10

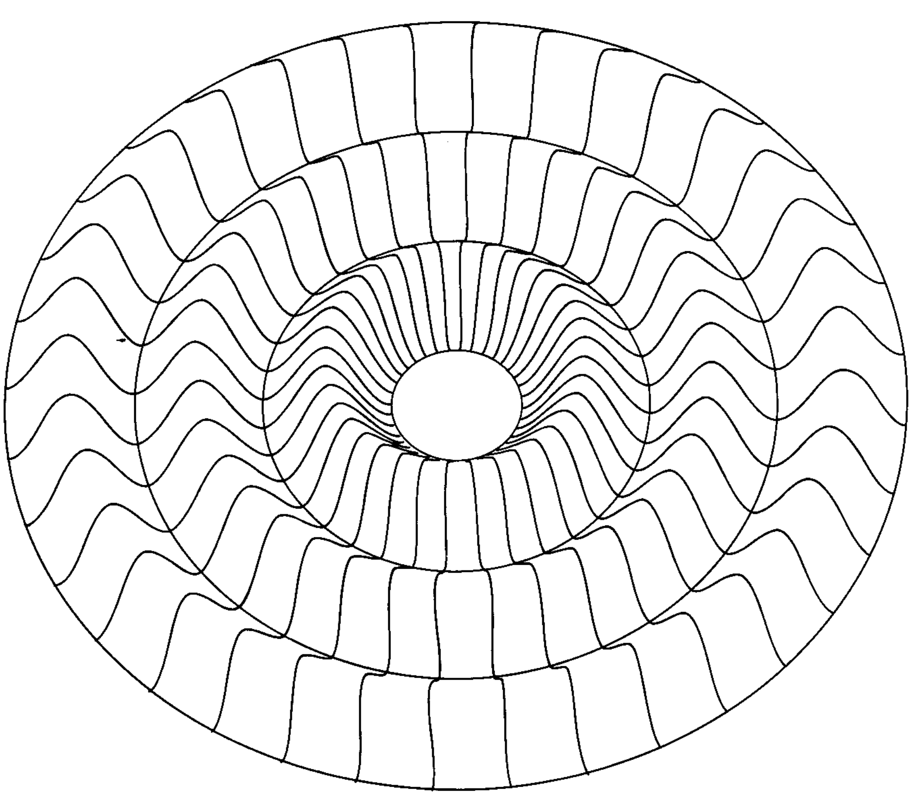
The Interpretation of Visual Illusions

by Donald D. Hoffman
December 1983

The visual system apparently organizes ambiguous retinal images according to rules of inference that exploit certain regularities in the external world

Vision is a process of inference. What you see when you look around depends not only on what is there to be seen but also on how your visual system organizes and interprets the images that fall on your retinas. An intriguing demonstration of this aspect of perception is presented by the apparent surface that is formed by rotating a cosine wave around a vertical axis viewed obliquely [*see illustration on this page*]. When you first look at the figure, it appears to be organized into a set of raised concentric rings, with the boundaries between the rings delineated approximately by the colored circular contours. If you turn the page upside down, however, the organization changes: now each colored contour, instead of lying in a trough between two rings, appears to trace the crest of a ring. (Try it.) Evidently the visual system does more than passively transmit signals to the brain. It actively takes part in organizing and interpreting them.

This finding raises three questions. First, why does the visual system need to organize and interpret the images formed on the retinas? Second, how does it remain true to the real world in the process? Third, what rules of inference does it follow? The answers to these questions call for a closer examination of such figures.



AMBIGUOUS SURFACE is made by rotating a cosine wave about a vertical axis. The surface initially appears to be organized into raised concentric rings, with the colored circular contours lying in the troughs between the rings. When the page is turned upside down, however, the organization appears to change: each colored contour is now seen to trace the crest of a ring.

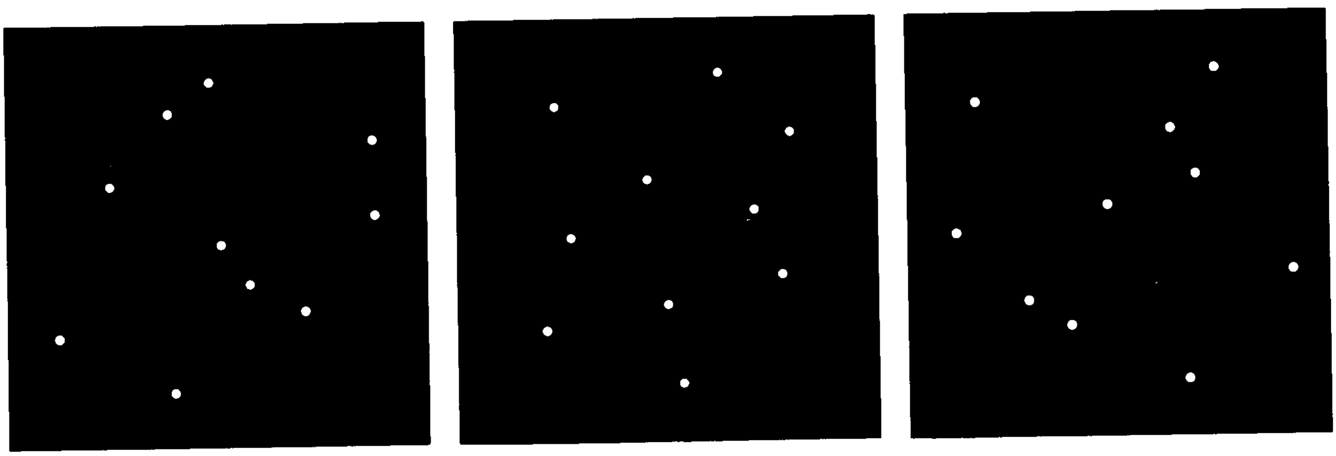
One reason the visual system organizes and interprets retinal images is simply that many possible configurations in the real world are consistent with any given retinal image. In other words, retinal images need organization and interpretation because they are fundamentally ambiguous. Their ambiguity is due in part to the fact that the world is three-dimensional and each retina is essentially two-dimensional. To describe the world in its full three-dimensional glory necessarily involves some rather sophisticated inferences by the visual system, inferences that for the most part proceed without any conscious awareness. For example, the cosine surface at the right, like your retinal image of it, is two-dimensional. Yet it appears, quite compellingly, as three-dimensional. The appearance of depth is entirely inferred, or, to put it another way, hallucinated. This conclusion should be cause for some concern. If, as I suggest, such hallucinations are not an exception but the rule, and if they are in fact a necessary concomitant of visual perception, how can one justify one's faith in perception? How is it still possible that in general seeing is believing?

