

# Ordinal Structure in the Visual Perception and Cognition of Smoothly Curved Surfaces

James T. Todd and Francene D. Reichel  
Brandeis University, and  
Center for Adaptive Systems, Boston University

In theoretical analyses of visual form perception, it is often assumed that the 3-dimensional structures of smoothly curved surfaces are perceptually represented as point-by-point mappings of metric depth and/or orientation relative to the observer. This article describes an alternative theory in which it is argued that our visual knowledge of smoothly curved surfaces can also be defined in terms of local, nonmetric order relations. A fundamental prediction of this analysis is that relative depth judgments between any two surface regions should be dramatically influenced by the monotonicity of depth change (or lack of it) along the intervening portions of the surface through which they are separated. This prediction is confirmed in a series of experiments using surfaces depicted with either shading or texture. Additional experiments are reported, moreover, that demonstrate that smooth occlusion contours are a primary source of information about the ordinal structure of a surface and that the depth extrema in between contours can be optically specified by differences in luminance at the points of occlusion.

For many higher organisms, including humans, a primary source of knowledge about objects and events in the surrounding environment is provided by vision. Because of the ecological importance of this knowledge, the processes of visual perception and cognition have been attracting a growing amount of attention in recent years from researchers in a variety of fields. There are two basic issues on which this research has primarily been focused: (a) identification of the relevant parameters of representation for visually based knowledge, and (b) determination of the specific sources of optical information from which it is obtained.

Consider, for example, a recent theory of object recognition by Biederman (1987) that addresses both of these issues. According to Biederman, most objects in the environment can be adequately represented to achieve recognition using a limited set of volumetric primitives, called *geons*, which are connected to one another in simple combinations, in much the same way that words can be composed from a relatively small alphabet of phonemes. The optical information from which these geons are perceptually specified is assumed to be based on easily measurable properties of image contours, such as the presence or absence of curvature, parallelism, or symmetry, and the coterminalization of contours at vertices.

Biederman (1987) has pointed out that his analysis is re-

stricted to concrete objects that have clearly defined volumetric structures (e.g., lamps, airplanes, or elephants), but it is important to keep in mind that there are some objects encountered in nature that do not conform to this restriction. There are, in particular, some continuous smoothly curved surfaces, such as landscapes or patterns of rumpled cloth, that are not well suited to a volumetric description. Although observers seem to have little difficulty perceiving the three-dimensional structures of these surfaces, the precise nature of their perceptual knowledge has yet to be determined.

## Local Orientation Maps

The first systematic analysis of the visual perception and cognition of smoothly curved surfaces was presented almost 40 years ago by Gibson (1950). For parameters of representation, Gibson proposed that our visual knowledge about the three-dimensional structure of curved surfaces can be described as a point-by-point mapping of depth and orientation for each local surface region within the field of view (see, however, Gibson, 1979). This same type of representation was adopted many years later by researchers in machine vision. Horn (1975, 1977), for example, in his pioneering work on *shape from shading*, defined *shape* as a local orientation map (see also Lee & Rosenfeld, 1985; Pentland, 1984; Stevens, 1981a, 1983; Witkin, 1981). A similar approach was adopted by Marr (1982) and Marr and Nishihara (1978). They argued that these local mappings of depth and orientation form the representational basis of our immediate perceptual awareness, and they named this representation the  $2\frac{1}{2}D$  sketch.

It is important to recognize that these theoretical analyses of the structure of visual knowledge have important methodological implications for empirical investigations of three-dimensional form perception. In any psychophysical experiment, an observer must be asked some question about the perceived

---

This research was supported in part by Air Force Office of Scientific Research Grant F49620-87-C-0018 and joint Grant BNS-8420143 from the Office of Naval Research and the National Science Foundation. The participation of Francene D. Reichel was also supported under a National Science Foundation Graduate Fellowship.

We are most grateful to Joe Lappin, Hal Sedgwick, Kent Stevens, and an anonymous reviewer for their comments on an earlier draft of the article.

Correspondence concerning this article should be addressed to James T. Todd, Department of Psychology, Brandeis University, Waltham, Massachusetts 02254.

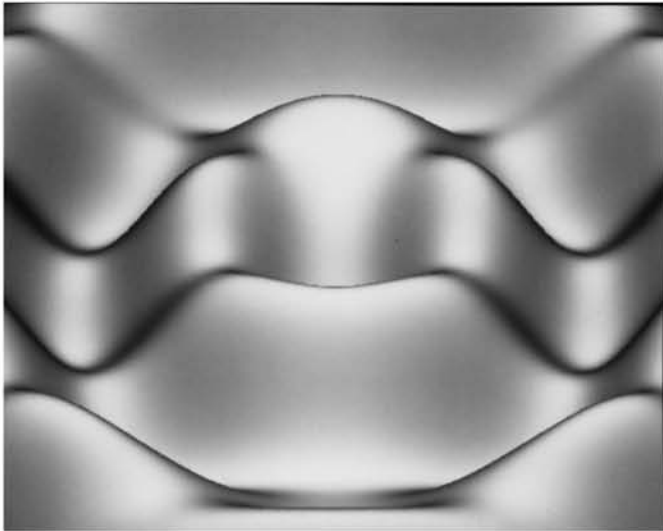


Figure 1. Shaded image of a smoothly curved surface used by Todd, Reichel, and Mingolla (1986) to investigate judgments of local orientation. (On each trial a single point on the surface was designated with a blinking light, and observers were required to estimate the local orientation at that point using a matching task. For many observers, the test-retest reliability was quite poor, so that multiple judgments for each individual point were often separated by large angular deviations.)

structure of a stimulus pattern. Which question to ask can only be determined by the researcher's assumptions (either implicit or explicit) about the salient dimensions of an observer's knowledge. If it is assumed, for example, that our immediate awareness of smoothly curved surfaces is best described as a local orientation map, then the most sensible psychophysical procedure for studying the perception of smoothly curved surfaces would involve judgments of local orientation.

This type of reasoning has provided the theoretical foundation for several recent experiments on the visual perception of surface shape from shading. In one such experiment by Mingolla and Todd (1986), observers were required to make numerical estimates of local orientation (i.e., slant and tilt) for selected points on computer-simulated ellipsoid surfaces with variable reflectance and illumination. Several variations of this procedure have also been examined by Todd, Reichel, and Mingolla (1986), including the replacement of numerical judgments with a matching task and the use of more complicated surface structures, such as the one shown in Figure 1. The same basic pattern of results is obtained in all of these variations: (a) Observers systematically underestimate slant, as if the depicted surface were perceptually flattened into the plane of the display screen. (b) Almost all of the observers report that these tasks are quite difficult and that they have little confidence in their judgments. (c) This assessment is confirmed, moreover, by objective measures of test-retest reliability. When multiple orientation judgments are obtained for a given surface region within a single experimental session, the average angular deviation of these judgments can be as large as  $10^\circ$ .

Similar results have also been obtained for the visual perception of shape from texture. Following the original work of Gibson (1950), there have been numerous experiments reported in

the literature in which observers have judged the absolute slants of planar surfaces (see Braunstein, 1976, for an excellent review). The results of this research show clearly that observers' slant judgments can be highly unreliable and that they tend to be systematically underestimated (see, however, Stevens, 1983; Stevens & Brookes, 1987). This is also true for the judged metric depths of curved ellipsoid surfaces (Todd & Akerstrom, 1987), although it is interesting to note that observers are quite good at nonmetric discriminations between curved and flat (Cutting & Millard, 1984).

When considered as a whole, these results cast serious doubt on the widely held view that local mappings of depth or orientation are the primary form of representation for the visual perception and cognition of smoothly curved surfaces. Although human observers clearly have some degree of visual knowledge about local depths and orientations, the evidence suggests that this knowledge may be surprisingly coarse-grained. Moreover, when we contrast these findings with the compelling impressions of three-dimensional form that are often experienced from shaded or textured images of curved surfaces (e.g., see Figure 1), it is reasonable to speculate that much of our perceptual awareness may involve other, more abstract forms of visual knowledge.

### Nonmetric Descriptions of Surface Shape

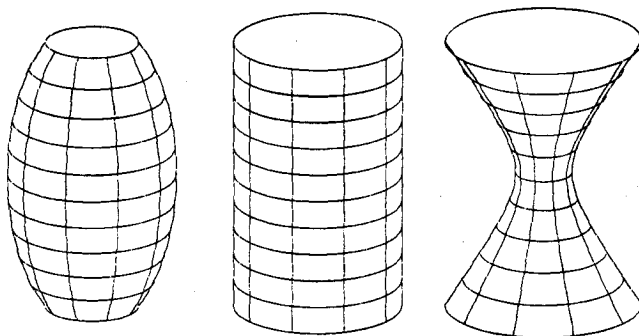
The most common alternative to local mappings of depth and/or orientation in the theoretical analysis of three-dimensional form perception is to decompose an observed object into a relatively small set of categorically distinct parts. There have been several variations of this approach described in the literature. For example, one possible strategy proposed by Koenderink and van Doorn (1976, 1980, 1982; see also Koenderink, 1984; Richards, Koenderink, & Hoffman, 1987; Stevens, 1981b) is to decompose a surface into bounded regions of positive (elliptic), negative (hyperbolic), or zero (parabolic) Gaussian curvature (see Figure 2). Koenderink and van Doorn have demonstrated mathematically that the local Gaussian curvature of a surface can be optically specified by certain types of image features such as smooth occlusion contours or singular points within the field of image intensity. Another strategy proposed by Hoffman and Richards (1984) and by Beusmans, Hoffman, and Bennett (1987) is to divide a surface into parts at loci of negative minima along lines of principal curvature. It is also possible in certain instances to decompose a surface using volumetric primitives as suggested by Biederman (1987).

To better appreciate how these approaches contrast with local mappings of depth or orientation, it is useful to consider a row of neighboring surface regions as schematically represented in Figure 3. Note in the figure that for a *metric description* of depth or orientation, each region would be assigned one or more numbers (e.g., representing depth, slant, or tilt) whose possible values would form a ratio scale (i.e., a surface region with a depth of 20 units would be twice as distant from the point of observation as some other region with a depth of only 10 units). For a *nominal description* of Gaussian curvature, in contrast, each region would be assigned to one of three nominal categories, labeled *positive* (elliptic), *negative* (hyperbolic), or *zero* (parabolic).

