**Perception**, **27**, 273-282

# On the Relative Salience of Euclidean, Affine and Topological Structure for 3D Form Discrimination

James T. Todd[1], Lin Chen[2], J. Farley Norman[3]

1: The Ohio State University
2: University of Science and Technology of China
3: Western Kentucky University

A match-to-sample task was performed, in which observers compared configurations of line segments presented stereoscopically in different 3D orientations. Several different structural properties of these configurations were manipulated, including the relative orientations of line segments (a Euclidean property), their co-planarity (an affine property) and their patterns of co-intersection (a topological property). Although the differences in these properties to be detected were all metrically equivalent, they varied dramatically in their relative perceptual salience, such that the error rates and reaction times in the three conditions varied by as much as 400%. Performance was highest in the topological condition, intermediate in the affine condition, and lowest in the Euclidean condition. These findings suggest that the relative perceptual salience of object properties may be systematically related to their structural stability under change, in a manner that is similar to the Klein hierarchy of geometries.

## Introduction

One of the most perplexing phenomena in the study of human vision is the ability of observers to perceive the 3D layout of the environment from patterns of light that project onto the retina. There are many different aspects of optical stimulation, such as shading, texture, motion and binocular disparity, that are known to provide perceptually salient information about 3D structure, but an effective computational analysis of this information has proven to be surprisingly elusive. One possible reason for this, we suspect, is that it is difficult to identify the specific aspects of an object's structure that form the primitive components of an observer's perceptual knowledge. After all, in order to compute shape, it is first necessary to define what "shape" is.

One important factor in evaluating potential primitives for the perceptual representation of 3D form is their relative stability. When an object is transformed in the natural environment, it is generally the case that only some of its properties will be altered, while others remain invariant. Consider, for example, some possible transformations of the diamond shaped object at the top of Figure 1. At the left of the middle row is a Euclidean transformation of this figure that was created by rotating it to a different orientation in the image plane. Note that the absolute orientations of all its line segments are altered, but that their lengths remain unchanged. If the same object is reduced in size by a similarity transformation, its line lengths will all be diminished, but its angles will be unaffected. If the object is subjected to an affine stretching transformation, the relative lengths and angles of its line segments will be distorted, but lines that are parallel will remain so. If the original object is rotated in depth, then the shadow it casts in the fronto-parallel plane will be deformed by a projective transformation. The parallelism of its opposing edges will be destroyed in that case, but they will continue to be straight lines. If the object were made of an elastic material and was stretched into the shape of an ice cream cone, then straight lines could become curved, but the number of bounded holes in the figure would remain constant. Finally, if one of its edges were cut with a pair of scissors, then the number of holes could be altered as well.

While considering the phenomenon of invariance under change, it is interesting to note its historical importance to the development of modern geometry. In 1872, the German mathematician Felix Klein gave a lecture at Erlangen University, in which he outlined a general principle for constructing different geometries that is now known as the Erlanger Programm. His basic idea was to

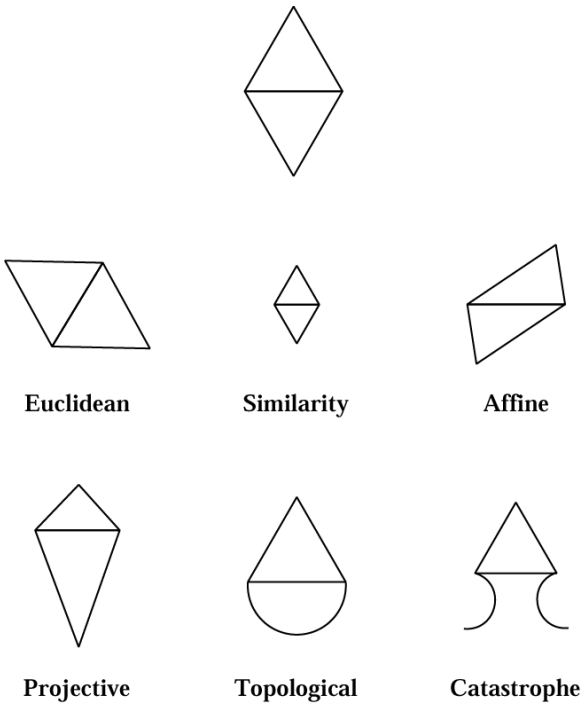


**Figure 1 --** Some possible transformations of a diamond shaped object. Note in each case that some of its structural properties are altered, while others remain invariant.