What is needed for an understanding of vision, therefore, is an explanation of why such visual inferences usually bear a nonarbitrary relation to the real world. A promising line of investigation begins with the observation that the visible world, far from being completely chaotic, obeys certain laws and exhibits numerous regularities. If the visual system is adapted to exploit these laws and regularities in its organization and interpretation of retinal images, and if it is constrained somehow to prefer the interpretation that is most credible, given both the image and a knowledge of these laws and regularities, then it might be possible to understand how it is that one's visual hallucinations bear a nonarbitrary and even useful relation to the external world.

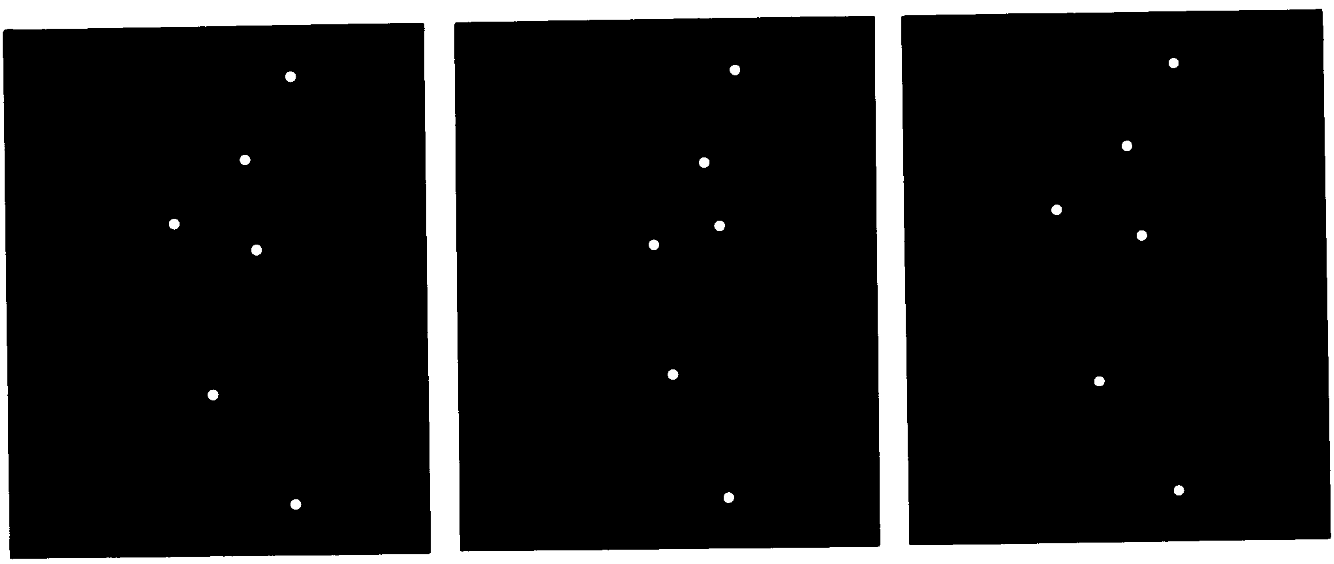
A particularly clear example of this approach is the research into visual motion perception done by Shimon Ullman of the Massachusetts Institute of Technology. Ullman has explored the remarkable ability of the human visual system to perceive the correct three-dimensional structure and motion of an object solely from its moving two-dimensional projection, an ability Hans Wallach and Donald N. O'Connell of Swarthmore College call the kinetic-depth effect. For instance, if a transparent beach ball with tiny light bulbs mounted randomly on its surface is set spinning in a dark room, one immediately perceives the correct spherical layout of the lights [*see upper illustration below*]. When the spinning stops, so does the perception of the spherical array. How does one see the correct three-dimensional structure when infinitely many three-dimensional structures are consistent with the moving two-dimensional retinal projection? Ullman showed mathematically that if the visual system exploits the laws of projection, and if it exploits the fact that the world contains rigid objects, then in principle a unique and correct interpretation can be obtained. In particular he showed that three views of four noncoplanar light bulbs are enough to solve the problem. The key point is that an inference rule, based on a law (the law of projection) and a regularity (namely the fact that the world includes rigid objects), enables the visual system to make a correct interpretation.

At this stage, however, a puzzle arises. The same mathematical precision that shows the rigidity regularity is sufficient in principle to interpret the rotating beach ball also shows the rigidity regularity by itself is insufficient to interpret a similar display. This display was first devised by Gunnar Johansson of the University of Uppsala as an example of what he calls biological motion [see "Visual Motion Perception," by Gunnar Johansson; SCIENTIFIC AMERICAN, June, 1975]. Johansson put small light bulbs on the major joints of a person and took motion pictures as the person moved about in a dark room. A single frame of such a film looks like a random collection of white dots on a black background. When the film is set in motion, however, one immediately sees the correct three-dimensional structure of the dots and recognizes that there is an invisible person walking about [*see lower illustration below*].