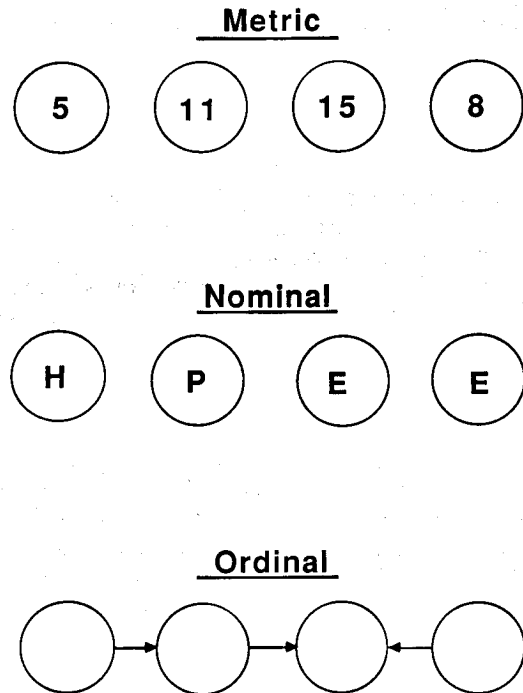
From a purely functional point of view, each of these descriptive frameworks has some specific advantages and limitations. The primary selling point for a metric description of depth or orientation is that it is able to represent the three-dimensional structure of a visible surface with arbitrary precision. Its primary disadvantage is its instability; when a surface moves relative to the observer (or vice versa), the values assigned to each individual surface region would be changing continuously. A nominal description of nonmetric Gaussian curvature has the opposite properties. Because it provides an intrinsic measure of surface structure, it is invariant over changes in viewing position. Its primary disadvantage is its lack of precision. Indeed, there are a wide variety of easily discriminable objects, such as spheres and ellipsoids, in which patterns of Gaussian curvature are identical.

Because neither of these methods of description seem to capture all of the essential characteristics of an observer's perceptual knowledge, we have recently begun to consider a third form of visual representation involving *ordinal descriptions* of surface structure. The basic idea, as shown in Figure 3, is to map out the order relations of depth and/or orientation among neighboring surface regions. This produces an intermediate level of representation—a form of compromise solution that balances the relative strengths and weaknesses of metric and nominal descriptions.

Our proposed analysis of ordinal structure has two important properties that need to be emphasized. First, the relevant depth-order relations are defined with respect to the point of observation within a viewer-centered coordinate system. An ordinal description is similar in this respect to local mappings of depth or



**Figure 2.** Three curved surfaces with different Gaussian curvatures. (The contours on each surface depict lines of principal curvature, which lie in the directions of maximum and minimum change in surface orientation. The principal curvatures at each surface point can be metrically represented along a ratio scale, so that convexities have positive values and concavities have negative values. The metric Gaussian curvature at a point is the product of its two principal curvatures. The nonmetric Gaussian curvature is simply the sign of this product: A positive [elliptic] Gaussian curvature occurs when the principal curvatures are both concave or both convex; a negative [hyperbolic] Gaussian curvature occurs when one of the principal curvatures is concave, whereas the other is convex; and a zero [parabolic] Gaussian curvature occurs when one or both of the principal curvatures are zero. These three surfaces are designed to exemplify each of these categories. From left to right, the depicted objects include an elliptic surface, a parabolic surface, and a hyperbolic surface.



**Figure 3.** Alternative representations for a row of neighboring surface regions. (For a *metric* representation of egocentric depth, as proposed by Marr, 1982, each region would be assigned a numeric value specifying its distance from the point of observation. For a *nominal* representation of nonmetric Gaussian curvature, as proposed by Koenderink & van Doorn, 1976, 1980, 1982, each region would be assigned to one of three possible categories labeled *elliptic* (E), *hyperbolic* (H), or *parabolic* (P). For an *ordinal* representation of egocentric depth as proposed by the present article, each pair of neighboring surface regions would be labeled in terms of which region is closest to the point of observation. As shown in Figure 3, each arrow points in the direction of decreasing depth.)

orientation, but it contrasts sharply with object-centered descriptions of intrinsic surface structure such as those proposed by Koenderink and van Doorn (1976, 1980, 1982). Because of this viewer-centered frame of reference, the depth-order relation between any two surface regions need not remain invariant over changes in viewing position. However, the global pattern of these relations throughout an entire image will generally exhibit a high level of stability. (This will be considered in more detail in the discussion of Experiment 3.)

The second important property of our proposed analysis of ordinal structure is that the depth-order relations on which it is based are only defined for adjacent regions within an arbitrarily small neighborhood. This too has some significant consequences. Suppose, for example, that we wish to determine the ordinal depth relation between two visible surface regions,  $R_1$  and  $R_n$ , that are not locally adjacent to one another. If an observer had knowledge of the metric depths of the surface, then this type of judgment would be quite trivial. The problem is more difficult, however, using a local representation of ordinal depth. The relative depths of  $R_1$  and  $R_n$  can only be determined in that case if there is a continuous chain of intervening regions that are ordinally transitive (i.e., if  $R_1 < R_2 < R_3 \dots < R_n$ , then

$R_1 < R_n$ ). If this restriction is violated (i.e.,  $R_1 < R_2 < R_3 \dots > R_n$ ), then the relative depths of  $R_1$  and  $R_n$  cannot be determined from an ordinal representation without providing some additional information.

To avoid having to search among the infinity of possible paths between two depicted surface points in a visual image, we will restrict our analysis of ordinal transitivity to the one path that projects to a linear segment in the image plane. Consider, for example, the regions highlighted by small black dots in the shaded images of smoothly curved surfaces presented in Figure 4. The two dots in the upper figure mark a pair of surface regions that are ordinally transitive: Moving between them in a straight line from right to left, the depicted surface depth decreases monotonically, so that each successive position is closer to the point of observation than is its preceding neighbor. Thus, for these regions, an ordinal description of the depicted surface is sufficient to determine their relative depths. The two small dots on the lower figure, in contrast, mark a pair of surface regions that are ordinally intransitive: Moving between these points in a straight line produces a nonmonotonic change in the depicted surface depth, and our proposed ordinal representation would therefore be inadequate to determine which region is closest to the point of observation.

Could an ordinal description of surface structure as outlined above have any psychological relevance to the processes of visual perception and cognition? The research described in the present article began as an effort to address this issue. Our initial investigation was focused on a strong empirical prediction from the preceding analysis: If human observers use ordinal knowledge of smoothly curved surfaces for judgments of relative depth, then ordinally transitive surface regions should be much easier to judge than ordinally intransitive surface regions. Experiment 1 was designed specifically to test this hypothesis.

## Experiment 1

### Method

**Subjects.** Subjects were 6 Brandeis graduate students who volunteered to participate in the experiment.

**Apparatus.** Stimuli were produced using a Lex-90 graphics system and were displayed on a 19-in. (48.5-cm) color monitor. Observers viewed the displays binocularly at a distance of approximately 115 cm (about 3.8 ft). The stimuli were presented in a rectangular region of the display screen that was 33 cm (13 in.) along the horizontal axis and 26 cm (10.2 in.) along the vertical axis. The spatial resolution within this viewing window was  $640 \times 512$  pixels. Head movements were not restricted.

**Stimuli.** On each trial, observers were presented with one of the two images of smoothly curved surfaces depicted in Figure 4. The surfaces were generated using a Cartesian coordinate system ( $X, Y, Z$ ), where  $X$  and  $Y$  were aligned with the horizontal and vertical axes of the display screen and  $Z$  was perpendicular to the screen in depth. For the large-scale surface at the bottom of Figure 4, the depth at each point was determined initially by the following equation:

$$Z = a \cos(r\pi/w),$$

where  $r = \sqrt{X^2 + Y^2}$ ,  $a = 100$ , and  $w = 100$ . An identical equation was also used for the small-scale surface at the top of Figure 4, except that there was a 25% reduction in its amplitude and period (i.e.,  $a = 75$  and  $w = 75$ ). In the actual stimulus displays (as shown in Figure 4), these

mathematically defined surfaces were depicted under parallel projection at a  $40^\circ$  slant from the image plane.

The intensity ( $I$ ) of each picture element was determined with seven-bit precision using the following equation:

$$I = 128(L \cdot N),$$

where  $L$  is a unit vector in the direction of the light source, and  $N$  is a unit vector that is perpendicular to the surface at the depicted point (see Todd & Mingolla, 1983). This relation is a reasonable approximation of how light reflects from a pure matte (Lambertian) surface. The pattern of illumination in each display simulated an infinitely distant point light source at a  $20^\circ$  angle directly above the simulated direction of view.