consider arbitrary groups of single valued transformations, and to investigate the properties of objects they leave invariant. Using this principle, it is possible to build a hierarchy of geometries (i.e., Euclidean, affine, conformal, etc.) in which structural properties can be stratified with respect to their stability in a formally precise way.

There is some evidence to suggest that a similar type of stratification may occur in human perception. For example, Chen (1983, 1989) has reported that the relative salience of different geometric properties in a texture segregation task is remarkably consistent with the hierarchy of geometries in the Klein Erlanger programm. Some representative stimulus patterns from his study are shown in Figure 2. When the elements in the segregated region were topologically distinct from those in the remaining portions of the pattern, observers could identify the disparate quadrant with a mean reaction time of only 801 msec. When the disparate quadrant was defined by a projective property (i.e., co-linearity) or an affine property (i.e., parallelism) the reaction times increased to 968 msec and 1465 msec, respectively. Finally, when the disparate region could only be distinguished by the Euclidean property of relative orientation, the reaction times increased still further to 1941 msec.

The invariance of figural properties under change is also an important factor in the ability of human observers to recognize objects in different 3D orientations. For example, in a recent experiment by Biederman and Bar (1999), observers were presented with pairs of objects presented in sequence, and were asked to judge as quickly as possible whether their 3D structures were the same or different. Two different types of manipulation were performed to create pairs of objects that were structurally different. On half of these trials, one object was created from the other by altering properties that are viewpoint invariant. These included the co-terminations of its contours (a topological property), whether its contours were straight or curved (a projective property) or whether they were parallel to one another (an affine property). On the remaining trials, one object was created from the other by altering the lengths or curvatures of its contours without affecting their co-terminations, linearity or parallelism. The two sets of foils were carefully matched so that they were equally detectable when both objects were presented at the same 3D orientation. When they were presented at different orientations, however,

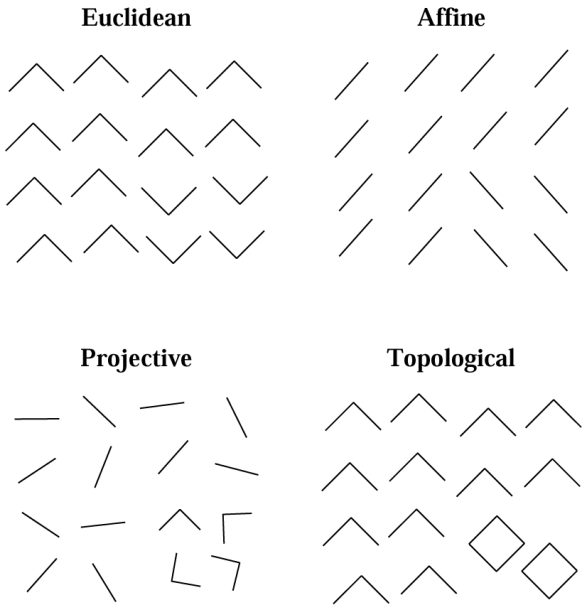


**Figure 2 --** Some example stimulus patterns used by Chen (1983, 1989).

the observers' responses were significantly more accurate and had shorter latencies for those pairs that differed in their viewpoint invariant structure (see also Cooper & Biederman, 1993; Liu, Knill & Kersten, 1995).