When my colleague Bruce E. Flinchbaugh, who is now at Bell Laboratories, and I considered this problem, what puzzled us was that it is possible to see the correct three-dimensional structure even though, according to Ullman's results, one lacks the appropriate information to do so. To infer a correct three-dimensional structure on the basis of the rigidity regularity it is necessary to have three snapshots of at least four nonco-



ROTATING SPHERE is seen when the three dot patterns represented here are shown in rapid succession. The visual system seems to be adopting the most rigid three-dimensional interpretation for the moving dots that is consistent with the two-dimensional projections.



WALKING PERSON is seen when these dot patterns are shown in rapid succession. In this case the visual system seems to adopt the most rigid and planar three-dimensional interpretation that is consistent with the two-dimensional motions of the dots. The display is based on an experiment conducted by Gunnar Johansson of the University of Uppsala in which small light bulbs were put on a person's major joints (shoulder, elbow, wrist, hip, knee and ankle) and a motion picture was made as the person moved about in a dark room.

planar points in a rigid configuration. In biological motion displays, on the other hand, at best only pairs of points are rigidly connected, such as the ankle and the knee or the knee and the hip. Rigid quadruplets of points just do not exist.

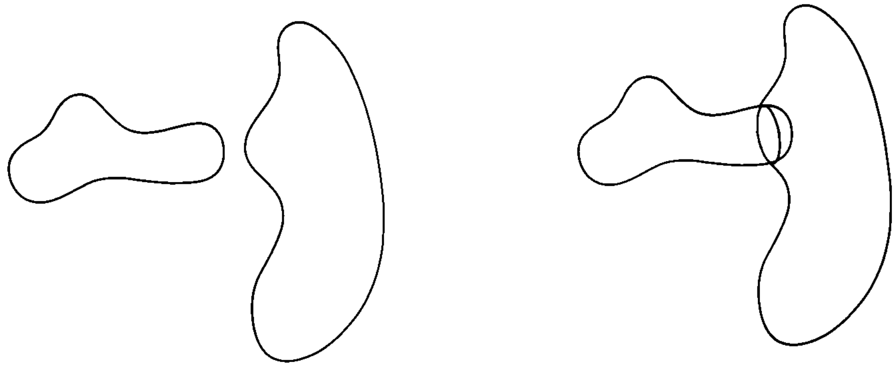
The rigidity regularity, then, is insufficient by itself, leading us to ask: What further regularity might the visual system be exploiting? After several false starts it occurred to us there is an anatomical regularity that might do the trick. Each weight-bearing limb of most animals is constrained, because of the construction of its joints, to swing in a single plane in a normal gait. We call this the planarity regularity.

In fact, the planarity regularity is sufficient to correctly interpret biological-motion displays of gait. The correct three-dimensional structure can be inferred either from three snapshots of two points swinging rigidly in a plane or from two snapshots of three points (such as an ankle, a knee and a hip) forming rigid pairs and swinging in one plane. These results comport nicely with Johansson's observation that only two or three frames of his films need be seen before subjects correctly perceive the biological motion. In addition it turns out not only that all three-dimensional motions governed by the planarity regularity can be given a correct interpretation but also that whenever an interpretation is found for image motion based on the planarity regularity or the rigidity regularity the interpretation is correct.

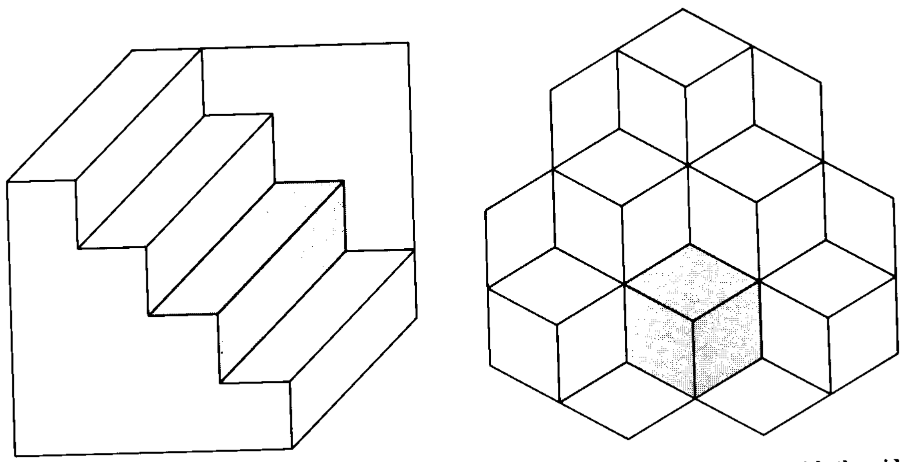
In short, the probability that the interpretation is wrong is zero, assuming infinite resolution in the image, or slightly greater than zero given less than perfect resolution. Hence nonrigid structures cannot masquerade as rigid ones, and nonplanar motions cannot be misconstrued as planar ones. Once again laws and regularities prove to be central in explaining how the visual system achieves a unique and correct interpretation of a retinal image.

Let us now return to the cosine surface. Its main interest is that it reveals the visual system organizing shapes into parts, an organization that is quite useful for the task of recognizing an object from its shape. The cosine surface also reveals that turning a shape upside down can alter this organization. Is the visual system, then, capricious in its organization? That is unlikely. If it is not governed by whim, however, it must be governed by rules for defining parts. And if the rules are not to be arbitrary, they must be grounded in some law or regularity in the external world.

