenfeld, A. (Eds), *Human and machine vision* (pp. 459–480). New York: Academic Press.

Wallach, H. & Moore, M. E. (1962). The role of slant in the perception of shape. *American Journal of Psychology*, *75*, 289–293.

Weiss, I. (1988). Projective invariants of shapes. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (pp. 291–297).

Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence*, *17*, 17–45.

Zusne, L. (1970). *Visual perception of form*. New York: Academic Press.

Acknowledgements—I am very grateful to Azriel Rosenfeld, Robert Steinman and Isaac Weiss for many discussions, to Eileen Kowler, Robert Melara, Robert Steinman and anonymous reviewers for comments on earlier drafts of this paper and to Julie Epelboim for collaboration on a number of aspects of this research. A part of this research was presented at an annual of the *Association for Research in Vision and Ophthalmology*, in Sarasota, 1992. This research was supported, in part, by grant 91-0124 from the Life Sciences Directorate of the AFOSR to R. M. Steinman.