The linearity, parallelism and co-terminations of image contours are theoretically important in object recognition because of the inherent loss of information that occurs due to optical projection. The utility of these properties is that they remain largely invariant over the projective mapping from a 3D environmental structure to a 2D visual image. That is to say, if the edges of an object are straight, if they are parallel to one another, or if they co-terminate at a vertex, then the projected images of those edges will exhibit the same characteristics. Although it is mathematically possible to adopt a viewpoint where curved edges appear straight, nonparallel contours appear parallel, or unconnected edges appear to co-terminate, the probability of encountering such a viewpoint by accident in a natural environment is vanishingly small. Thus, they are often referred to as non-accidental properties (Biederman, 1987; Lowe, 1987).

There are similar theoretical issues that arise in the perceptual analysis of 3D structure from motion or binocular disparity. Unlike the case of static monocular vision, where physical space (x, y, z) is projectively mapped onto a 2D visual image (x', y'), additional dimensions are required to adequately represent the structure of moving or stereoscopic images. One way of conceptualizing this higher order structure is that physical space (x, y, z) is optically transformed into a 3D image space (x', y', z'), where z' represents projected velocity or binocular disparity. Although the domain and range of this mapping may have commensurate dimensionalities, it is important to keep in mind that the relation between them is non-Euclidean. Some aspects of 3D structure will be systematically distorted by this transformation, while others will remain invariant, and these latter properties could be especially useful as potential sources of information for the perceptual analysis of 3D form.

There have been several experiments reported in the literature that have compared sensitivity to various aspects of 3D structure for objects depicted with motion or binocular disparity, including Euclidean properties, such as line lengths, and affine properties, such as co-planarity (e.g., see Todd & Bressan, 1990; Tittle, Todd, Perotti & Norman, 1995). There is, however, a problem in interpreting the results of such experiments. In order to compare perceptual sensitivity on such disparate tasks, it is necessary to devise some form of common currency for evaluating observers' performance. One way of addressing this issue that has been employed to investigate the relative salience of various properties in static monocular images is to carefully match the variations in those properties so that they are all metrically equivalent (e.g., see Biederman & Bar, 1999; Cooper & Biederman, 1993). The research described in the present article used a similar technique to examine the importance of invariance under change for the binocular form perception of configurations of line segments in 3D space. Our goal was to compare the relative perceptual salience of several different structural properties, including the relative orientations of line segments (a Euclidean property), their co-planarity (an affine property) and their patterns of co-intersection (a topological property).

## Methods

### Apparatus

The experiment was performed using a Silicon Graphics Indigo Extreme workstation with stereoscopic viewing hardware. The displays were viewed through LCD (liquid crystal) shuttered glasses that were synchronized with the monitor's refresh rate. The different views of a stereo pair were displayed at the same position on the monitor screen, but they were temporally offset. The left and right lenses of the LCD glasses shuttered synchronously with the display at an alternation rate of 60 Hz, so that each view could only be seen by the appropriate eye. The spatial resolution of the monitor was 1280 X 1024 pixels, which subtended 25.2 by 20.3 degrees of visual angle when viewed at a distance of 76 cm.

### Stimuli

Each stimulus display contained a triangular arrangement of three wire-frame objects (see Figure 3), all of which contained four connected line segments. The upper object in the triangular configuration was designated as the standard, and the two lower objects were designated as test figures. The projected images of these objects on the

