

Contents

1	Seeing	5
1.1	Miracle Cures	6
1.2	Illusions	9
1.3	Summary	21

List of Figures

1.1	Shepard Table/Depth Reconstruction	10
1.2	Boring size illusion	11
1.3	Adelson Transparency/Brightness Illusion	12
1.4	Craik-O'Brien-Cornsweet Illusion	14
1.5	Rotated Volcanoes	15
1.6	Fraser Spiral	16
1.7	Kanizsa Figures	17
1.8	Occlusion Helps us to See Parts as Unified	19
1.9	Margaret Thatcher Illusion	20
1.10	Gregory Muller-Lyer	22

Chapter 1

Seeing

Seeing is a collection of inferences about the world. Motion, color and depth are important individual judgments. To see, however, we must connect these inferences into a unified explanation of the image. Until we integrate the separate inferences of pattern, color, motion and depth into a description of objects and surfaces, the world remains a disconcerting jumble of unconnected events.

It is easy to recognize the importance of integrating our visual inferences into a coherent view of the scene, but it is much harder to understand the process by which we perceive objects and surfaces. Because there is no current consensus on a theoretical approach to this topic, I have chosen to spend this chapter reviewing phenomena that I believe will be important in defining a computational theory of seeing.

In the first section of this chapter I will discuss clinical cases that illustrate the importance of being able to integrate information from different locations within an image and images acquired at different times. To those who are sighted from birth, the ability to integrate image information acquired from different viewpoints at different points in time is easy and automatic. As we walk about, we see a single object and not a collection of independent images. The computational complexity of the visual inference that integrates the different information is made quite plain, however, when we read about the difficulties of patients who were blind as infants but “cured” later in life. The tragic stories of these individuals, as they struggle to learn to see, provide us with some understanding of the complexity of object perception.

In the second section of this chapter, I will review a set of visual illusions. Visual illusions help us understand how the visual pathways organize images into objects. I have selected a series of illusions that show how the visual system uses integrates image information concerning occlusion, transparency, and boundaries to integrate judgments of brightness and shape. While I have mentioned the significance of

many of these aspects of vision in earlier chapters, the illusions we will review here provide some clues about the rules for combining visual inferences into a complete description of the scene. And, there is a second reason for devoting this time to studying illusions: they are fun.

1.1 Miracle Cures

In 1963 Gregory and Wallace wrote a monograph describing a miracle cure. As an infant, the patient SB had lost effective sight in both eyes from a corneal disease. At the age of 52, he received a corneal graft that restored his optics. After living most of his life without sight, SB looked upon his wife for the first time.

While the case of SB is one of the best studied, there have been a few similar cases described over the last few centuries. There is considerable uniformity, and some real surprises, concerning several aspects of these “miracle cures” (von Senden, 1960; Valvo, 1971; Saks, 1991).

First, patients who have been blind most of their lives do not see well after their optics have been repaired. Even after months or years, they continue to struggle at tasks those blessed with sight at birth find effortless. Some visual measures, such as acuity and color vision, can be within the normal range. But patients do not perceive depth, motion, or the relationship among features effortlessly. They have difficulty recognizing a face, or judging the movement of traffic. Their visual world is a jumble from which they can occasionally glean a useful pattern or bit of information. The description of these cases suggests that many patients never acquire a good facility at grouping together features from different positions within the image, or features scene at different points in time from different perspectives. They have great difficulty integrating information from different visual perspectives, over time, into a coherent description of the scene.

The difficulty in integrating information is not a small thing. The restoration of the elements of sight without this integrative ability is a disconcerting emotional experience. Most of the patients experience severe depression, and even those patients who overcome the depression, wonder whether the returned sight was worth the effort. In summarizing the cases he studied, Valvo (1971) wrote

The congenitally blind person especially, has to face the prospect of a difficult struggle before reaching a stage at which his vision permits him to understand the world around him. For a period of time varying with each patient, these people experience a confusing proliferation of perceptions, and they must learn to see as a child learns to walk. Moreover, personalities and character armors built up as a blind person have to be shed, and they often find it difficult to change their ways of

living. As one of our patients put it, “I had to die as a blind person to be reborn as a seeing person.” [page 4]

Gregory and Wallace heard about SB’s restoration of sight from a story in a London newspaper. They managed to get to the hospital after the first operation, in which the optics of one eye were repaired, but before the operation on the second eye¹. They continued to visit with SB and examine his vision, when his health and mood permitted. Fairly quickly, SB managed to recognize various forms including upper case letters and the face of a clock. His ability to recognize such patterns quickly was apparently due to his ability to transfer his understanding of these shapes based on touch into a corresponding visual sensation. This happened automatically and quickly, at a rate that astonished Gregory and Wallace. It suggested to them that he had a good facility for integrating information into objects and patterns when the information corresponded to his tactile experience.

Equally surprisingly, SB could recognize the shapes in the Ishihara color plates quite easily. He learned to identify color names, and in fact some colors were already known to him because, even though when blind he could not see pattern, he could detect the difference between light and dark. Also, during ophthalmological exams during his blindness the strong light gives yields a red appearance that was probability familiar to him as well.

Many of our most important perceptual abilities, however, were beyond SB’s reach. We take for granted our ability to judge the shape of objects as we change our viewpoint. As we walk around a house, or a tree, or a person, each image that we see is different. Yet, we integrate the information we acquire into a single unified description of an object or a person. But, SB seemed to experience a different world as he moved around an object.