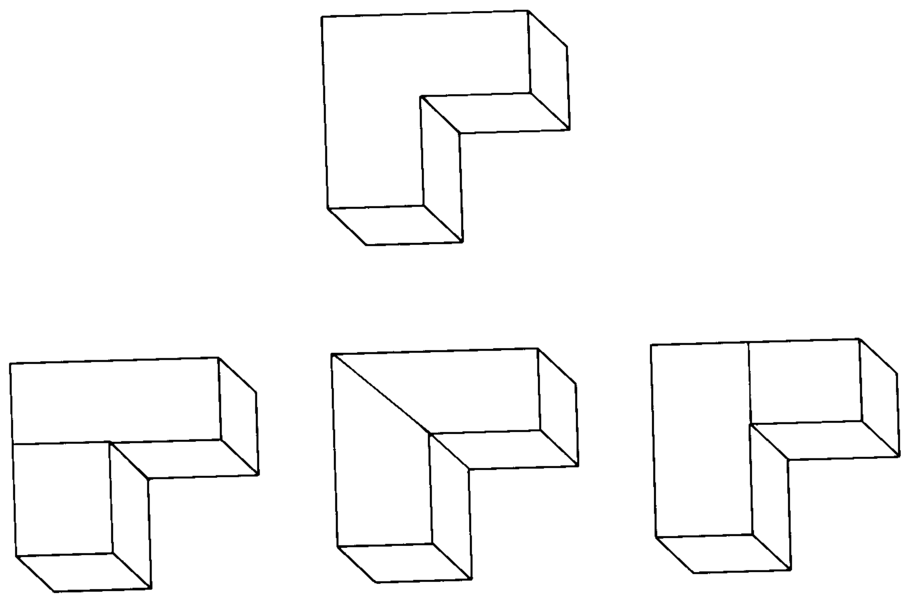
This line of reasoning led Whitman A. Richards of M.I.T. and me to seek a law or regularity that could motivate a set of rules for partitioning surfaces. The regularity we found to be relevant is the following transversality regularity:



TRANSVERSALITY, a kind of regularity commonly observed in the external world, underlies a unified account of several visual illusions. According to the rule of transversality (as defined by Whitman A. Richards and the author), when any two surfaces penetrate each other at random, they always meet at a concave discontinuity, indicated here by the colored contour.

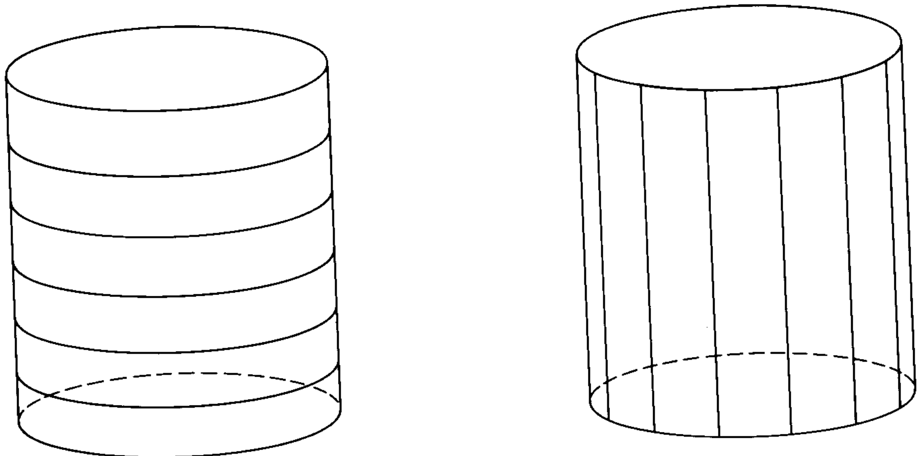


PARTITIONING RULE based on the transversality regularity is demonstrated with the aid of these two figures that reverse when they are looked at steadily. In both cases the apparent boundaries of the different parts of the perceived shape change when the "figure" becomes the "ground" and vice versa. For example, in the case of the reversing-stairway illusion (*left*), first published by H. Schröder in 1858, the two colored panels, which in one view appear to be parts of one step, suddenly seem to be parts of two adjacent steps when the stairway reverses. Similarly, in the stacked-cube illusion (*right*) the three diamond-shaped colored panels can be seen either as the faces of one cube or, when the figure reverses, as the faces of three different cubes.



ELBOW-SHAPED BLOCKS show that the rule partitioning shapes at concave discontinuities is appropriately conservative. The rule does not give a closed contour on the top block because three different perceived partitions seem possible, as illustrated by the bottom three blocks.

When two arbitrarily shaped surfaces are made to penetrate each other at random, they always meet at a contour of concave discontinuity of their tangent planes [*see top illustration on page 97*]. Although the transversality regularity may sound esoteric, it is actually a familiar part of everyday experience. A straw in a soft drink, for instance, forms a circular concave discontinuity where it meets the surface of the drink. A candle in a birthday cake, the tines of a fork in a piece of steak, a cigarette in a mouth—all are examples of this ubiquitous regularity.



LINES OF CURVATURE are easily pictured on an idealized cylindrical drinking glass. The lines of greatest curvature (*left*) are circles; the lines of least curvature (*right*) are straight lines.

On the basis of the transversality regularity one can propose a first rule for partitioning a surface: Divide a surface into parts along all contours of concave discontinuity. This rule cannot help with the cosine surface because it is entirely smooth. The rule must first be generalized somewhat, as will be done below. In its nongeneralized form, however, it can elucidate several well-known perceptual demonstrations.