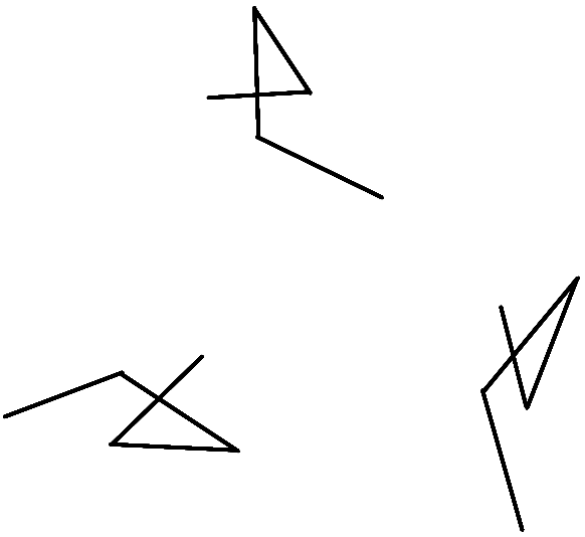

**Figure 3 --** An example stimulus configuration with a standard and two possible test objects.

plane of the display screen all had the same 2D topology as shown in Figure 4. The four line segments – labeled in the figure as *a*, *b*, *c*, and *d* -- were oriented in three-dimensional space so that *b* was connected to *a* and *c*; *c* was connected to *b* and *d*; and the projected images of *b* and *d* intersected one another. In addition, the angle between adjacent segments and between each segment and the line of sight was always greater than 25°. Within those constraints, the lengths and angles of the standard object were generated at random on each trial.

One of the test figures we shall refer to as the target had a 3D Euclidean structure that was identical to the standard. The other, which we shall refer to as the foil, had three line segments that were identical to the standard, but the fourth segment was bent by 40º relative to its corresponding segment in the standard. That is to say, the two test figures were structurally identical except for one pair of corresponding segments that differed in orientation by 40º. Both test figures were presented at randomly selected 3D orientations (subject to the constraints described above), so that the projected 2D Euclidean structure in the image plane would be different for all three objects. The observer's task on each trial was to indicate which of the two test figures had the same 3D structure as the standard.

There were three distinct experimental conditions, in which we manipulated the topological, affine and Euclidean relations between the targets and the foils. In the topological condition, the foil was created from the standard by bending line segment *d*, so that the two test figures had different 3D topologies (see Figure 4). On half the trials, the standard was constructed such that *d* intersected *b* in 3-dimensional space, and the foil was structured such that *d* was slanted by 40º relative to the plane defined by *b* and *c*. On the remaining trials this relationship was reversed. That is to say, *b* intersected *d* in the foil but not in the standard.

In the affine condition, the foil was created from the standard by bending line segment *a* so that the two test figures differed in the affine property of planarity. Segments *b*, *c*, and *d* were always coplanar in this condition. On half the trials, the standard was constructed such that a was coplanar with the other segments, and the foil was structured such that *a* was slanted by 40º relative to the plane defined by *b*, *c* and *d*. On the remaining trials this relationship was reversed – i.e., all of the segments were coplanar in the foil but not in the standard.

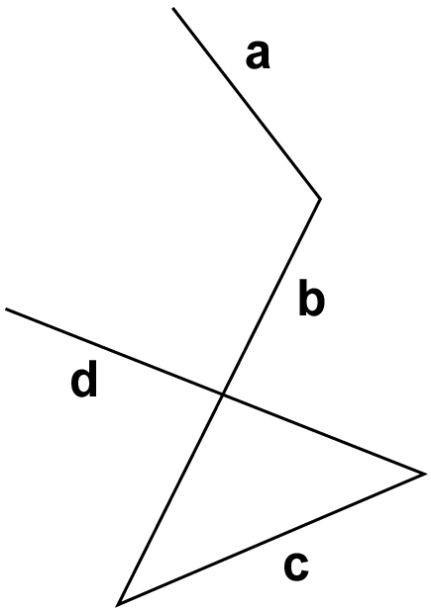


**Figure 4** -- Each standard and test object contained four connected line segments (a, b, c, and d) and their projected images all had the same 2D topology.