Using an exhaustive search over the large-scale surface, 36 horizon-

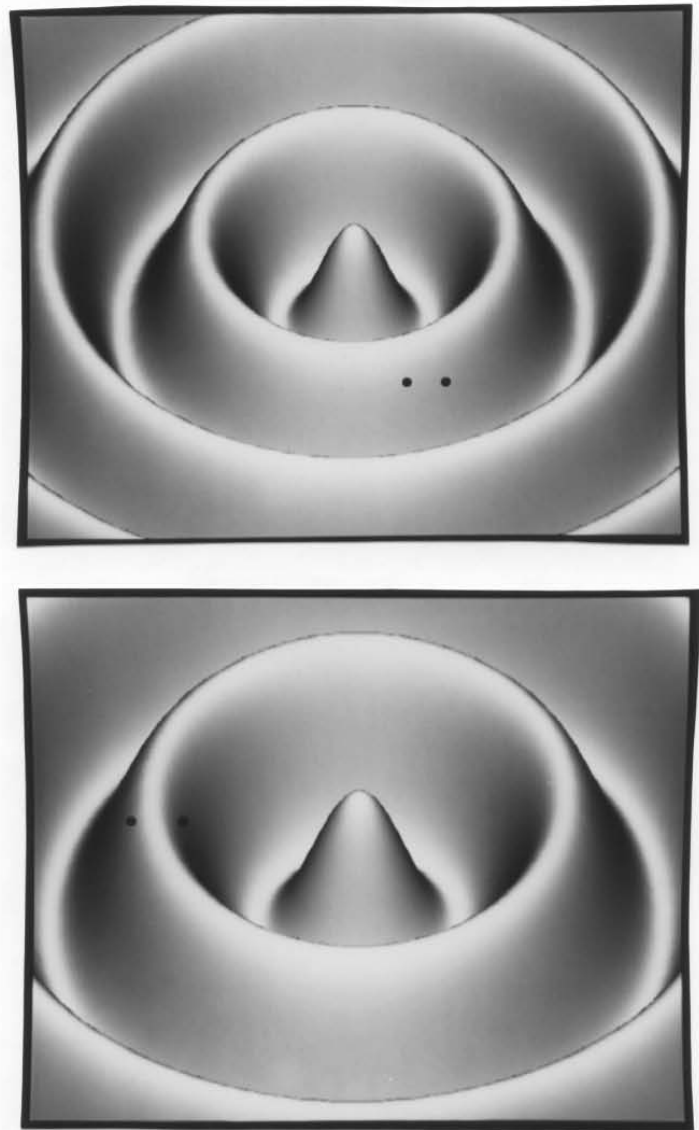


Figure 4. Two shaded images of smoothly curved surfaces used in Experiment 1. (The two small dots on the upper, small-scale surface are ordinally transitive, whereas the two dots in the lower, large-scale surface are ordinally intransitive. Note that the relative depth is much easier to discern for the upper pair of dots, even though their metric depth difference is 25% smaller.)

tally aligned point pairs were selected for use in a relative depth-judgment task. One half of these pairs were ordinally transitive, whereas the remaining half were ordinally intransitive. These two sets of point pairs were perfectly matched for their horizontal separations in the image plane and for their simulated separations in depth. There were three possible horizontal separations of 50, 75, and 100 pixels, and three possible depth separations of 25, 50, and 75 pixels. For each combination of transitivity, horizontal separation, and depth separation, two exemplars were selected from the total set of possible pairings. This selection was constrained so that the pairs would be balanced in terms of which point (right or left) was closest to the observer in depth and would be evenly distributed over the visible surface. Because the images were bilaterally symmetric, moreover, additional variation was achieved by presenting each point pair with equal frequency at two possible locations on either side of the display. The same sample of 36 point pairs was also used for the small-scale surface, although it is important to note that as a result of the scale change, their spatial separations were reduced by 25%, both horizontally and in depth.

*Procedure.* Each trial of the experiment began with a countdown against a homogeneous black background: In 1-s intervals, the number 3 was displayed, followed by the number 2, followed by a single pair of small red dots presented in isolation, which allowed the observers to direct their gaze to the appropriate region of the display screen. Following this countdown the two small dots remained in view, but the background was replaced abruptly by a single shaded image of a smoothly curved surface (see Figure 4). Observers were required to indicate which of the two surface regions marked by the dots (right or left) appeared closest in depth. They were instructed to respond as rapidly as possible, without sacrificing accuracy, by pressing one of two response keys on the computer keyboard. The depicted surface remained on the screen with the two red dots until an appropriate response was recorded. No feedback about the accuracy of their responses was provided until after the experiment was completed.

All 72 possible point pairs (2 exemplars  $\times$  2 types of transitivity  $\times$  3 horizontal separations  $\times$  3 depth separations  $\times$  2 surface scales) were presented four times each over two experimental sessions. The observers were given 20 trials of practice at the beginning of the first session to familiarize themselves with the procedure. All of them reported at the conclusion of this practice that the task was straightforward and that they were highly confident in the accuracy of their judgments.

### Results and Discussion

Our primary goal in designing Experiment 1 was to compare observers' relative depth judgments for ordinally transitive and ordinally intransitive surface regions. Although none of the observers reported having consciously noticed this particular stimulus manipulation until it was explicitly pointed out to them, its effect on their performance was quite dramatic. For the ordinally transitive point pairs, the observers were 98.0% accurate, with an average response time of 719 ms. For the ordinally intransitive point pairs, their accuracy was reduced to 88.3%,  $F(1, 355) = 64.62, p < .001$ , and their average response time was increased to 1,093 ms,  $F(1, 355) = 414.57, p < .001$ . Indeed, among all of the various stimulus manipulations in this experiment, the ordinal transitivity of the point pairs was by far the most significant determinant of the observers' performance, accounting for 34% of the variance in accuracy between the different point pairs and more than 68% of the variance in reaction time.<sup>1</sup>

With respect to the other stimulus manipulations, performance was improved slightly by increasing the depth separation between the point pairs,  $F(2, 355) = 4.29, p < .05$ , for the accu-

racy data, and  $F(2, 355) = 4.31, p < .05$ , for the reaction-time data, or by decreasing their horizontal separation,  $F(2, 355) = 3.86, p < .05$ , for the accuracy data, and  $F(2, 355) = 9.76, p < .001$ , for the reaction-time data. All of these effects were quite small, however, and accounted for less than 5% of the variance. The average response time decreased significantly between the first and second experimental sessions,  $F(1, 355) = 107.67, p < .001$ , but there was no change in accuracy. The variation in surface scale had no significant effect.

It is important to keep in mind when considering these results that the ordinally transitive and ordinally intransitive point pairs were perfectly matched for their horizontal separations in the image plane and their simulated separations in depth. What distinguished these conditions was the pattern of depth change in the intervening surface regions between each pair of points. That is to say, there were monotonic depth changes between the ordinally transitive point pairs and nonmonotonic depth changes between the ordinally intransitive point pairs.

We now examine how this difference would be expected to affect performance for various types of visual representations. One possibility to consider is that observers' perceptions of smoothly curved surfaces are defined in terms of local mappings of absolute metric depth. To determine relative depth order from this type of representation, it would only be necessary to compare the depth values at the two marked locations. The intervening pattern of depth change would be irrelevant. Alternatively, observers' perceptions could also be defined in terms of local mappings of surface orientation or by relative metric depth relations between adjacent regions in small local neighborhoods. To determine the depth order of a pair of points using either of these representations, it would be necessary to integrate over the intervening surface regions through which they are connected (cf. Stevens & Brookes, 1987). From the perspective of an information-processing analysis, it would be reasonable to expect that the time required for performing this integration would be likely to increase with the spatial separation between the two designated points. This hypothesis provides no obvious explanations, however, for why the ordinal transitivity of the intervening surface regions should have any effect at all on either accuracy or reaction time.

Suppose, on the other hand, that visual knowledge of smoothly curved surfaces is defined in terms of local depth-order relations, so that neighboring surface regions are perceptually distinguished by which region is closer to the point of observation, without specifying how much closer. Unlike the other possibilities considered earlier, this hypothesis makes a strong prediction about the effects of ordinal transitivity. Indeed, with this type of ordinal representation, the relative depth order of a spatially separated pair of points can only be determined if the intervening surface regions through which they are

<sup>1</sup> The accuracy and reaction-time data were analyzed separately using a one-way analysis of variance (ANOVA) for the 72 different point pairs, with a set of orthogonal planned comparisons to evaluate differences among specific subgroups. Because all of these comparisons had the same error term as in the overall ANOVA, their relative proportions of the variance can easily be determined by a direct comparison of their corresponding  $F$  values.



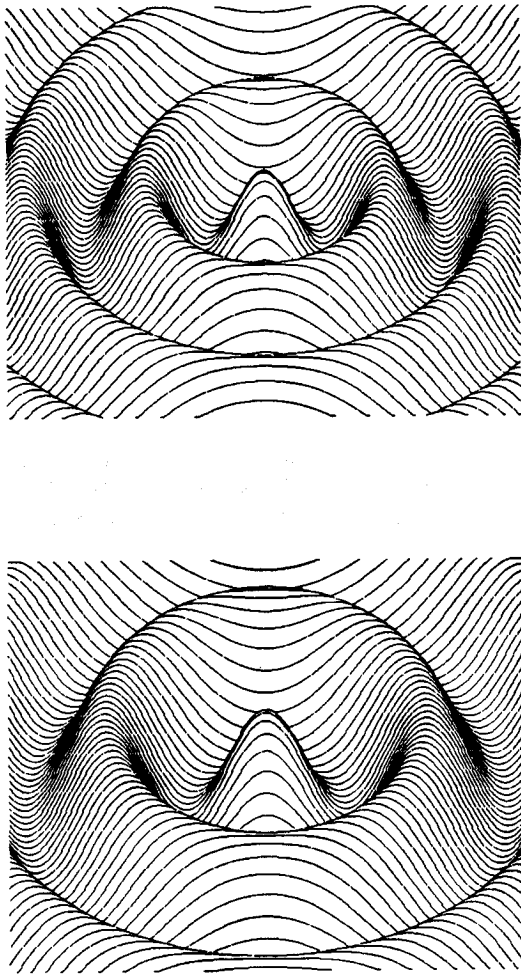


Figure 5. Two patterns of image contours used in Experiment 2. (The depicted surfaces are structurally identical to those shown in Figure 4 using patterns of shading.)

connected are ordinally transitive. We would therefore expect that performance should deteriorate whenever this restriction is violated—a prediction that was confirmed in the present experiment by the reduced accuracy and increased reaction times of the observers' judgments.

These findings provide strong evidence that local depth-order relations can be of great importance to the visual perception of smoothly curved surfaces, but there is also evidence to suggest that observers are capable, if necessary, of basing their responses on other aspects of surface structure. Although performance was significantly impaired for the ordinally intransitive point pairs, the correct depth orders in that condition could still be determined with more than 88% accuracy. One possible explanation of this result is that visual images of smoothly curved surfaces are perceptually analyzed at multiple levels of description, including a rapid fine-grained analysis of ordinal structure and a much slower course-grained analysis of metric structure. Within this framework, observers would presumably rely on whichever analysis is most appropriate for the required judgment.

## Experiment 2

One relevant issue for interpreting the results of Experiment 1 concerns the domain of generality for the ordinal transitivity effect. Is it, for example, restricted to patterns of image shading, or does it reflect a more general property of visually based knowledge? In Experiment 2, we attempted to address this issue by examining an alternative method of depiction involving patterns of image contours (see Figure 5).