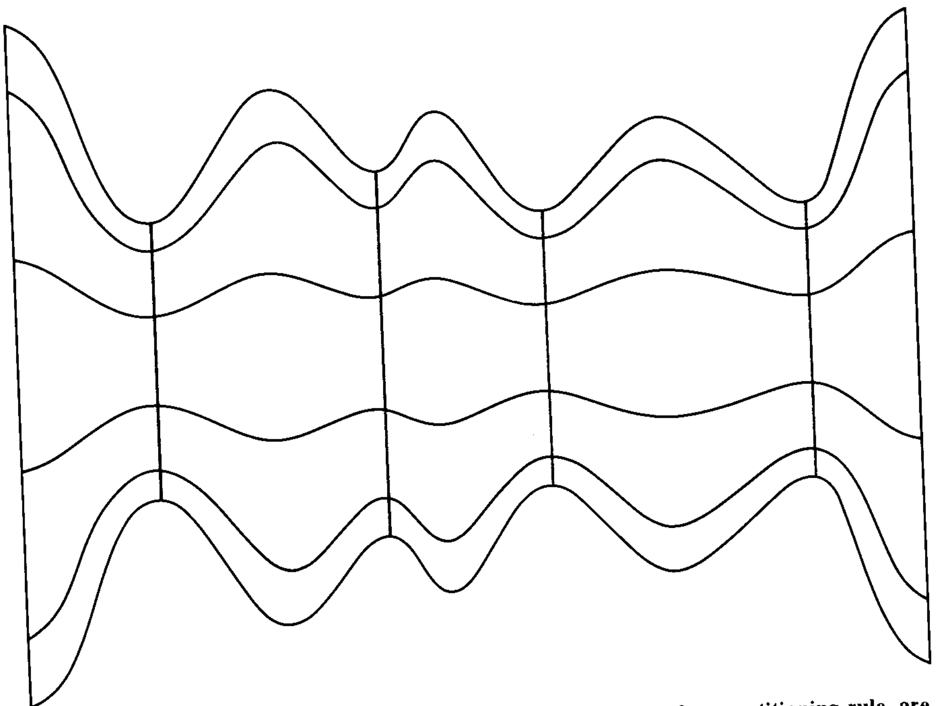
For example, the rule makes the obvious prediction that the parts of the staircase shown in the middle illustration on page 97 are its steps, each step lying between two successive lines of concave discontinuity in the staircase. The rule also makes a less obvious prediction. If the staircase undergoes a perceptual reversal, such that the "figure" side becomes "ground" and vice versa, then the step boundaries must change. This conclusion follows because only concave discontinuities define the step boundaries, and what looks like a concavity from one side of a surface must look like a convexity from the other. Thus when the staircase reverses, convex and concave discontinuities must reverse roles, leading to new step boundaries. You can test this prediction yourself by looking at the step that has color on each of its two faces. When the staircase appears to reverse, note that the colored panels are no longer on a single step but rather on adjacent steps.

This prediction can be confirmed with a more complicated demonstration such as the stacked-cubes test seen in the same illustration. The three colored faces, which at first appear to be on one cube, are seen to be on three cubes when the figure reverses.

A further prediction follows from this simple partitioning rule. If the rule does not define a unique partition of some surface, then the appropriate way to divide the surface into parts should be perceptually ambiguous (unless there are additional rules that can eliminate the ambiguity). A clear confirmation of this prediction can be seen with reference to the elbow-shaped block in the bottom illustration on page 97. The only concave discontinuity is the vertical line in the crook of the elbow. As a consequence the rule does not define a unique partition of the block. Perceptually there are three plausible ways to cut the block into parts. All three ways rely on the contour defined by the partitioning rule, but they complete it along different paths.

Even this simple partitioning rule leads to interesting insights into the perception of shape. To explore the cosine surface and other smooth surfaces, however, the rule must be generalized. This requires a brief digression into the differential geometry of surfaces in order to understand three important concepts: surface normal, principal curvature and line of curvature. Fortunately, although these concepts are quite technical, they can readily be given an intuitive characterization.

The surface normal at a point on a surface can be thought of as a needle of unit length sticking straight out of the surface at that point, much like the spines on a sea urchin. All the surface normals at all points on a surface are collectively called a field of surface normals. Usually there are two possible fields of surface normals on a surface; they can be either outward-pointing or inward-pointing. For example, a sphere can have the surface normals all pointing radially out like spines or all pointing in toward its center. Let us adopt the convention that the field of surface normals is always chosen to point into the figure. Thus a baseball has inward normals whereas a bubble under water has outward normals. Reversing the choice of figure and ground on a surface im-



PART BOUNDARIES, as defined by the generalized, smooth-surface partitioning rule, are represented by the colored contours on this arbitrarily shaped surface. The black lines are the lines of greatest curvature whose minimums give rise to the colored partitioning contours.

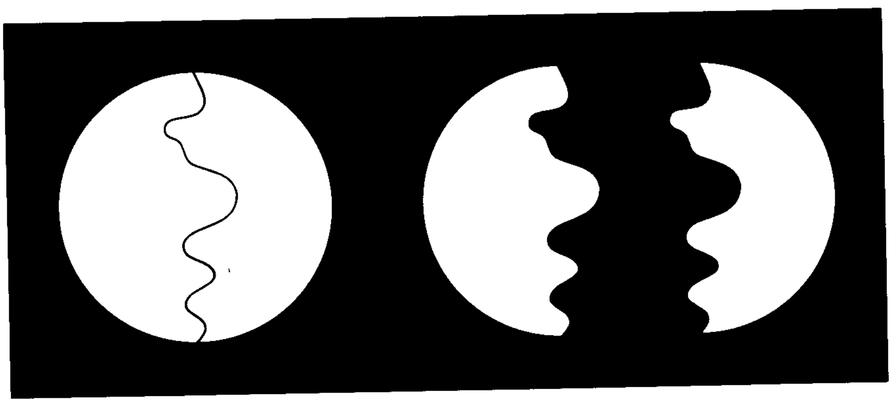
plies a concomitant change in the surface normals. A reversal of the field of surface normals induces a change in sign of each of the principal curvatures at every point on the surface.