Quite recently he had been struck by how objects changed their shape when he walked round them. He would look at a lamp post, walk round it, stand studying it from a different aspect, and wonder why it looked different and yet the same [Gregory, 1974, p. 111].

To see a moving object, we must also see the connection between the object at different moments in time. As the object moves further and further, we often see it from different perspectives and we must be able to integrate the different retinal images of the object into a single coherent description. Patients with restored sight have a difficult time learning to perceive motion and depth. Saks quotes from a patient with restored sight, Virgil, who wrote in his journal,

¹The original monograph is difficult to obtain. But, it is reprinted, along with additional material in a collection of Gregory’s writings “Concepts and Mechanisms of Vision.”

During these first weeks [after surgery] I had no appreciation of depth or distance; street lights were luminous stains stuck to the window panes and corridors of the hospital were black holes. When I crossed the road the traffic terrified me, even when I was accompanied. I am very insecure while walking; indeed I am more afraid now than before the operation.

Gregory's description of SB is striking in its similarity.

He [SB] found the traffic frightening, and would not attempt to cross even a comparatively small street by himself. This was in marked contrast to his former behaviour, as described to us by his wife, when he would cross any street in his own town by himself. In London, and later in his home town, he would show evident fear, even when led by a companion whom he trusted, and it was many months before he would venture alone. We heard that before the operation he would sometimes injure himself by walking briskly into a parked vehicle, or other unexpected obstruction, and he generally did not carry a white stick. As a blind man he was unusually active and aggressive. We began to see that this assurance had at least temporarily left him; he seemed to lack confidence and interest in his surroundings.

To perceive motion, the visual system must be able to integrate information over space and time. To perform this integration, then, requires a means of short term visual storage that can be used to represent recent information and visual inferences. If this visual storage fails, perhaps because it did not develop normally during early blindness, motion perception will be particularly vulnerable; more so, say, than color perception. One of Valvo's patients, HS, describes his difficulties with short term visual memories as he learned to read. He wrote in his journal,

My first attempts at reading were painful. I could make out single letters, but it was impossible for me to make out whole words; I managed to do so only after weeks of exhausting attempts. In fact, it was impossible for me to remember all the letters together, after having read them one by one. Nor was it possible for me, during the first weeks to count my own five fingers: I had the feeling that they were all there, but ... it was not possible for me to pass from one to the other while counting.

These clinical cases are important for the qualitative information they provide us. We learn that patients can identify colors, or even individual letters. Yet, they have difficulty integrating their visual experiences into a single whole. As they walk around an object, it appears to be a series of different shapes, not a single unitary thing. Moving objects do not have any continuity of existence. Distance, which also

requires a relative judgment, is impossible to judge accurately. The experience of these patients shows us how important the processes that integrate information over space and time are to seeing. To understand seeing, we must understand the processes that link our inferences of pattern, color, motion and depth into a unified description of the world.

1.2 Illusions

Illusions are fun. They draw people into our discipline, they inspire new algorithms, they fill us with wonder. They are the children of our professional lives. And like children, illusions are a bit unruly. They do unpredictable things and defy a simple organization. You can try to insist that an illusion clean up its room, but a few minutes later you will discover another idea thrown haphazardly on the floor, or a theory turned upside down.

Of the many illusions known to vision scientists, only a fraction are suitable for the printed page. Of that portion, I have included mainly illusions to make some points about how we see objects. To understand seeing we must understand how we integrate all of the different inferences concerning pattern, color, motion and depth into a single description of the world.

Seeing the Three-Dimensional World

A central premise of object perception is that we see objects in a three-dimensional world. If there is an opportunity to interpret a drawing or an image as a three-dimensional object, we do. This principle is illustrated by the drawing created by Shepard (1990) shown in Figure 1.1. The two table tops have precisely the same two-dimensional shape on the page, except for a rigid rotation. Nobody believes this when they first look at the illusion. To convince yourself that the shapes of the table tops are truly the same, trace one of them on an overhead transparency or tracing paper, and then rotate the tracing around. Or, make a cutout that covers one table-top and then rotate it and place it on the other. The illusion shows that we don't see the two-dimensional shape drawn on the page, but instead we see the three-dimensional shape of the object in space. This experience, which is inescapable for us, appears to be unattainable for individuals like patient SB whose case was described in the previous section.

Boring (1964) illustrated the way we automatically interpret size and depth using an image like the one shown in Figure 1.2. When we copy the image of the distant figure and place it next to the closer figure, we are surprised to see the size of the distant figure on the page. Boring and Shepard's illusions show that we interpret the