### Method

The same 6 observers who had participated in Experiment 1 again volunteered their services.<sup>2</sup> The apparatus, stimuli, and procedure were identical in all respects to those used in the previous study, except that the surfaces were depicted using patterns of image contours rather than shading (see Figure 5). As described for Experiment 1, the depth  $Z$  of each surface point was defined initially as a function of  $X$  and  $Y$ :  $Z = f(X, Y)$ . A horizontally aligned contour on this surface could be defined quite simply by holding  $Y$  constant at a fixed value  $Y_i$ . A pattern of contours was generated using multiple values of  $Y_i$  in 10 pixel increments. These mathematically defined contours were then rotated 40° with respect to the initial  $X, Y$  coordinate system in the plane of the display screen. They were presented under parallel projection as black lines against white background with their occluded portions removed.

### Results

The results were remarkably similar to those obtained for shaded surfaces in Experiment 1. For the ordinally transitive point pairs, the observers were 98.3% accurate, with an average response time of 542 ms. For the ordinally intransitive point pairs, their accuracy was reduced to 86.8%,  $F(1, 355) = 67.97$ ,  $p < .001$ , and their average response time was increased to 988 ms,  $F(1, 355) = 579.59$ ,  $p < .001$ . These effects accounted for 34% of the variance in accuracy between the different point pairs and more than 76% of the variance in reaction time.

The effects of the other stimulus manipulations were also similar to those obtained in Experiment 1. Performance was improved by increasing the depth separation between the point pairs,  $F(2, 355) = 3.8$ ,  $p < .05$ , for the accuracy data, and  $F(2, 355) = 10.9$ ,  $p < .001$ , for the reaction-time data, or by decreasing their horizontal separation,  $F(2, 355) = 16.05$ ,  $p < .001$ , for the accuracy data, and  $F(2, 355) = 32.95$ ,  $p < .001$ , for the reaction-time data. The average response time decreased significantly between the first and second experimental sessions,  $F(1, 355) = 20.30$ ,  $p < .001$ , but there was no change in accuracy. The variation in surface scale had no significant effect.

These results show clearly that the perceptual distinction between ordinally transitive and ordinally intransitive surface regions is not restricted to shaded images and can occur over multiple methods of depiction. This conclusion has recently been reinforced, moreover, in a closely related experiment by Todd (1989), in which surfaces were depicted using patterns of optical motion, with careful controls to eliminate all possible informa-

<sup>2</sup> Because the same observers participated in several experiments, they were not debriefed about the overall pattern of results or informed of the authors' hypotheses until after the entire series of experiments was completed.

tion from shading or texture. Taken as a whole, these findings provide strong evidence that local order relations are a fundamental parameter of description for visually based knowledge in general, irrespective of the particular source of optical structure from which that knowledge is obtained.

### Experiment 3

The remaining experiments in the present series were designed to discover some of the specific stimulus variables by which order relations are visually specified in patterns of image shading. In Experiment 3, for example, we attempted to determine the minimum amounts of image structure that are required for accurate depth-order judgments.

### Method

The apparatus and general procedure were identical to those used in Experiments 1 and 2, and the same 6 observers again volunteered their services. Two shaded images of smoothly curved surfaces—the one shown in Figure 6 and its mirror reflection—were used as the basis for all of the experimental displays. The observers never saw these images in their entirety, however, because the displays were restricted to small rectangular subregions that were randomly repositioned from their corresponding locations in the base image. There were four possible subregion sizes of 0.88%, 3.52%, 7.90%, and 14.06% of the total image area (see Figure 7).

In generating the depicted surface, the depth  $Z$  of each point was defined as a function of  $X$  and  $Y$ , using the following series of equations:

$$r = \sqrt{X^2 + Y^2},$$

$$g_1 = 200e^{[Y^2 - (X-50)^2]/(800\pi)},$$

$$g_2 = 400e^{[(Y+60)^2 - (X+150)^2]/(1250\pi)},$$

and

$$Z = 150 \cos(r\pi/150) + g_1 + g_2.$$

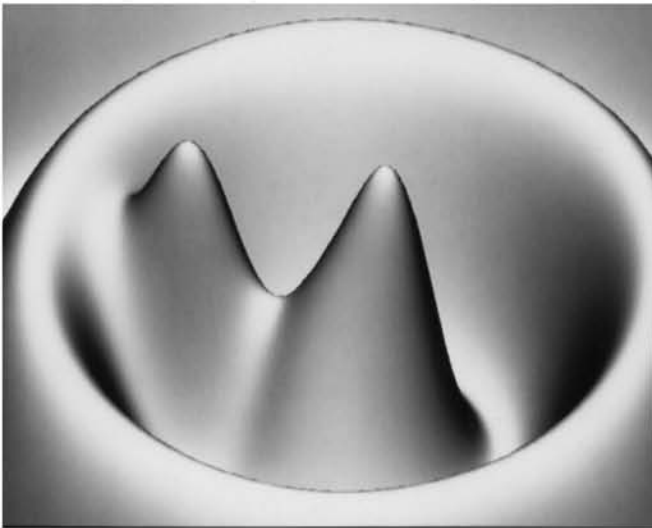


Figure 6. Shaded image of a complex surface used in Experiment 3. (The observers never saw this surface in its entirety because the stimulus displays were restricted to small rectangular subregions. See Figure 7.)

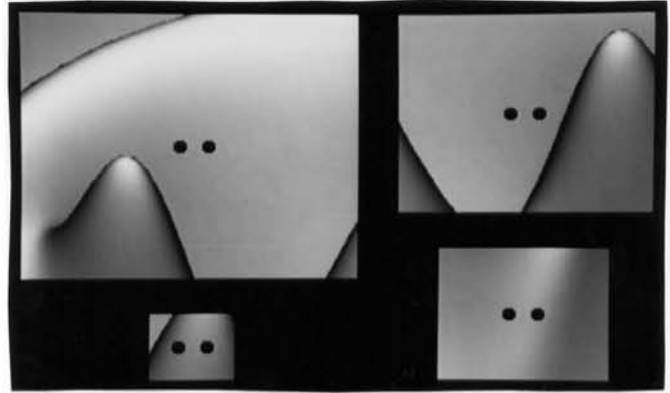


Figure 7. Four example subregions of variable size from Experiment 3. (The two subregions on the left are from the attached contour condition. The one on the upper right is from the unattached contour condition, and the one on the lower right is from the no contour condition.)

The surface was then rotated  $40^\circ$  with respect to the initial  $X, Y$  coordinate system. It was displayed under parallel projection with a simulated pattern of illumination that formed a  $20^\circ$  slant with the image plane.

Using an exhaustive search over the image in Figure 6, 400 ordinal transitive point pairs were selected, so that their horizontal separations would be 20 pixels and their depth separations would be between 6 and 23 pixels. On each trial, observers were presented with one of these point pairs selected at random, within one of its four possible subregions, at a randomly selected position on the display screen. An experimental session consisted of 144 trials, and each observer participated in two sessions.

### Results and Discussion

It quickly became apparent while piloting this experiment that the presence or absence of occlusion contours within each subregion would have a large effect on the observer's performance (cf. Ikeuchi & Horn, 1981). The occlusion contours of a smoothly curved surface are defined mathematically as the set of all visible points whose surface normals are perpendicular to the observer's line of sight (see Koenderink & van Doorn, 1976, 1982, for an elegant analysis of their mathematical properties). To better appreciate how occlusions relate to ordinal structure, it is useful to consider a planar cross-section of a smoothly curved surface as schematically represented in Figure 8. Note in Figure 8 that an occlusion contour separates a visible scene into two distinct regions (labeled  $A$  and  $B$ ) but that its relationship with these regions is anisotropic—that is to say, the depicted occlusion point is *attached* to region  $A$  but is *unattached* to region  $B$ . As we move from an occlusion contour in an attached region, the ordinal depth of the surface must decrease monotonically until a depth minimum is reached. This need not be the case, however, for an unattached region, as is exemplified in the figure.

In analyzing the results of the present experiment, each possible combination of point pair and subregion size was coded into three separate categories: an *attached contour condition*, in which a visible occlusion contour was attached to the area of the depicted point pair (see Figure 7, left); an *unattached contour condition*, in which occlusion contours were clearly visible but

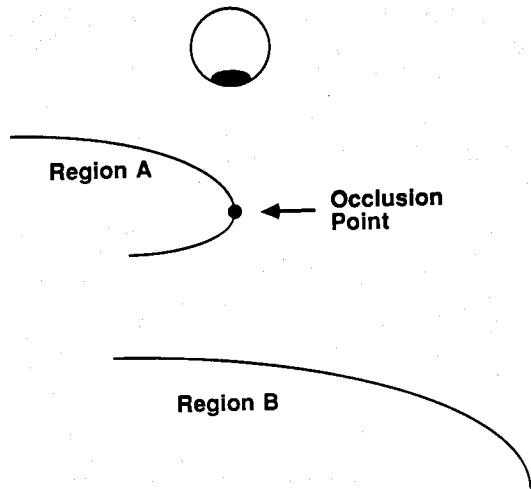


Figure 8. An occlusion point separates a visible scene into two distinct regions. (In this particular example, the occlusion point is attached to Region A but is unattached to Region B. As we move from an occlusion point in an attached region, the surface depth must decrease monotonically until a depth minimum is reached.)

were all unattached to the area of the depicted point pair (see Figure 7, upper right); and a *no contour condition*, in which there were no visible occlusion contours at all (see Figure 7, lower right). It is important to note that these conditions did not occur with equal frequency, inasmuch as the probability of encompassing an occlusion contour varied with subregion size (see Table 1).

Figures 9 and 10, respectively, show the accuracy and reaction times of the observers' judgments as a function of subregion size for all three occlusion categories. It is clear from these data that the presence of unattached occlusion contours did not improve performance relative to the no contour condition.<sup>3</sup> The attached contours, on the other hand, had a dramatic influence, producing a significant increase in accuracy,  $F(1, 75) = 67.29, p < .001$ , and a significant decrease in reaction time,  $F(1, 75) = 27.06, p < .001$ . The only other significant effect was a small improvement in performance with increasing subregion size,  $F(3, 75) = 6.89, p < .001$ , for the accuracy data, and  $F(3, 75) = 6.35, p < .001$ , for the reaction-time data.