It is often important to know not only the surface normal at a point but also how the surface is curving at the point. The 18th-century Swiss mathematician Leonhard Euler discovered that at any point on any surface there is always a direction in which the surface curves the least and a second direction, always at right angles to the first, in which the surface curves the most. (In the case of a plane or a sphere the surface curvature is identical in all directions at every point.) These two directions are called principal directions, and the corresponding surface curvatures are called principal curvatures. By starting at some point and always moving in the direction of the greatest principal curvature one traces out a line of greatest curvature. By moving instead in the direction of the least principal curvature one traces out a line of least curvature. On a drinking glass the family of lines of greatest curvature is a set of circles around the glass. The lines of least curvature are straight lines running the length of the glass [*see top illustration on preceding page*].

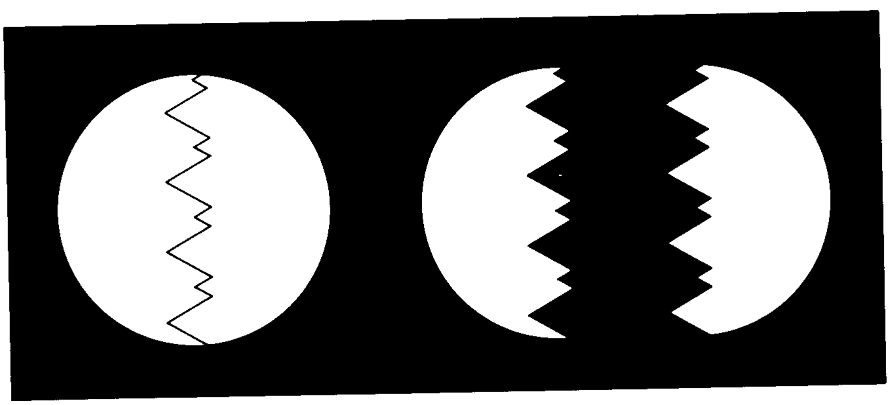
With these concepts in mind the transversality regularity extends easily to smooth surfaces. Suppose wherever a surface has a concave discontinuity one smoothes the discontinuity somewhat, perhaps by stretching a taut skin over it. Then a concave discontinuity becomes, roughly speaking, a contour where the surface has locally the greatest negative curvature. More precisely, the generalized version of transversality suggests the following generalized partitioning rule for surfaces: Divide a surface into parts at negative minimums of each principal curvature along its associated family of lines of curvature [*see bottom illustration on preceding page*].

This rule partitions the cosine surface along the colored circular contours. It also explains why the parts are different when the page is turned upside down: the visual system then reverses its assignment of figure and ground on the surface (perhaps owing to a preference for an interpretation that places the object below the observer's viewpoint rather than above it). When figure and ground reverse, so does the field of surface normals, in accordance with the convention mentioned above. Simple calculations show, however, that when the normals reverse, so does the sign of the principal curvatures. As a result minimums of the principal curvatures must become maximums and vice versa. Since minimums of the principal curvatures are used for part boundaries, it follows that these part boundaries must also move. In sum, parts appear to change because the partitioning rule, motivated by the transversality regularity, uses minimums of the principal curvatures, and because these minimums relocate on the surface when figure and ground reverse.

The transversality regularity, in short, provides an underlying unity for explanatory accounts of the perception of parts in both smooth and rough surfaces. It also underlies an explanation of another well-known class of visual illusions: reversing plane curves. A good example of this phenomenon is the reversing figure devised by Fred Attneave of the University of Oregon [*see upper illustration on this page*]. He found that by simply scribbling a line through a circle and separating the two halves one can create two very different-looking contours. Evidently, as Attneave points out, the appearance of the contour depends on which side is taken to be part of the figure, not on any prior familiarity with the contour [see "Multistability in Perception," by Fred Attneave; SCIENTIFIC AMERICAN, December, 1971].



REVERSING PLANE CURVE, constructed by Fred Attneave of the University of Oregon by scribbling a line through a circle and separating the two halves, shows that the apparent shape of the resulting contour depends on which side of the line is perceived as the figure.