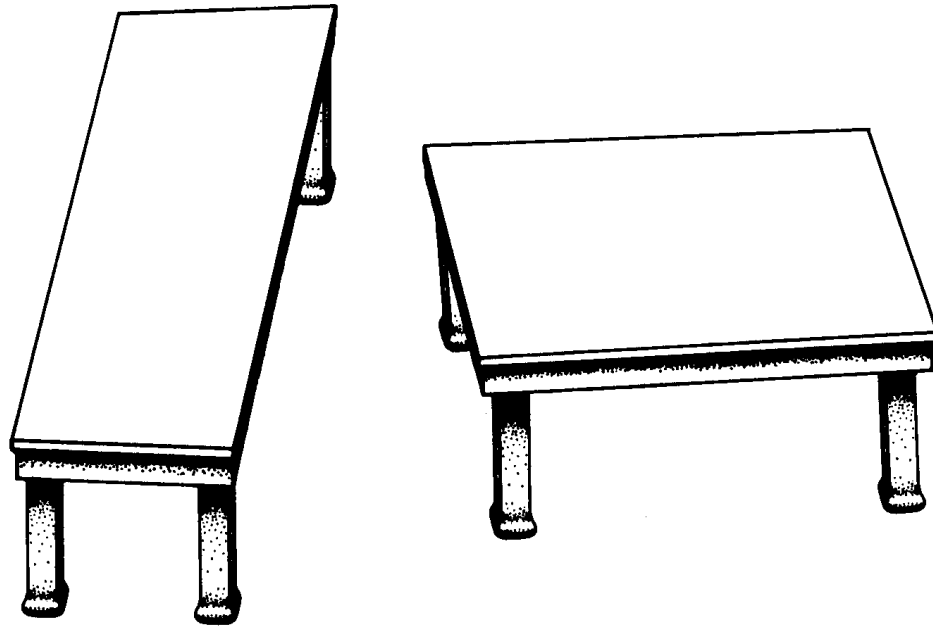


Figure 1.1: *We assume two-dimensional shapes describe three dimensional objects.* Drawn on the two-dimensional page, the table tops are the same except for a rotation. Convince yourself that the shapes are the same on the page by making a cut-out equal in size to one of the tables. Then, rotate the cut-out and place it on the other table. (Source: Shepard, 1990).



Figure 1.2: *Judging size*. Seen in its proper context, we can use the image to infer the man's height accurately and we are unaware of the size of the man's image on the page. We are made aware that the man's image is small when we translate the image to a new position with improper depth cues (After Boring, 1964).

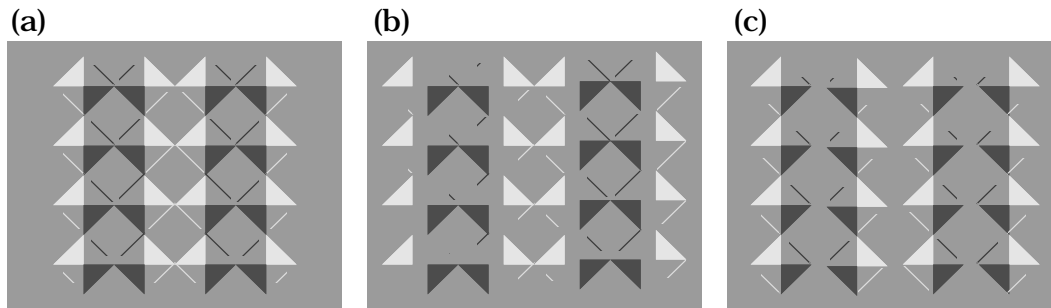


Figure 1.3: *Brightness and shadows*. (a) The intensity of the light reflected by the diamond regions in the middle and right columns is the same. Yet, the diamonds in the middle column appear darker than the diamonds in the right column. (b) When we displace the columns and destroy the interpretation of the image as containing shadows, the brightness illusion is decreased greatly. (c) When we displace the columns but maintain the perceived shadows, the brightness illusion remains strong. (After: Adelson, 1993)

size of the distant figure in terms of the three-dimensional cues in the image. It is hard for us to see the image on the page because, in most cases, we infer the size of things as if they were projections of three-dimensional objects.

Shadows and Edges

Not just size, but most visual inferences are based on the interpretation of image data as arising from objects in a three-dimensional world. Even judgments that seem simple, such as brightness, may depend on interpreting the scene as consisting of objects in a three-dimensional world.

Figure 1.3 is an example of a brightness judgment that depends on our interpretation of the objects in the image (Adelson 1993). Consider the middle and right columns of diamond shapes in Figure 1.3a. The physical intensity of the light reflected these two sets of diamonds is the same. But, the diamonds in the middle column appear darker than the diamonds in the right column.

Adelson (1993) suggests the brightness difference between the columns arises because of a transparency, that is some columns appear to be seen through light and dark strips overlaid on the image. Another interpretation of the differences between the columns is that some columns are seen under a cast shadow (Marimont, personal communication). In either event, the brightness of the local regions appears to depend on the global interpretation of the image. This is shown by the images in Figure 1.3bc, which are variations of the image in (a). The image in (b) has no shadow edge, (b) while the image in (c) changes the image without destroying the perception of a shadow (c). The brightness difference is diminished when the

shadow is destroyed, but the difference is maintained in when the shadow is present (c).

As I described in Chapter ??, brightness and color appearance are better predicted by reflectance than the light incident at the eye. If the visual system's objective is to associate brightness with reflectance, then the visual system should take transparency into account when judging an object's brightness. If the physical intensity of the light from a surface seen through the semi-transparent object has the same intensity as light from a surface seen directly, then the surface behind the transparency (right column) must be more reflective and hence judged brighter. The example in Figure 1.3a shows that even image interpretations as complex as shadows or transparency can influence the brightness of a target.