Finally, in the Euclidean condition, the foil was again created from the standard by bending line segment *a* 40° relative to the plane defined by *b* and *c*. In this case, however, there were no constraints on planarity or whether *b* intersected *d* in 3-dimensional space. Some example objects from these different conditions are shown in Figure 5. The upper and lower stereograms of this figure show a possible standard object and target at different 3D orientations. The middle three stereograms depict potential foils for this object in the topological, affine and Euclidean conditions. Note that it is virtually impossible to distinguish these objects from their 2D projections in the image plane, but that they are relatively easy to discriminate when viewed stereoscopically.

**Procedure**

At the beginning of each trial, observers pressed a key to initiate the presentation of a standard and two test objects. They were instructed to examine all three objects, and to indicate which of the test figures had the same 3D structure as the standard by pressing the left or right button on a hand held mouse. They were informed that the accuracy and reaction time of each response would be recorded, and they were asked to make their judgments as quickly and accurately as possible. To provide immediate feedback after every trial, an auditory beep was presented after every correct response.

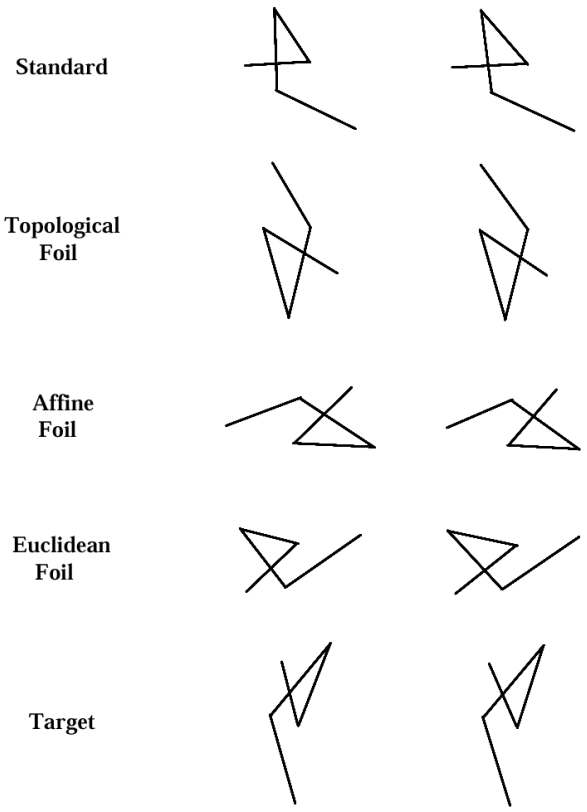


**Figure 5** -- An example stereogram of a possible standard object, a target, and three possible foils from the topological, affine and Euclidean conditions.

The topological, affine and Euclidean conditions were run in separate blocks in a randomly determined order. Each block contained 150 trials, the first 50 of which were considered as practice and excluded from subsequent analyses. To make the task easier, observers were told at the beginning of each condition which line segment – *a* or *d* – would distinguish the foils from the standard.

**Observers**

The observers included ten volunteers from the students and staff at the Ohio State University and the University of Science and Technology of China. All had normal or corrected to normal vision.

**Results**

Several of the observers reported spontaneously during their debriefings that the difficulty of the task varied dramatically across the three different conditions, and this was confirmed in subsequent analyses of their accuracy and reaction times. Figure 6 shows the percentage of correct responses in the Euclidean, affine and topological conditions averaged over all 10 observers. Note that the observers were almost perfectly accurate in the topological condition, with a mean error rate of only 4%. This increased to 11% in the affine condition, and jumped to over 21% in the Euclidean condition. Because the distribution of errors did not pass a normality test, the data were analyzed using a Friedman repeated measures analysis of variance on ranks. The results revealed that performance was significantly different across the three conditions, $\chi^2$ (2) = 18.2, $p < .0001$, and that all pair-wise comparisons were significant as well, $p < .01$.