The perceptual significance of attached occlusion contours, as demonstrated by these results, can sometimes be observed in an alternative manner by simply reorienting a surface. Two relevant examples of this phenomenon are shown in Figure 11. The surfaces depicted in these images are identical to those in

Table 1  
Total Number of Trials Across All Six Observers for Each of the Different Stimulus Combinations in Experiment 3

Condition	Subregion size (% of total area)			
	0.88	3.52	7.90	14.06
No contours	225	57	15	0
Unattached contours	117	190	187	107
Attached contours	90	185	230	325

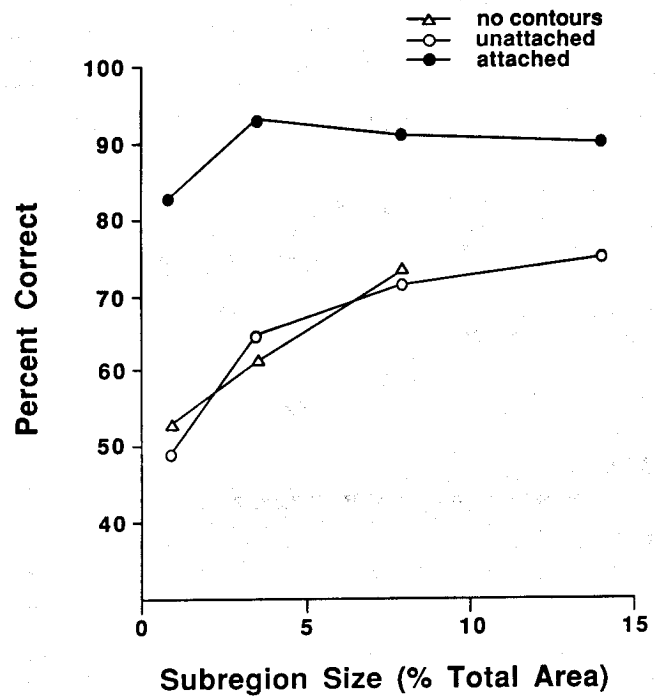


Figure 9. Percentage of correct responses for 6 observers as a function of subregion size for the three occlusion categories of Experiment 3.

Figures 1 and 4, but they are shown from a vantage point from which there are no occlusions. When viewed from this particular perspective, the visual impression of a smoothly curved surface is much less compelling.

#### Localization of Depth Extrema

Although occlusion contours may be a primary source of visual information about the ordinal structure of smoothly curved surfaces, it is important to recognize that they cannot be the sole basis of ordinal knowledge. Consider, for example, the planar cross-section of a smoothly curved surface shown in Figure 12. Note in Figure 12 that there are two occlusion points,  $O_1$  and  $O_2$ . We know from the earlier discussion that the ordinal depth of a surface must decrease monotonically as we move from an occlusion point in an attached region, and that this can provide perceptually salient information about the ordinal structure of the surface in the neighborhood of the occlusion. A complete ordinal representation cannot be achieved, however, without also identifying the depth extrema  $E_1, E_2$ , and  $E_3$ . These extremal points (i.e., the depth maxima and minima) define the boundaries of ordinal transitivity in which a monotonic depth change switches from positive to negative, or vice versa. Their optical projections  $E_1, E_2$  and  $E_3$ , together with those of the occlusion points  $O_1$  and  $O_2$ , are both necessary and sufficient for visually specifying the complete ordinal structure of the depicted cross section.

<sup>3</sup> Because of the empty cell in the no contour condition, the responses from this condition were excluded from further statistical analyses.



To better appreciate the lawful regularities of this type of ordinal representation, it is useful to consider the observed depth extrema along a fixed linear path through visual space. Because they are defined within a viewer-centered frame of reference, the positions of the depth extrema along any such path would be unlikely to remain invariant if the surface were moved relative to the point of observation. As noted in the introduction, however, the overall pattern of depth-order relations would tend to be highly stable. That is to say, the path would be partitioned by its depth extrema into a small number of bounded, ordinal transitive regions (cf. Hoffman & Richards, 1984), whose global organization would remain relatively stable over time, except for occasional accretions or deletions at points of occlusion. The ordinal structure would be especially constrained in regions bounded by attached occlusion points (e.g., see Figure 12), because the number of minima within such a region must always be one greater than the number of maxima.

Because the construction of an accurate ordinal representation is largely dependent on localizing the depth extrema along each surface cross-section, it is important to consider the specific sources of information for identifying their optical projections within a visual image. Surfaces depicted with patterns of image shading are particularly problematic in this regard, because of potential variations in the direction of illumination.<sup>4</sup> One relevant property of shaded images that is generally applicable over a wide range of illumination conditions is that the depth extrema (i.e., maxima and minima) along any surface cross-section are often in one-to-one correspondence with the luminance maxima on that cross-section.<sup>5</sup> This might suggest, as a working hypothesis, that perceived depth decreases monotonically from an attached occlusion contour and reverses direction at each luminance maximum. There are, however, some

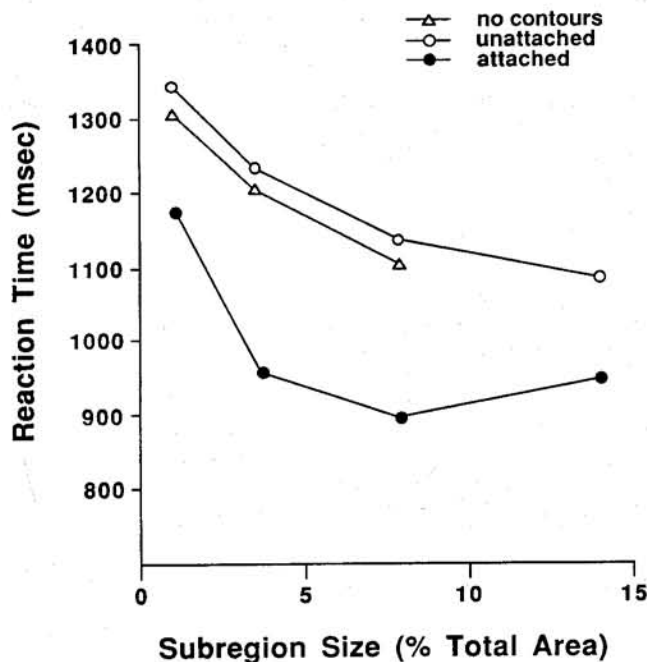


Figure 10. The mean reaction times of 6 observers as a function of subregion size for the three occlusion categories of Experiment 3.

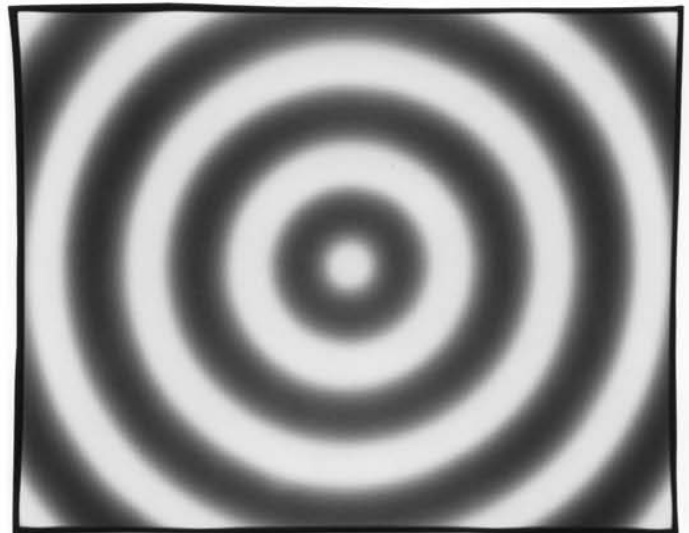
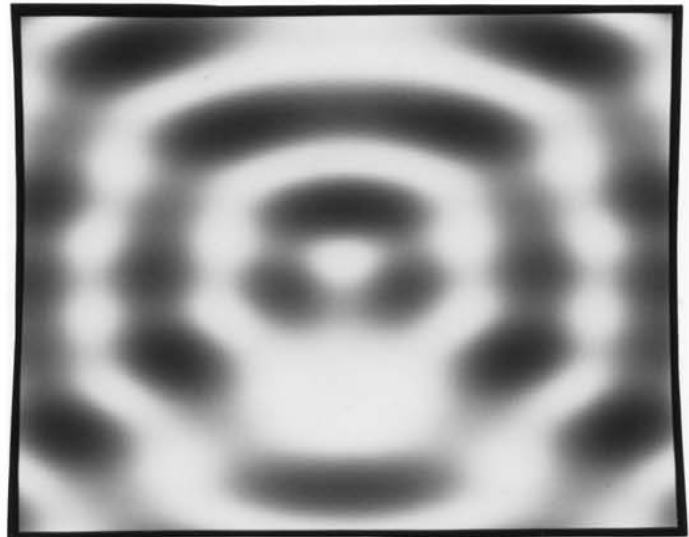


Figure 11. Two shaded images of smoothly curved surfaces as viewed from a vantage point from which there are no occlusions. (Note in this case that the visual impression of three-dimensional form is much less compelling than when the same surfaces are viewed from a different orientation as shown in Figures 1 and 4.)

serious difficulties with this strategy. Although they are often numerically equivalent, the depth extrema and luminance maxima along any linear segment of an image will not in general be spatially coincident unless the segment is orthogonal to the direction of illumination. To the extent that this condition is

<sup>4</sup> For other methods of depiction involving texture, motion, or stereopsis, the optical projections of the depth extrema can generally be determined from the maxima and minima of measurable image properties such as size, velocity, or disparity.

<sup>5</sup> This one-to-one correspondence breaks down if an image contains sharp specular highlights or if there are multiple sources of illumination that are widely separated in space.

violated, the luminance maxima will be pulled farther and farther from the depth extrema. Experiment 4 was designed, therefore, in an effort to determine how this affects observers' perceptions of three-dimensional form.

#### Experiment 4

##### Method

*Subjects.* Subjects were 5 observers who volunteered their services; 3 of the observers had also participated in Experiments 1, 2, and 3.