The illusion in Figure 1.4 is named for three individuals who discovered it separately: the *Craik-O'Brien-Cornsweet* illusion. The illusion shows that surface boundaries influence brightness. The two areas on opposite sides of the border have the same physical intensity. Yet, the region on the right appears darker. The reason for this is that the intensity pattern at the border, shown at the bottom of the figure, suggests a spatial transition from a light to dark edge. This transition only occupies a small part of the image, and the intensity within the two regions away from the edge is the same. But, the visual system extends the inference from the boundary to a brightness judgment of the two large regions. It is quite surprising that the inference made using the boundary transition overrides the intensity levels within the individual regions. The inference from the boundary spreads across a large region and influences our perception of the entire object (Craik, 1966; O'Brien, 1958; Cornsweet, 1970; Burr, 1987).

Figures 1.3 and 1.4 show that judgments of transparency and boundaries can influence judgments of brightness. Figure 1.5 shows that brightness judgments can influence the perception of shape. Panel (a) shows an image containing a mound of dirt with a small dimple at the top. Panel (b) shows a second image containing a small crater with a mound at the bottom. The images in Figure 1.5ab are the same except for being flipped (not rotated) up and down using a simple image processing program.

If you rotate this book by 180 degrees, you will see that the mound in Figure 1.5a changes into a crater, and conversely the crater in Figure 1.5b changes into a mound. When we interpret these shapes, we assume that the illuminant is elevated. This assumption about the position of the illuminant guides our inference about the shape of objects in the image. The distinction between mound and crater in these images is mediated mainly by the shading differences. Hence, rotating the images changes the shading relationship and we reinterpret the shape. Ramachandran (1988; see also Knill and Kersten, 1991) has demonstrated this phenomenon in a number of different ways. He argues further that the brain simplifies the interpretation of images by assuming the illumination consists of a single light

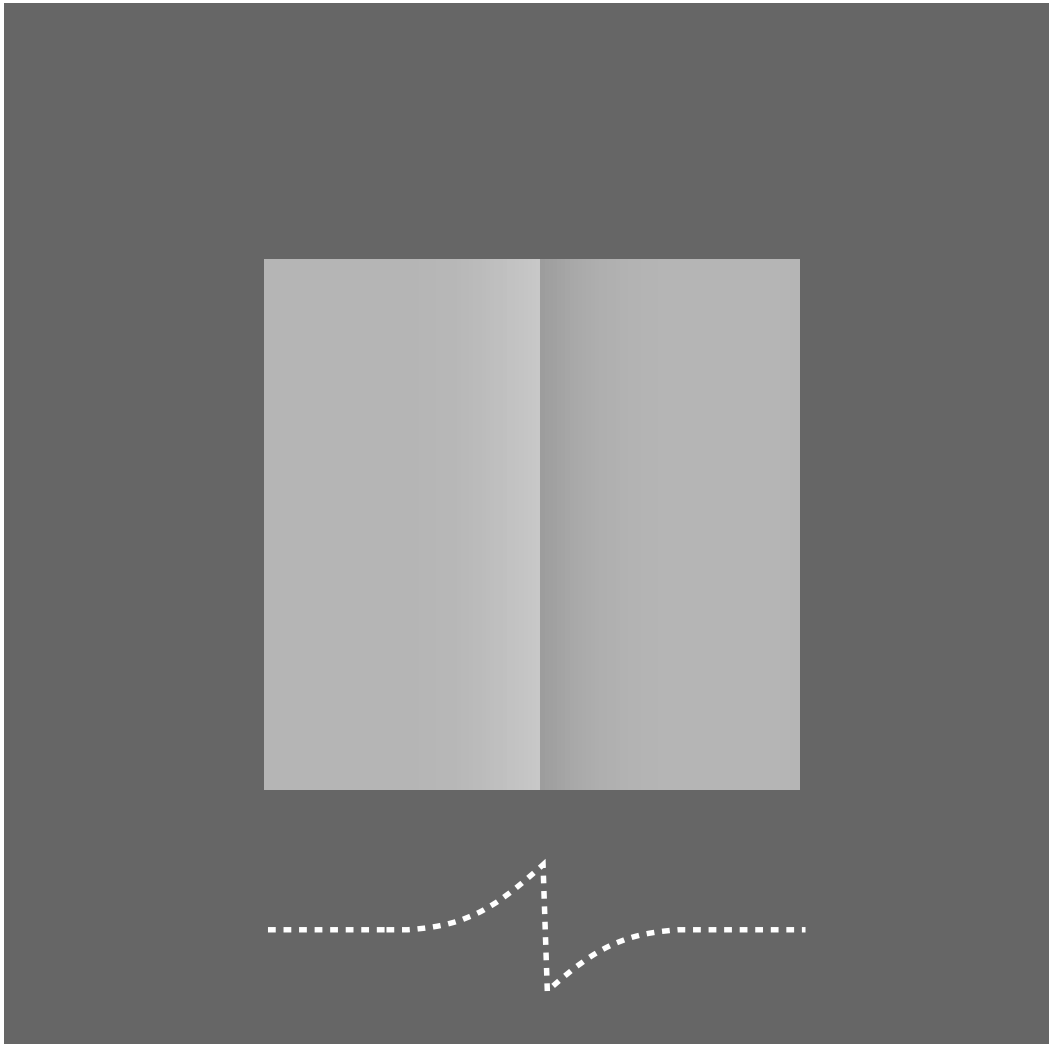


Figure 1.4: *Edges can influence the brightness of a large region.* The relative physical intensities of the inset region is shown by the trace. The intensities of the regions are equal but they are separated by a transient that defines an edge. Even though the intensities are equal, the region on the right appears darker. To confirm that the physical intensities of the areas are equal, cover the edge transient.