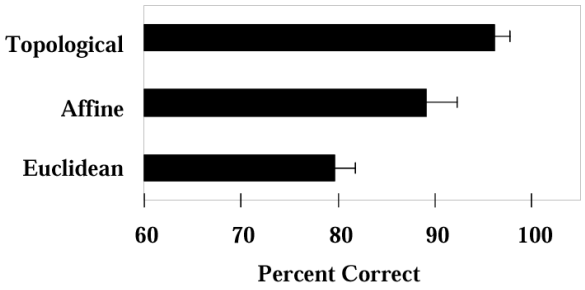


**Figure 6** -- The percentage of correct responses averaged over 10 observers in the topological, affine and Euclidean conditions.

A similar portrait of the difficulty of these tasks was revealed in the pattern of reaction times for correct responses. Figure 7 shows the mean response latencies in the different conditions averaged over all ten observers. It is important to keep in mind when considering these data that multiple eye movements were required in order to fixate the three separate objects presented on each trial, so it is not surprising that the response latencies were relatively large, ranging from one to three seconds. As is evident from the figure, the structural discriminations were easiest in the topological condition. The mean response latency in that case was only 1216 msec, or roughly 400 msec per object. The response times in the affine and Euclidean conditions were nearly two and three times larger (i.e., 2109 and 2818 msec), thus providing a clear indication of their relative difficulty. An analysis of variance revealed that the differences among these three conditions were statistically significant, $F(2,9) = 22.6$, $p < .00001$, as were all of their pairwise comparisons, $p < .01$.

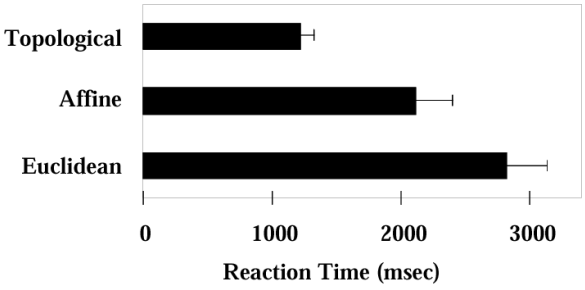


**Figure 7** -- The mean reaction time for correct responses, averaged over ten observers in the topological, affine and Euclidean conditions

## Discussion

Let us now consider some possible strategies by which observers may have performed these judgments. In order to determine which of two test objects is most closely matched to a standard it is necessary to achieve two goals: First, observers must somehow perceptually represent the structure of the standard and both test objects; and second, they must also have a way of comparing these different representations that is sufficiently powerful to detect structural similarities or differences over varying 3D orientations.

There has been considerable debate in recent years about how objects are perceptually represented within the human visual system, and much of this discussion has been focused on the issue of viewpoint invariance. One popular theoretical position is that object representations are primarily viewpoint dependent (e.g., see Edelman & Bülthoff, 1992; Tarr, 1995). According to this approach, recognition is achieved by mentally transforming one representation so that it can be matched with another. It is assumed that the duration of this process varies with the magnitude of the required transformation, and that response times and errors should increase monotonically with increasing differences in orientation between the two object viewpoints to be compared (e.g., Shepard & Metzler, 1971; Tarr & Pinker, 1989).

An alternative theoretical position that has been promoted by Biederman and his colleagues (e.g., Biederman, 1987; Hummel & Biederman, 1992) is that observers can recognize objects at varying orientations in depth by exploiting certain structural properties that are viewpoint invariant, such as the linearity, parallelism and co-termination of image contours. According to this account, the specific orientation difference between two objects to be compared should have a negligible effect on performance, provided that the set of objects to be distinguished vary in their viewpoint invariant structure, and that those structural properties are not obscured by occlusion (see Biederman & Gerhardstein, 1993).

An orthogonal issue to whether object representations are viewpoint dependent or viewpoint invariant, is whether the underlying data structure is primarily 2D or 3D. For example, within the viewpoint dependent approach, there are some alignment models in which object representations

are mentally rotated in 3D space in order to be compared with an input image (Huttenlocher & Ullman, 1987), and others in which a 2D representation of an image is transformed (Bienenstock & Von der Malsburg, 1987). Similarly, within the viewpoint invariant approach, it is possible to define features within the space of a 2D image or a higher dimensional 3D space defined by motion or binocular disparity.