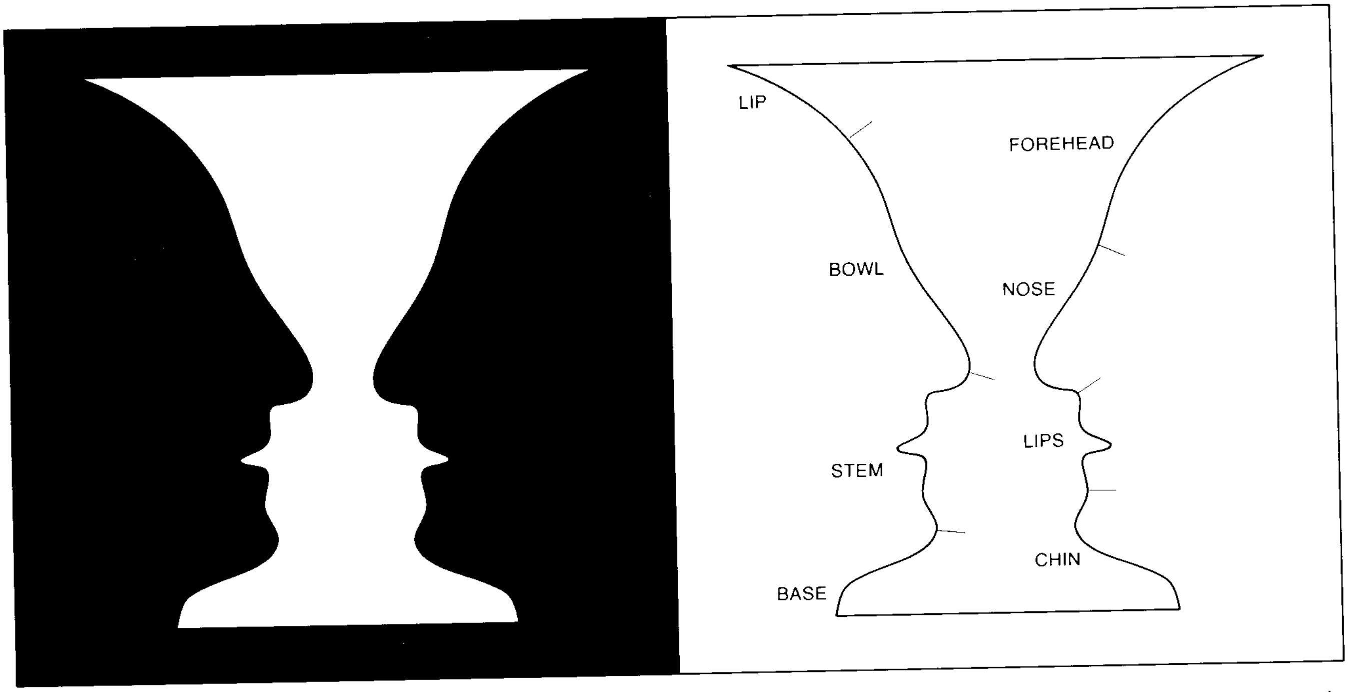


SIMILAR REVERSING FIGURE can be made with a plane curve that is not smooth. One can see the resulting jagged contour either as an alternating chain of tall and short mountains or, in the reversed figure-ground assignment, as a chain of tall mountains with twin peaks.

How does the transversality regularity explain this phenomenon? The answer involves three steps: (1) a projection of the transversality regularity from three dimensions onto two dimensions, (2) a brief digression on the differential geometry of plane curves and (3) the formulation of a partitioning rule for plane curves.

The two-dimensional version of the transversality regularity is similar to the three-dimensional version. If two arbitrarily shaped surfaces are made to penetrate each other at random, then in any two-dimensional projection of their composite surface they will always meet in concave cusps. To paraphrase it loosely, concave cusps are always formed in a silhouette at points where one part stops and another begins. This suggests the following partitioning rule for plane curves: Divide a plane curve into parts at concave cusps. This rule cannot apply to Attneave's demonstration because his demonstration relies on a contour that is everywhere smooth. The rule must again be generalized. Nevertheless, in its nongeneralized form it can account for a version of Attneave's demonstration that is not everywhere smooth.

In the lower illustration at the left the same jagged contour can look either like an alternating chain of tall and short mountains or, for the reversed



FACE-GOBLET ILLUSION, devised by Edgar Rubin in about 1915, can be seen either as a pair of facial profiles or as a goblet (*left*). If a face is taken to be the figure, partitioning the figure by reference to minimums of curvature divides the contour into chunks corresponding to a forehead, a nose, a pair of lips and a chin; if the goblet is taken to be the figure, defining the part boundaries by minimums of curvature divides the contour into a lip, a bowl, a stem and a base (*right*). Principal-normal lines in both cases point into the figure.

figure-ground assignment, like a chain of tall mountains with twin peaks. The contour is carved into parts differently when figure and ground reverse because the partitioning rule uses only concave cusps for part boundaries. What is a concave cusp if one side of the contour is figure must become a convex cusp when the other side is figure, and vice versa. There is a parallel between this example and the reversible staircase discussed above.

Before generalizing the rule to smooth contours let us briefly review two concepts from the differential geometry of plane curves: principal normal and curvature. The principal normal at a point on a curve can be thought of as a unit-length needle sticking straight out of the curve at that point. All the principal normals at all points on a curve together form a field of principal normals. Usually there are two possible fields of principal normals, one field on each side of the curve. Let us adopt the convention that the field of principal normals is always chosen to point into the figure side of the curve. Reversing the choice of figure and ground on a curve implies a concomitant change in the field of principal normals. What is important to note is that because of the convention forcing the principal normals to point into the figure, concave parts of a smooth curve have negative curvature and convex parts have positive curvature.

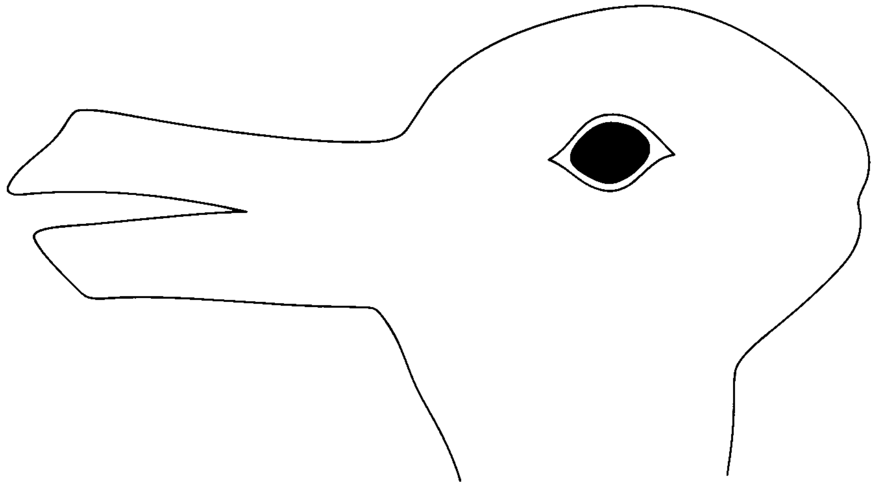
It is an easy matter now to generalize the partitioning rule for plane curves. Suppose wherever a curve has a concave cusp one smoothes the curve a bit. Then a concave cusp becomes a point of negative curvature having, locally, the greatest absolute value of curvature. This observation leads to the following generalized partitioning rule: Divide a plane curve into parts at negative minimums of curvature.