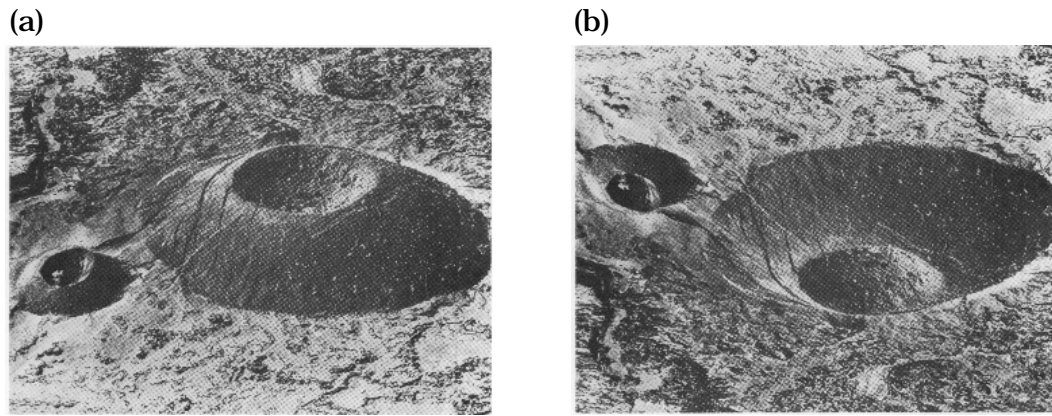


Figure 1.5: *Shading influences shape*. The image in (a) has the appearance of mound of dirt with a small indentation. The image in (b) appears to contain a crater with a mound at the top. Yet, the two images are the same except for an up-down flip. If you rotate the book 180 deg, the image containing the mound will now appear to contain a crater, and conversely the image with a crater will appear to contain a mound. The spatial relationship between the light and dark regions of the mound/crater is the main source of information defining it as convex or concave. Rotating the image rotates the shading cue and thus changes the shape we infer (After: Rittenhouse, 1786).

source.

Shapes

The Fraser *spiral* is named after the *perceived* form in Figure 1.6. In fact, there is no spiral in the Figure at all; the apparent spiral is really a set of concentric circles. (To persuade yourself of this, take your finger and carefully trace one of the patterns that you believe to be part of the spiral.) The light dark structure of the patterns within each circle suggest an inward spiral. But this curvature is not present in the global shape. Visual inferencing mechanisms fail to notice that the local features do not join properly into a single global spiral. This image, like the many famous drawings by Escher (1967), show that the visual mechanisms for interpreting objects in images can yield globally inconsistent solutions.

The objects you see in Figure 1.7 are visual inferences derived by integrating cues concerning occlusion and transparency. Figure 1.7a show an image of a white triangle occluding three disks. We see the triangle even though no edges are present in the image to support the hypothesis that the triangle is present. Compare this figure with Figure 1.4. There, boundary information influenced the judgment of brightness. Here, occlusion information influences the judgment of a boundary and