Although the present experiment was designed to investigate the relative salience of viewpoint invariant 3D features, we must also consider the possibility that observers may have adopted one of the other strategies described above for determining which test object best matched the standard. It is important to keep in mind with respect to this issue that all of the displays were created so that the metrical difference between the target and the foil were identical on every trial. If appropriately oriented in 3D space, the two test figures could be perfectly aligned except for one pair of corresponding segments whose orientations would differ by 40º. Given that this was a constant factor for all of the different structural configurations, it is difficult to imagine how an alignment strategy could produce such large variations in performance among the topological, affine and Euclidean conditions. It is also important to note in this regard that the displays were all constructed so that the targets and foils could not be distinguished by their viewpoint invariant 2D features. That is to say, they all contained four non-parallel connected line segments with the same 2D image topology (see Figure 4).

Because of these various constraints, the most straightforward strategy for performing these judgments is to search for a specific 3D feature that can reliably distinguish the target from the foil, and that is the subjective impression one gets while participating in this paradigm. Depending upon the experimental condition, there were three possible features that could potentially be employed for this purpose. First, observers could estimate the specific 3D orientation of the critical line segment relative to other components of the configuration. Second, in the affine condition, they could examine whether all four line segments of a configuration were co-planar; and finally, in the topological condition, they could evaluate whether line segments *b* and *d* intersected one another in 3D space.

Although the structural deviations of these features that distinguished the targets from the foils were all metrically equivalent, they varied dramatically in their relative perceptual salience, such that the error rates and reaction times in the three conditions differed by as much as 400%. This is also confirmed by observers' subjective experiences while performing these tasks. When the targets and foils differ in their 3D topology, the task seems almost effortless, and the primary factor that limits performance is having to move one's eyes as rapidly as possible in order to fixate on the different objects. However, when targets and foils can only be distinguished by their 3D Euclidean structures, the task seems virtually impossible. Indeed, the average response time in that case is almost three seconds, with an error rate of over 20%.

These findings are consistent with a theoretical hypothesis proposed by Chen (1983, 1989), Todd & Bressan (1990) and Tittle, et. al. (1995) that the relative perceptual salience of object properties is systematically related to their structural stability under change, in a manner that is similar to the Klein hierarchy of geometries. According to this hypothesis, observers should be most sensitive to those aspects of an object's structure that remain invariant over the largest number of possible transformations. Previous evidence to support for this prediction has been largely confined to the perceptual organization of 2D patterns (Chen, 1983, 1989), but the present research extends this result to the perceptual analysis of 3D structure from binocular disparity (see also Tittle, et. al., 1995). Among the three types of 3D properties that were examined, the differences in topological structure were easier to recognize than differences in affine structure, but performance was even worse when the task could only be performed based on differences in Euclidean structure.

One likely reason why the human visual system might be biased toward object properties that are invariant under change is to facilitate shape constancy. There are a number of different problems in natural vision for which this approach might be particularly useful. Perhaps the most obvious of these is the inevitable distortion that occurs when a 3D structure is optically projected onto a 2D visual image. Such distortions should be irrelevant, however, if a task can be performed based on non-accidental properties that are invari-

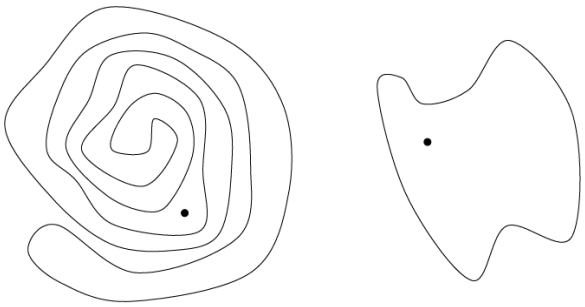

**Figure 8 --** The difficulty of judging whether a probe dot is inside or outside of a closed boundary can be significantly influenced by its structural complexity.

ant over projective transformations (e.g., see Biederman & Bar, 1999; Biederman, 1987; Biederman & Gerhardstein, 1993). Other distortions can occur in the mathematical relationship between the physical environment and stereoscopic space, and objects themselves can also be distorted. The human body provides many good examples, such as the bending of the arms and legs during locomotion, or changes in facial expression. Object properties that remain invariant under these changes are a potentially rich source of information, which we suspect may be essential for the perceptual experience of environmental stability.