Now it is possible to explain why the two halves of Attneave's circle look so different. When figure and ground reverse, the field of principal normals also reverses in accordance with the convention, and when the principal normals reverse, the curvature at every point on the curve must change sign. In particular, minimums of curvature must become maximums and vice versa. This repositioning of the minimums of curvature leads to a new partitioning of the curve by the partitioning rule. In short, the curve looks different because it is organized into fundamentally different chunks, or units. Note that if one chooses to define part boundaries by inflections, or by both maximums and minimums of curvature, then the chunks would not change when figure and ground reverse.

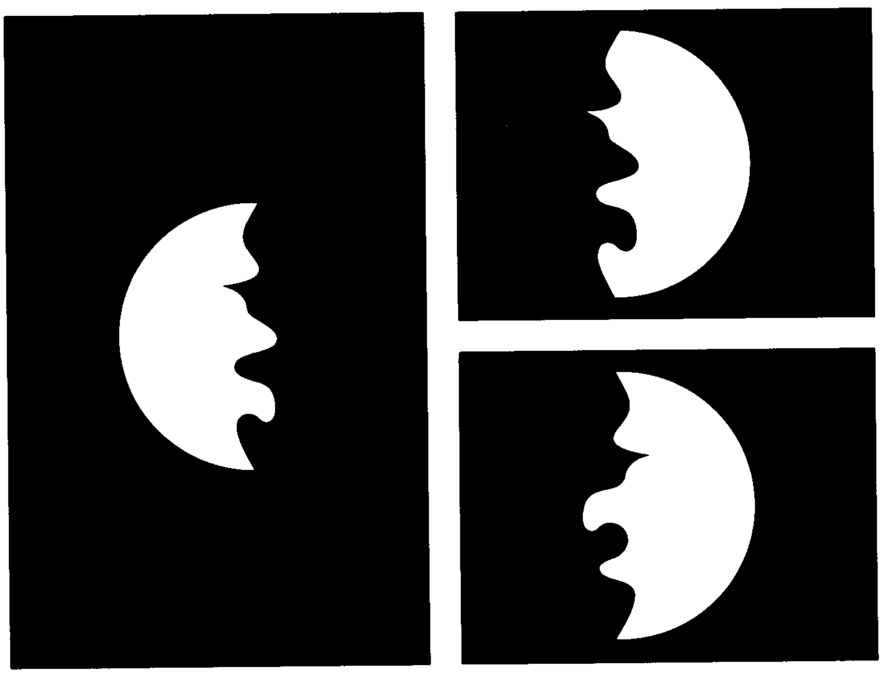
A clear example of two very different partitions for one curve can be seen in the famous face-goblet illusion devised by Edgar Rubin in about 1915 [*see illustration above*]. If a face is taken to be figure, the minimums of curvature divide the curve into chunks corresponding to a forehead, nose, upper lip, lower lip and chin. If instead the goblet is taken to be figure, the minimums are repositioned, dividing the curve into new chunks corresponding to a base, a couple of parts of the stem, a bowl and a lip on the bowl. It is probably no accident that the parts defined by minimums are often easily assigned verbal labels.

Demonstrations have been devised that, like the face-goblet illusion, allow more than one interpretation of a single contour but that do not involve a figure-ground reversal. A popular example is the rabbit-duck illusion [*see top illustration on next page*]. Because such illusions do not involve a figure-ground reversal, and because as a result the minimums of curvature never change position, the partitioning rule must predict that the part boundaries are identical for both interpretations of each of the contours. This prediction is easily confirmed. What is an ear on the rabbit, say, becomes part of a bill on the duck.

If the minimums rule for partitioning curves is really obeyed by the human visual system, one would expect it to predict some judgments of shape similarity. One case in which its prediction is counterintuitive can be seen in the bottom illustration on the next page. Look briefly at the single half-moon on the left side of the illustration. Then look quickly at the two half-moons at the right and decide which one seems more like the first one. In an experiment done on several similar figures, Aaron F. Bobick of M.I.T. and I found that almost all subjects chose the half-moon at the lower right as the more similar one. Yet if you look again, you will find that the bounding contour for the half-moon at



REVERSING-ANIMAL ILLUSION does not involve a reversal of figure and ground. Accordingly the part boundaries defined by the minimums of curvature do not change position when the interpretation changes. The rabbit's ears turn into the duck's bill without moving.



HALF-MOON TEST demonstrates that judgments of the similarity of shapes can be correctly predicted by the minimums-of-curvature partitioning rule. At first glance the half-moon at the lower right seems to resemble the single half-moon at the left more than the one at the upper right does. Closer inspection, however, reveals that the bounding contour of the upper-right half-moon is identical with that of the half-moon at the left, whereas the bounding contour of the lower-right half-moon has been mirror reversed and has also had two parts interchanged.

the upper right is identical with that of the left half-moon, only it is figure-ground reversed. The bounding contour of the lower half-moon has been mirror reversed, and two parts defined by minimums of curvature have been swapped. Why does the lower one still look more similar? The minimums rule gives a simple answer. The lower contour, which is not figure-ground reversed from the original contour, has the same part boundaries. The upper contour, which is figure-ground reversed from the original, has different part boundaries.

In summary, vision is an active process whose function is to infer useful descriptions of the world from changing patterns of light falling on the retinas. These descriptions are reliable only to the extent that the inferential processes building them exploit regularities in the visual world, such as rigidity, planarity and transversality. The discovery of relevant regularities and the mathematical investigation of their power in guiding visual inferences are promising directions for the investigator seeking an understanding of human vision.