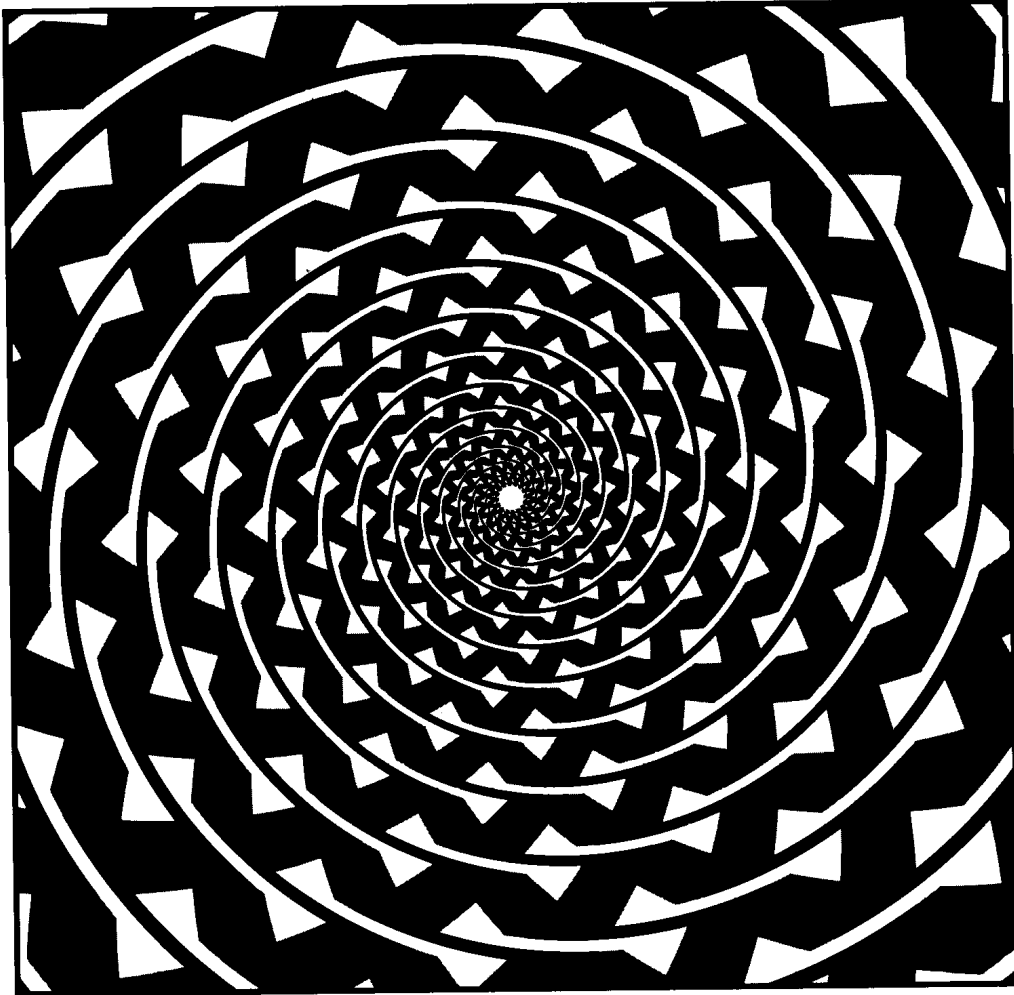


Figure 1.6: *The Fraser spiral*. The figure consists of a set of concentric circles, not a spiral at all. Yet, because of the local pattern within the circles, we perceive the overall pattern as if it were a single spiral.

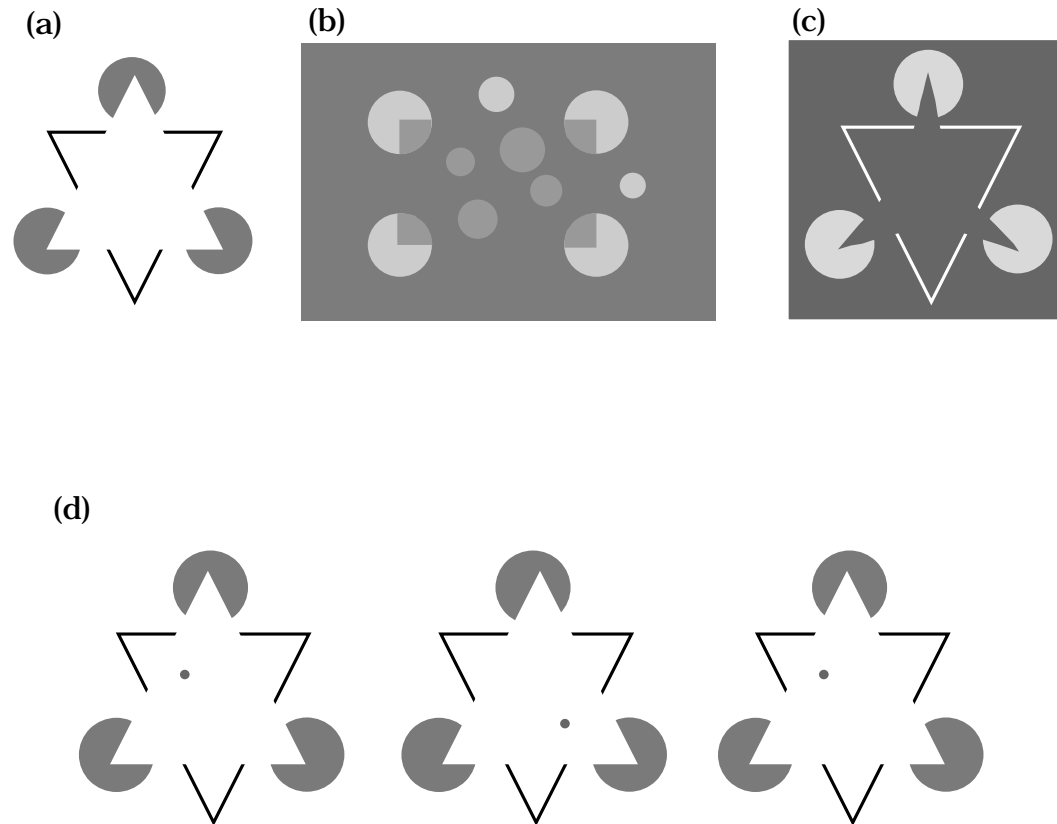


Figure 1.7: *Subjective contours*. These subjective contours are inferred from occlusion and transparency cues in the images. (a,b,c) A triangle is suggested by occlusion, a rectangle is suggested by transparency, and a curved object is suggested by occlusion. (After: Kanizsa, 1976) (d) Stereo pairs of subjective contours. By diverging your eyes beyond the page, the image pair on the right (left) will fuse and you will see the subjective contours of a triangle in front (behind) of the circles. The subjective contour is somewhat more vivid seen the depth cue is added. If you converge your eyes to fuse, the depth relationships will reverse. (After: Ze and Nakayama, 1994).

brightness. Figure 1.7a shows that visual inferences accept the occlusion information as highly informative, even though there is missing edge and brightness information.

The transparency cues in Figure 1.7b are enough to infer the presence of a rectangle. Figure 1.7c shows that occlusion information can be used to infer rather complicated curved shapes, not just straight edges.

Figure 1.7d contains stereo pairs of the subjective contour in panel (a). When the depth cue is added, the subjective contour becomes somewhat more compelling (He and Nakayama, 1994). Fusing a stereo pair takes some practice. Try placing a piece of paper perpendicular to the page and between the two images you wish to fuse. Put your nose against the edge of the paper so that each eye sees only one of the patterns. If you then relax, and look through the page, the two images will fuse into a single depthful image. If you see both dots on the two figures, you will know that you have merged the two images, not just suppressed one of them. If you fuse the pair on the right, the triangle will appear to be in a plane floating above the page. The pair on the left shows the subjective triangle behind the page ².