Although the Klein hierarchy of geometries provides a useful framework for assessing the invariance of object properties to various types of change, it is best not to take it too literally as a model of human perception. Suppose, for example, that an observer is asked to judge the topological property of whether a small probe dot is inside or outside a closed boundary (see Ullman, 1996). For many possible boundary shapes, such as the one shown in the right portion of Figure 8, this type of judgment seems subjectively trivial and automatic. If, however, the boundary is made sufficiently convoluted as in the left portion of Figure 8, then the task can become quite difficult – even though the two figures are topologically equivalent. It is likely in this case that performance is limited by the need to integrate local topological relations over an extended region of visual space. For other types of topological judgments that do not impose this requirement, such as determining if a dot is on or off the boundary, the global complexity of the figure will have no effect on performance.

**References**


Biederman, I. & Bar, M. (1999) One-shot viewpoint invariance in matching novel objects. Vision Research, 39, 2885-2900

Biederman, I. (1987) Recognition-by-components: A theory of human image interpretation. Psychological Review, 94, 115-147.

Biederman, I. & Gerhardstein, P. C. (1993) Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. Journal of Experimental Psychology: Human Perception and Performance, 19, 1162-1182.

Bienenstock, E. & Von der Malsburg, C. (1987) A neural network for invariant pattern recognition. Europhysics Letters, 4, 121-126.

Chen, L. (1983) What are the units of figure perceptual representation? (Studies in Cognitive Science No. 22). Irvine, CA: University of California, School of Social Sciences.

Chen, L. (1983) What are the units of figure perceptual representation? (Studies in Cognitive Science No. 22). Irvine, CA: University of California, Irvine, School of Social Sciences.

Chen, L. (1989) Topological perception: A challenge to computational approaches to vision. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulie & L. Steels (Eds.), Connectionism in Perspective (pp.317-329). Amsterdam: Elsevier.

Cooper, E. E. & Biederman, I. (1993) Geon differences during recognition are more salient than metric differences. Poster presented at the meeting of the Psychonomic Society, Washington, D. C.

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

Hummel, J. E. & Biederman, I. (1992) Dynamic binding in a neural network for shape recognition. Psychological Review, 99, 480-517.

Huttenlocher, D. P. & Ullman, S. (1987) Object recognition using alignment. In Proceedings of the International Conference on Computer Vision (pp. 102-111). London: IEEE.

Liu, Z., Knill, D. C. & Kersten, D. (1995) Object classification for human and ideal observers. Vision Research, 35, 549-568.

Lowe, D. G. (1987) The viewpoint consistency constraint. International Journal of Computer Vision, 1, 57-72.

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

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

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

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995) The systematic distortion of perceived 3D structure from motion and binocular stereopsis. Journal of Experimental Psychology: Human Perception and Performance, 21, 663-678.

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

Ullman, S. (1996) High level vision: Object recognition and visual cognition. Cambridge, MA: MIT Press.

**Acknowledgments**

This collaboration was supported in part by the United Nations Development program in China and the National Natural Science Foundation of China. James Todd and Farley Norman were also supported by the Air Force Office of Scientific Research (AFOSR grant #F49620-93-1-0116) and the National Science Foundation (SBR-9514522); Correspondence should be addressed to James T. Todd, Department of Psychology, 142 Townshend Hall, The Ohio State University, Columbus, OH 43210.