*Apparatus.* Stimuli were produced using an Iris 3130 graphics workstation and were displayed on a 19-in. (48.5-cm) color monitor. Observers viewed the displays binocularly at a distance of approximately 100 cm (about 3.4 ft). The stimuli were presented within a rectangular region of the display screen that was 36 cm (about 14 in.) along the horizontal axis and 27 cm (about 11 in.) along the vertical axis. The spatial resolution within this viewing window was  $1,024 \times 780$  pixels. Head movements were not restricted.

*Stimuli.* On each trial, observers were presented with a smoothly shaded image of a computer-simulated ellipsoid surface such as those shown in Figure 13 (see Mingolla & Todd, 1984). The surfaces were generated using a Cartesian coordinate system ( $X, Y, Z$ ), where  $X$  and  $Y$  were aligned with the horizontal and vertical axes of the display screen and  $Z$  was perpendicular to the screen in depth. The surfaces were constrained so that they would all produce identical occlusion boundaries when viewed under parallel projection. Each occlusion boundary appeared on the display screen as a vertically oriented ellipse with horizontal and vertical semiaxes of 150 and 300 pixels, respectively. Given this constraint, the only remaining degrees of freedom for uniquely defining the surface were to identify the position of the depth extremum  $E = (X_E, Y_E, Z_E)$  within the elliptical outer boundary.

Each surface was defined using the following equation:

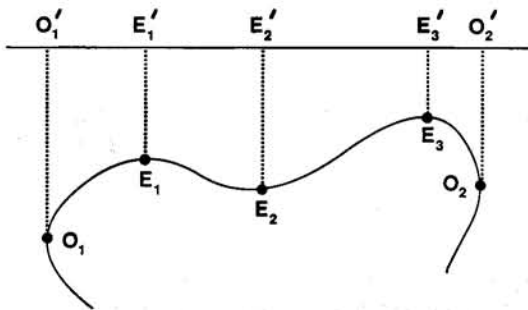
$$C_{11}X^2 + C_{22}Y^2 + C_{33}Z^2 + C_{13}XZ = 1,$$

where

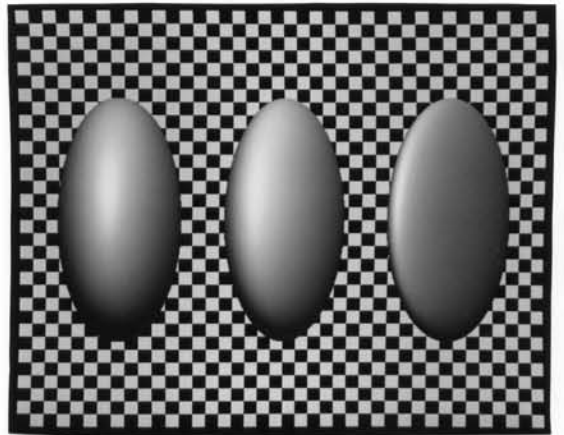
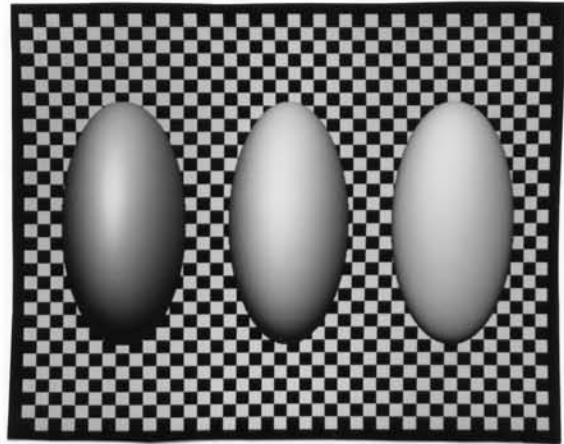
$$C_{13} = X_E / [(X_E^2 - 150^2)Z_E],$$

$$C_{33} = (1 - C_{13}X_E Z_E) / Z_E^2,$$

$$C_{22} = 1/300^2,$$



*Figure 12.* The ordinal structure of any surface is completely determined by the optical projections of its occlusion points and depth extrema. (In this particular example, there are two occlusion points,  $O_1$  and  $O_2$ , and three depth extrema,  $E_1$ ,  $E_2$ , and  $E_3$ . Their corresponding optical projections in the picture plane are labeled  $O_1'$ ,  $O_2'$ ,  $E_1'$ ,  $E_2'$ , and  $E_3'$ , respectively.)



*Figure 13.* Six shaded images of ellipsoid surfaces used in Experiment 4. (The three surfaces in the lower figure vary in both their ordinal and metric structures. Moving from left to right the depicted values of  $X_E$  are 25, 75, and 125 pixels, respectively, with  $Z_E = 300$  pixels and  $L_0 = [0, .5, .866]$ . The three surfaces in the upper figure vary in their metric structures, but are ordinally identical. From left to right, the depicted values of  $Z_E$  are 300, 200, and 100 pixels, respectively, with  $X_E = 25$  and  $L_0 = [0, .5, .866]$ .)

and

$$C_{11} = -C_{13}Z_E/X_E.$$

This produces a family of ellipsoids with depth extrema that are all vertically centered (i.e.,  $Y_E = 0$ ) but that can vary both horizontally and in depth depending on the parameters  $X_E$  and  $Z_E$ .

In generating the displays, the horizontal component  $X_E$  had three possible values of 25, 75, and 125 pixels, relative to the center of the elliptical occlusion boundary. These variations (see Figure 13, bottom) significantly altered both the metric and ordinal structures of the depicted surfaces. The depth component  $Z_E$  had three possible values of 100, 200, and 300 pixels, relative to the plane of the display screen. These variations (see Figure 13, top) could significantly alter the metric structure of a surface, but had no effect on its ordinal structure (i.e., the ordinal structure is invariant over changes in  $Z_E$ ).

To better appreciate the relative structures of these surfaces in depth, it is useful to consider a horizontal cross-section through the center of each ellipsoid perpendicular to the image plane, as shown in Figure 14. Each curve in Figure 14 shows the visible portion of a surface cross-section, with depth represented along the vertical axis. Note, in particu-

lar, how the curves vary in shape because of the different possible combinations of  $X_E$  and  $Z_E$ . For the sake of comparison, the top row and left column of Figure 14 show the horizontal cross-sections for the different ellipsoids depicted in Figure 13.

The image intensity ( $I$ ) at each point on a surface was determined with eight-bit precision using the following equation (see Todd & Mingolla, 1983):

$$I = 76 + 128(L \cdot N) + 51(H \cdot N)^{20},$$

where  $L$  is a unit vector toward the light source,  $N$  is a unit surface normal, and  $H$  is a unit vector that bisects  $L$  and  $N$ . This provides a reasonable approximation of a moderately specular surface illuminated by a point light source, with a 30% component of diffuse ambient illumination. There were three possible directions of illumination relative to the observer's line of sight:  $L_{-30} = (-.5, .433, .75)$ , where the light source was displaced to the left by  $30^\circ$ ;  $L_0 = (0, .5, .866)$ , where the light source was horizontally centered; and  $L_{30} = (.5, .433, .75)$ , where the light source was displaced to the right by  $30^\circ$ . In each case, these horizontal displacements of the light source were accompanied by an additional upward displacement of  $30^\circ$ .

*Procedure.* Each trial of the experiment began with the presentation of a single ellipsoid surface against a checkerboard background on the left half of the display screen. Observers were required to make two separate judgments about the perceived three-dimensional structure of a horizontal cross-section through the center of each ellipsoid. First, they evaluated the ordinal structure of this cross-section by indicating the point that appeared to be closest in depth to the point of observation. This was accomplished with a mouse-controlled cursor that was constrained to move along the horizontal cross-section between the two occlusion boundaries. Once the cursor was positioned correctly, the response was recorded by pressing an appropriate button on the mouse. This initiated the presentation of a  $180^\circ$  elliptical arc, similar to those

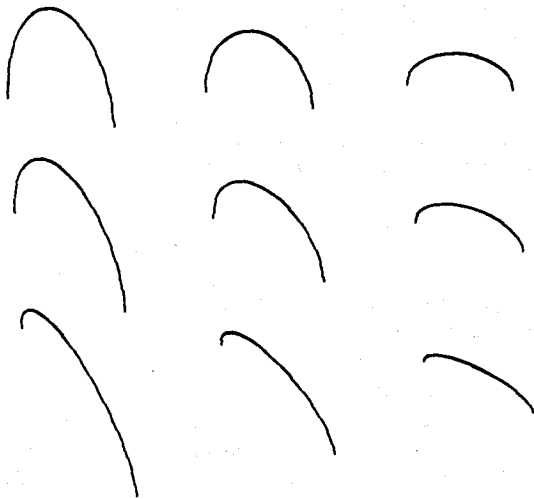


Figure 14. The horizontal cross-sections through the centers of the ellipsoids used in Experiment 4. (Each curve depicts the visible portion of a surface cross-section, with depth represented along the vertical axis. The curves vary in shape because of the different possible values of  $X_E$  and  $Z_E$  from which they were generated. They are arranged in a matrix so that the horizontal position,  $X_E$ , of the depth extremum varies down the columns and its position in depth ( $Z_E$ ) varies across the rows. For the sake of comparison, the top row and left column, respectively, show the relative three-dimensional structures of the different ellipsoids depicted in Figure 13.)