Normally, we think of occlusion as removing information and thereby making it harder to detect an object. However, the two examples Figure 1.8 show the presence of an occluding object can help us explain image information and see an object that might otherwise be difficult to discern. The pattern on the left of Figure 1.8a appears to be a set of two-dimensional drawings. When the gaps between the drawings are filled in by an occluding object, however, we can integrate the different drawings into a single three-dimensional shape of a cube. The only difference between the two drawings in panel (a) is that the white gaps separating the sections on the left have been filled in by the dark bars.

The patterns on the left of Figure 1.8b are drawn as if they were separate parts. Precisely the same patterns are present on the right, but this time they are separated by dark bars that suggest an occluding object. Again, it is much easier to recognize the pattern as a collection of B's when the occlusion is made visually explicit. Occlusion is a very important clue for visual inferences having to do with objects.

Integrating cues

Much of object perception requires us to integrate image information for image features that are separated in space or in time. In some cases, when we integrate separate visual features into an object, we rely on certain implicit assumptions about the object. An interesting example that reveals our implicit assumptions can be

²If you fuse these stereo pairs by converging, rather than by diverging, your eyes, the depth relationships reverse.

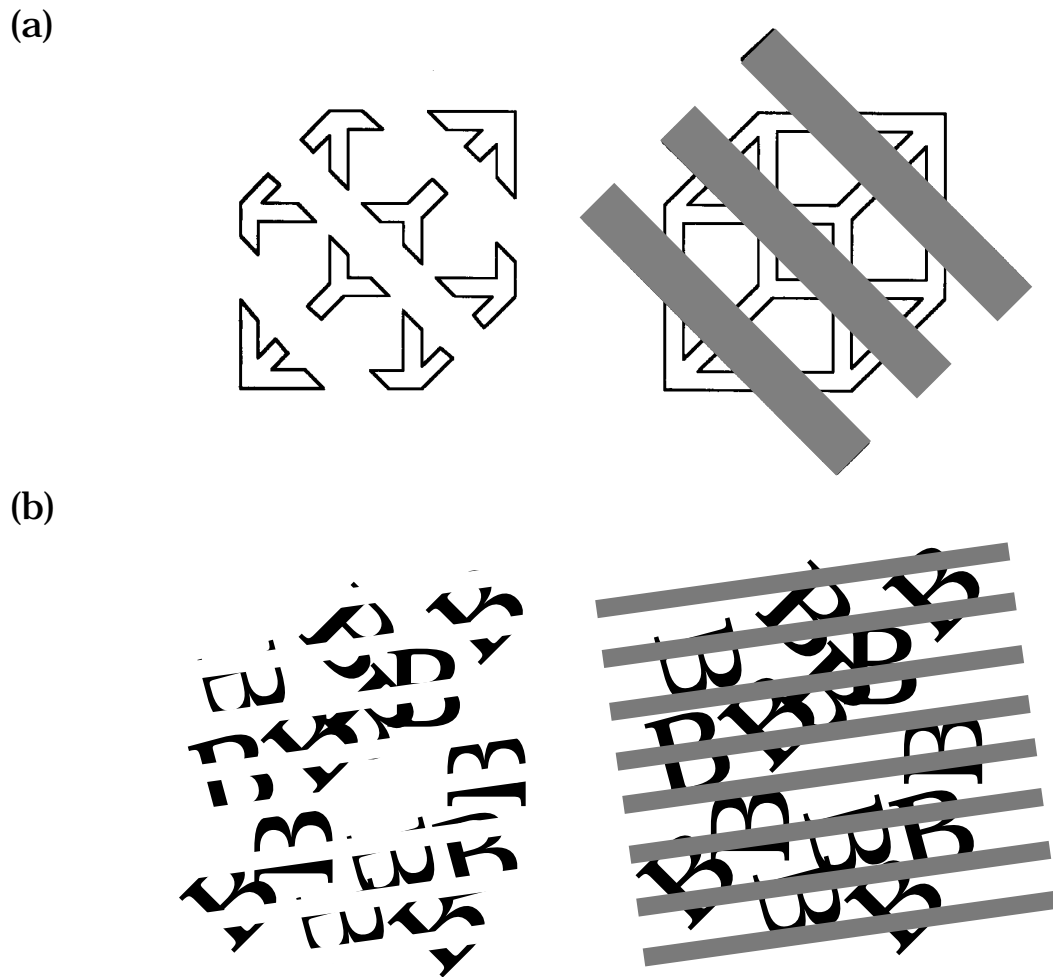


Figure 1.8: *Occlusion and object recognition.* The presence of a clearly visible occluding surface helps us to integrate otherwise fragmentary image components. (a) When the line segments are seen without an occlusion cue, they appear as a set of uncorrelated two-dimensional patterns. By overlaying occluding boundaries, the pattern is seen as part of an object, namely a three-dimensional cube (After: Kanizsa, 1979). (b) When the pattern on the left is seen on its own, it appears as a jumble of unconnected curves and lines. By placing an occluding object over the white spaces, it is much easier to see that the occluded pattern is a collection of “B’s.” (After: Bregman, 1981).