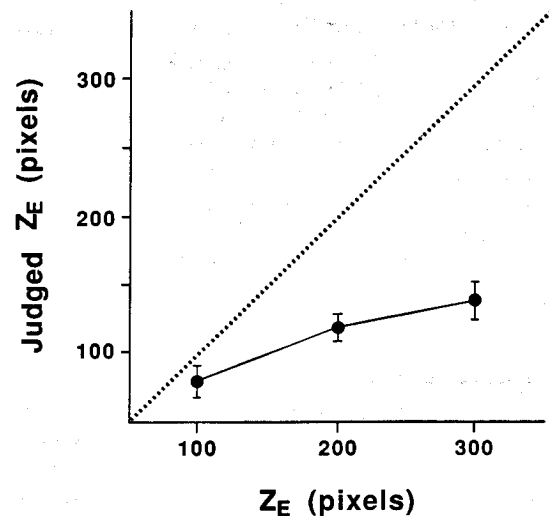


Figure 15. Average judged depth for 5 observers as a function of the actual depth  $Z_E$  collapsed over all of the other conditions in Experiment 4. (The dotted line represents perfect performance.)

in Figure 14, against a dark blue background on the right half of the display screen. (The shaded image of the ellipsoid remained in view.) The precise shape of this arc could be varied by manipulating the two parameters  $X_E$  and  $Z_E$ . The value of  $X_E$  was determined by the previous judgment of ordinal structure, but the value of  $Z_E$  could be controlled in real time by moving the mouse. The observers attempted to adjust the curve until it most closely matched the perceived metric structure of a horizontal cross-section through the depicted ellipsoid surface. This response was also recorded by pressing an appropriate button on the mouse, at which time a new trial was initiated.

The stimulus set included all possible combinations of  $X_E$ ,  $Z_E$ , and  $L$ , together with their mirror reflections for a total of 54 possible displays. These were presented four times each in a single experimental session. The observers were given several trials of practice prior to the actual experiment to familiarize themselves with the procedure. All of them reported at the conclusion of this practice that they felt comfortable with the task, although they all commented that the second of the two judgments was much more difficult.

### Results and Discussion

We begin with the observers' judgments of metric depth. Figure 15 shows the average perceived depth of the surfaces as a function of their actual depth ( $Z_E$ ), collapsed over all of the other conditions. An analysis of variance for these data revealed that the observers were significantly above chance in discriminating among the different possible depth values,  $F(2, 104) = 109.36, p < .001$ . It is important to keep in mind that variations in  $Z_E$  have no effect on ordinal structure. Thus, the observers' ability to discriminate these surfaces confirms the conclusion of Experiments 1 and 2 that ordinal structure cannot be the sole basis of our perceptual knowledge. It is also clear from the data, however, that in terms of their metric precision, the observers' estimates of surface depth were highly inaccurate. Indeed, the actual values of  $Z_E$  used to generate the displays were systematically underestimated by over 44% (see also Mingolla & Todd, 1986; Todd & Mingolla, 1983).

We now contrast these findings with the observers' judgments of ordinal structure. Figure 16 shows the average perceived horizontal position of the depth extrema as a function of their actual position ( $X_E$ ), collapsed over all of the other conditions. A correlational analysis of these data revealed that the observers' average judgments of ordinal structure were essentially perfect: The correlation of judged versus actual position was .961, the slope of the regression line was .954, and the intercept was  $-.206$ . In other words, whereas the metric depths of these surfaces were systematically underestimated by 44%, their ordinal depths could apparently be determined with almost no error at all.

A primary goal in designing Experiment 4 was to determine the effect of changes in illumination on observers' perceptions of ordinal structure. As was described earlier, our initial working hypothesis was that the perceived position of the depth extremum along any surface cross-section would be optically determined by the position of the luminance maximum on that cross-section. Because changing the direction of illumination significantly alters the position of the luminance maximum (see Figure 17), we would therefore expect such changes to significantly alter the perceived value of  $X_E$ . The results showed clearly, however, that this hypothesis is incorrect. The direction of illumination and its various interactions had almost no effect at all, accounting for only 4% of the between-displays variance, as opposed to more than 37% that would have been expected if the responses had all been positioned perfectly at the luminance maxima.

To appreciate how the observers compensated for changes in illumination, it is useful to consider a more specific example. Figure 17 shows the patterns of luminance along a horizontal cross-section of an ellipsoid surface for the three different directions of illumination used in the present experiment. Note in Figure 17 that there is a very simple (and remarkably general) rule for constraining the possible position of the depth extre-

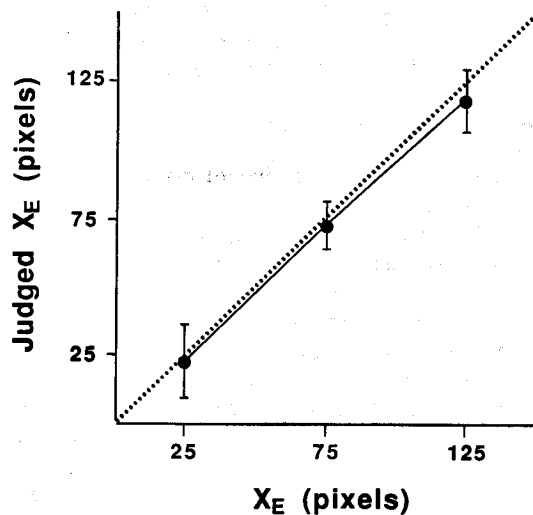


Figure 16. Average judged horizontal position of the depth extremum for 5 observers as a function of its actual value,  $X_E$ , collapsed over all of the other conditions in Experiment 4. (The dotted line represents perfect performance.)

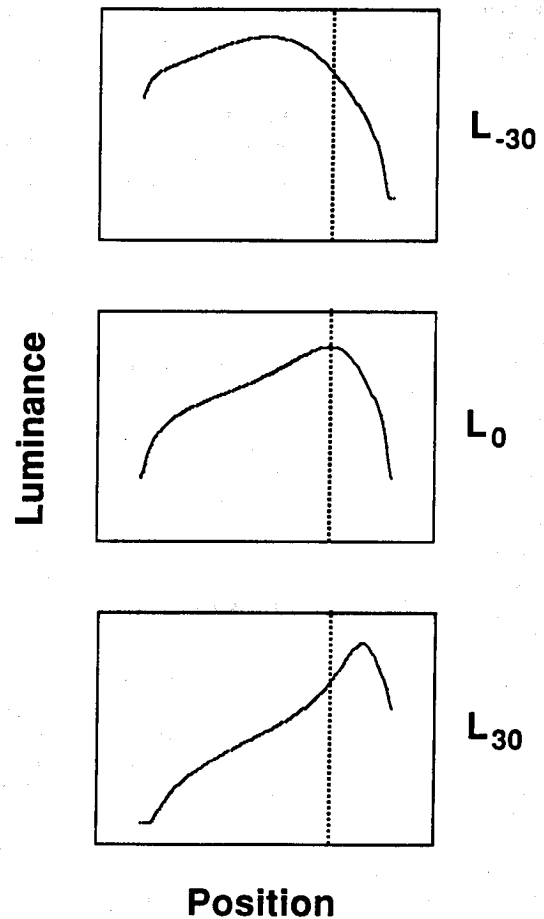


Figure 17. Pattern of luminance as a function of position for a horizontal cross-section of an ellipsoid surface ( $X_E = 75$ ,  $Z_E = 100$ ) with three different directions of illumination:  $L_{-30} = (-.5, .433, .75)$ ,  $L_0 = (0, .5, .866)$ ,  $L_{30} = (.5, .433, .75)$ . (The horizontal position of the depth extremum is represented in each curve by a vertical dotted line. Note that in the figure the displacement of the luminance maximum relative to the depth extremum is optically specified by the difference in luminance at the two occlusion points.)

num relative to the luminance maximum: If the two occlusion points on either side of a horizontal cross-section are of equal luminance, then the direction of illumination will be horizontally centered and the depth extremum will be coincident with the luminance maximum. If one of the occlusion points has a higher luminance than the other, then the direction of illumination will be tilted toward the lighter occlusion point and the depth extremum will be displaced away from the luminance maximum toward the darker occlusion point. Within broad limits, moreover, the size of this displacement will be positively correlated with the magnitude of the luminance difference at the two occlusion points, relative to the overall range in luminance.

We have experimented informally with a wide variety of equations in an effort to provide a more precise implementation of these general observations. Our best fit to the data has been obtained by estimating the specific luminance ( $I_E$ ) of the depth

extremum along each surface cross-section using the following equation:

$$l_E = l_o + (l_m - l_o) \sqrt{1 - ((l_l - l_o)/(l_m - l_o))^2},$$

where  $l_o$  is the luminance of the darker occlusion point,  $l_l$  is the luminance of the lighter occlusion point, and  $l_m$  is the luminance maximum along the surface cross-section. When the estimated luminances are localized in the image, their positions are highly correlated with the observers' judgments ( $r = .962$ ). It is important to keep in mind, however, that the details of this equation were arrived at through trial and error, with little or no theoretical justification. Thus, in the absence of more extensive data to test its generality, it is best to be cautious about drawing any firm conclusions concerning its relevance to human perception. It is presented here only to demonstrate the potential feasibility of estimating the position of a depth extremum from variations in luminance at points of occlusion.

### Discussion

The research described in the present article has been designed to explore some of the fundamental characteristics of our visual knowledge of smoothly curved surfaces. Most of the previous work in this area has been based on the idea that smoothly curved surfaces are perceptually represented using local mappings of metric depth and/or orientation relative to the point of observation. Indeed, from the dozens of articles that have been published in the literature on the computational analysis of metric structure from shading, texture, motion, or stereopsis, it would appear at first blush that the psychological relevance of this approach has a strong theoretical foundation. A closer examination reveals, however, that many of these analyses are based on extremely dubious assumptions. In the computational analysis of shape from shading, for example, it is typically assumed that an observed surface region has a homogeneous reflectance with homogeneous illumination and that there are no transparencies or specular highlights. Some models also require that the specific characteristics of surface reflectance and/or illumination must be determined in advance through prior knowledge. The problem, of course, in applying such models in an uncontrolled natural environment is that the required assumptions may seldom if ever be satisfied.

The psychological importance of metric representations can also be questioned on the basis of empirical data from human psychophysics. When observers are asked to make judgments about local depths or orientations of smooth surface regions (e.g., see Mingolla & Todd, 1986; Todd & Akerstrom, 1987), they generally report that the tasks are quite difficult and that they have little confidence in their judgments. With relatively few exceptions (e.g., Stevens, 1983; Stevens & Brookes, 1987), objective performance on these tasks is typically characterized by low reliability and relatively poor accuracy, even for displays that are perceptually quite compelling (e.g., see Figure 1).

Taken as a whole, these findings provide strong evidence that visual knowledge of local surface depth or orientation may be too coarse-grained to adequately account for the richness of our perceptual experiences. Thus, in an effort to overcome this deficiency, we have recently begun to explore some alternative nonmetric representations that, hopefully, might provide a

more complete understanding of the visual perception and cognition of smoothly curved surfaces (see also Hoffman & Richards, 1984; Koenderink & van Doorn, 1976, 1980, 1982; Stevens, 1981b). One such alternative involves local mappings of nonmetric order relations, in which neighboring surface regions are perceptually distinguished by which region is closer to the point of observation, without specifying how much closer (see Figure 3). This type of representation captures some but not all of the potentially measurable properties of an object's three-dimensional structure, but the available evidence suggests that its limitations are consistent with the perceptual capabilities of actual human observers.

One source of evidence for the psychological reality of ordinal representations was provided in Experiments 1 and 2 using a relative depth-judgment task. When an ordinal description of surface structure was sufficient to determine relative depth (i.e., for the ordinally transitive point pairs), observers exhibited a high level of performance; when an ordinal description was not sufficient (i.e., for the ordinally intransitive point pairs), performance was significantly impaired. A second source of evidence, as described in Experiment 4, involved judging the perceived position of a surface's depth extrema. It is important to note in this regard that the ordinal structure of a surface is only affected by spatial variations of its depth extremum that are parallel to the image plane; comparable variations perpendicular to the image plane significantly alter the metric structure of a surface but have no effect at all on its ordinal structure. A similar spatial anisotropy was also evident in the observers' judgments: The average perceived horizontal component of the depth extremum was almost perfectly correlated with its actual value, whereas the depth component was systematically underestimated by more than 44%.

In addition to analyzing the generic properties of ordinal representations, we have also considered some of the specific characteristics of shaded images by which ordinal structure is perceptually determined. Of particular importance in this regard is the role of smooth occlusion contours. It can be demonstrated theoretically (see Koenderink & van Doorn, 1976, 1982) that as we move from an occlusion contour in an attached region, the ordinal depth of a surface relative to the observer must decrease monotonically until a depth minimum is reached. Thus, occlusion contours provide potential information about the ordinal structure of attached surface regions in their immediate local neighborhood. The evidence suggests that human observers rely heavily on this information. We showed in Experiment 3, for example, that relative depth judgments are significantly impaired if the surface regions to be judged are not visibly connected to an attached occlusion contour. Similarly, if a surface is viewed from a particular orientation from which there are no occlusions, the perceptual salience of its three-dimensional structure is dramatically reduced (e.g., compare Figures 1, 4, and 11).

It is important to keep in mind while considering these phenomena that there are other varieties of image contours in addition to the self-occlusions of smoothly curved surfaces. Contours can arise from specular highlights, abrupt changes in surface reflectance, abrupt changes in illumination (e.g., cast shadows), or by discontinuities of surface orientation (e.g., the edges of a polyhedron). Because it is only attached occlusion



contours that provide information about ordinal structure, a critical problem for future research is to develop a theoretical analysis that can adequately distinguish one type of contour from another and, for occlusion contours, to adequately distinguish the attached and unattached regions.

Although occlusion contours may be a primary source of visual information, they are not by themselves sufficient for optically specifying the complete ordinal structure of a surface without also identifying the boundaries of ordinal transitivity (i.e., the depth extrema) in which a monotonic depth change switches from positive to negative, or vice versa. Patterns of image shading are particularly troublesome in this regard. For other sources of information such as texture, motion, or stereopsis, the depth extrema are always coincident with the maxima and minima of measurable image properties, such as size, velocity, or disparity. Patterns of shading are more complicated, however, because of the effects of direction of illumination. Thus, although the depth extrema of a surface cross-section are often in one-to-one correspondence with its luminance maxima, they will not in general be spatially coincident with one another.

Despite this theoretical difficulty, human observers are apparently quite good at judging the ordinal structure of a surface, even under conditions of variable illumination. In Experiment 4, for example, the horizontal positions of the depth extrema were identified with almost perfect accuracy over a 60° range in the direction of illumination. One possible source of information that may account for this phenomenon of illumination constancy is the variation in luminance at points of occlusion. If the variation is zero (i.e., all occlusion points have the same luminance), then the depth extrema will be coincident with the luminance maxima; as the variation increases, the depth extrema along any surface cross-section will be displaced farther and farther from the luminance maxima toward the darker occlusion points.

An additional complication for the analysis of ordinal structure from image shading is that the depth maxima and minima are locally indistinguishable and must therefore be disambiguated by some other source of information. There are several different biases that may influence whether a depth extremum is labeled as a maximum or a minimum: Illumination from above may be preferred over illumination from below (Berbaum, Bever, & Chung, 1984; Ramachandran, 1988); convex surfaces may be preferred over concave surfaces; and ground surfaces may be preferred over ceiling surfaces (Reichel & Todd, 1989). Occlusion contours can also provide perceptually useful information in this context. As we move from an occlusion point in an attached region, the first depth extremum encountered must be a minimum, the second a maximum, the third a minimum, and so on. The evidence suggests, moreover, that human observers may be quite sensitive to this information. Indeed, there have been several demonstrations reported in the literature of how the bounding contour of a surface can significantly influence its perceived pattern of relief (Barrow & Tenenbaum, 1981; Howard, 1982; Ramachandran, 1988; Reichel & Todd, 1989).

Although the present article has focused primarily on the ordinal structure of smoothly curved surfaces, it is important to keep in mind that our results show clearly that local order relations cannot be the sole basis of visual knowledge. In Experiment 1, for example, observers were able to determine the rela-

tive depths of ordinally intransitive point pairs with more than 88% accuracy. Similarly, in Experiment 4, they were able to perceptually distinguish between ellipsoid surfaces whose ordinal structures were identical. These findings suggest that visual knowledge may exist at multiple levels of description and that the predominant level for any given task is dependent on the particular judgment an observer is asked to perform. Our results indicate that performance is most difficult for tasks that require a metric knowledge of surface structure. Ordinal judgments, in contrast, are performed significantly faster and significantly more accurately; and, by extension, it is reasonable to suspect that performance may be improved still further for tasks such as object recognition involving nominal or categorical judgments (e.g., see Biederman, 1987).

There are several important issues that remain to be investigated concerning the potential interactions among these different levels of representation. The available evidence does not suggest, for example, whether they are all affected by the same sources of optical information or whether they must always be consistent with one another. The answers to these questions remain for future research.

## References

- Barrow, H. G., & Tenenbaum, J. M. (1981). Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence*, *17*, 75-116.
- Berbaum, K., Bever, T., & Chung, C. S. (1984). Extending the perception of shape from known to unknown shading. *Perception*, *13*, 479-488.
- Beusmans, J. M. H., Hoffman, D. D., & Bennett, B. M. (1987). Description of solid shape and its inference from occluding contours. *Journal of the Optical Society of America A*, *4*, 1166-1167.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-147.
- Braunstein, M. L. (1976). *Depth perception through motion*. New York: Academic Press.
- Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General*, *113*, 198-216.
- Gibson, J. J. (1950). The perception of visual surfaces. *American Journal of Psychology*, *63*, 367-384.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Hoffman, D. D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, *18*, 65-96.
- Horn, B. K. P. (1975). Obtaining shape from shading information. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 115-156). New York: McGraw Hill.
- Horn, B. K. P. (1977). Understanding image intensities. *Artificial Intelligence*, *8*, 201-231.
- Howard, I. P. (1982). Occluding edges in apparent reversal of convexity and concavity. *Perception*, *12*, 85-86.
- Ikeuchi, K., & Horn, B. K. P. (1981). Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, *17*, 141-184.
- Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception*, *13*, 321-330.
- Koenderink, J. J., & van Doorn, A. J. (1976). The singularities of the visual mapping. *Biological Cybernetics*, *24*, 51-59.
- Koenderink, J. J., & van Doorn, A. J. (1980). Photometric invariants related to solid shape. *Optica Acta*, *27*, 981-996.
- Koenderink, J. J., & van Doorn, A. J. (1982). The shape of smooth objects and the way contours end. *Perception*, *11*, 129-137.

- Lee, C. H., & Rosenfeld, A. (1985). Improved methods of estimating shape from shading using the light source coordinate system. *Artificial Intelligence*, 26, 125-143.
- Marr, D. (1982). *Vision*. San Francisco: Freeman.
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of three dimensional shapes. *Proceedings of the Royal Society of London, Series B*, 197, 441-475.
- Mingolla, E., & Todd, J. T. (1984). Computational techniques for the graphic simulation of quadric surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 740-745.
- Mingolla, E., & Todd, J. T. (1986). Perception of solid shape from shading. *Biological Cybernetics*, 53, 137-151.
- Pentland, A. P. (1984). Local shading analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 170-187.
- Ramachandran, V. S. (1988). Perceiving shape from shading. *Scientific American*, 252, 76-83.
- Reichel, F. D., & Todd, J. T. (1989). Perceived depth inversion of smoothly curved surfaces. *Investigative Ophthalmology*, 30, 252.
- Richards, W. A., Koenderink, J. J., & Hoffman, D. D. (1987). Inferring three-dimensional shapes from two-dimensional silhouettes. *Journal of the Optical Society of America A*, 4, 1168-1175.
- Stevens, K. A. (1981a). The information content of texture gradients. *Biological Cybernetics*, 42, 95-105.
- Stevens, K. A. (1981b). The visual interpretation of surface contours. *Artificial Intelligence*, 17, 47-74.
- Stevens, K. A. (1983). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics*, 46, 183-195.
- Stevens, K. A., & Brookes, A. (1987). Probing depth in monocular images. *Biological Cybernetics*, 56, 355-366.
- Todd, J. T. (1989). [Ordinal structure in the visual perception of structure from motion.] Unpublished raw data.
- Todd, J. T., & Akerstrom, R. A. (1987). Perception of three-dimensional form patterns of optical texture. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 242-255.
- Todd, J. T., & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 583-595.
- Todd, J. T., Reichel, F. D., & Mingolla, E. (1986). Perception of curved surfaces from shading and occlusion contours. *Bulletin of the Psychonomic Society*, 24, 349.
- Witkin, A. P. (1981). Estimating shape from texture. *Artificial Intelligence*, 17, 17-45.

Received August 9, 1988

Revision received January 25, 1989

Accepted March 3, 1989 ■