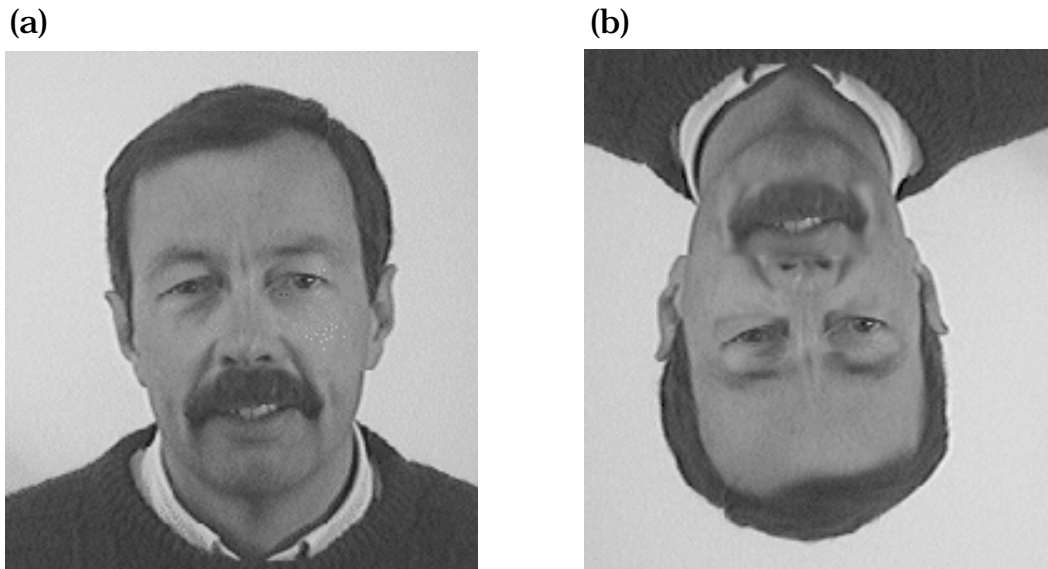


Figure 1.9: *Face recognition*. (a) A face. (b) An edited version of the face in (a). Can you integrate the features when the face is inverted and predict the expression that will appear when you rotate the book? (After: Thompson, 1980).

revealed by the images in Figure 1.9. The image in (a) shows a face in its normal upright pose. The image in (b) shows the same face with several of the features, namely the eyes and mouth, edited. In this form, it is recognizable as the same face, and it seems clear that there is something amiss with the individual features. But, we cannot infer what the expression on the face will be when the inverted face is rotated into the upright position (Thompson, 1980). You will be surprised at the appearance of the face in (b) when you rotate the book and see the face in its upright pose.

This may be an illusion that has to do with faces. The clinical syndrome of *prosopagnosia*, the inability to recognize familiar faces, is further evidence that our brain has specialized circuitry for integrating the components of an image when we recognize and interpret faces. It seems more likely to me, however, that this illusion has to do with the integration of spatially segregated features. When we see a familiar object made up of many separate features in an unlikely pose, it is very difficult to judge the object's structure (Kanizsa, 1979).

Geometric Illusions

Gregory (1966) suggested that the simple and unassuming Muller-Lyer illusion, shown in Figure 1.10a, results from basic perceptual assumptions we make when we perceive depth. The two vertical lines shown in the illusion have the same length, but the line segment on the right appears longer. Gregory argues that the lines

appear of different length because we cannot escape interpreting even such trivial images as three-dimensional objects. As the image in Figure 1.10b illustrates, in the natural environment the edges that define a near corner are similar to the lines on the left of Figure 1.10a. The edges that define a far corner are similar to the lines on the right. Gregory explains the illusion as a consequence of our relentless interpretation of images as arising from a three-dimensional world. Lines that sweep out equal retinal angle, but that are at different depths, must have different physical size. Gregory suggests that even the impoverished stimulus in the Muller-Lyer illusion invokes the visual systems inferences of objects and depth. The improper application of a good principle causes the equal line segments to appear different.

Gregory's hypothesis is important because it reminds us that the basic function of visual inferences is to see objects in three-dimensions. The interpretation of visual illusions, however, is never straightforward. To test Gregory's suggestion, DeLucia and Hochberg (1991) created the version of the Muller-Lyer shown Figure 1.10c. In the two separations between the three dots are equal. Yet, just as in the Muller-Lyer the separation between the two dots attached by the closed arrow form appears smaller than the separation between the dots connected by the open arrow form. In this schematic three-dimensional image, there is little chance that the separation between the dots has to do with different depths. Gregory's hypothesis and this counter-example are a wonderful exchange that illustrates how qualitative hypotheses concerning the mechanisms of visual illusions can be tested and become part of the scientific study of vision.

1.3 Summary

Throughout this chapter, we have seen the importance of integrating separate visual inferences into a single explanation of the contents of a scene. Patients who cannot integrate visual information may identify color or a shape, but they feel that they cannot see because they cannot integrate the separate inferences into a sensible interpretation of the objects and surfaces in the scene. Many of the illusions we have reviewed show us that to see as a whole, we must resolve conflicting information. We do not see a two-dimensional shape properly because we insist on interpreting the data as a three-dimensional shape. We do not see the physical intensity properly because we insist on interpreting a shadow or an edge.

I have chosen the illusions here to emphasize several important principles that the visual system uses to combine different visual inferences. We integrate shape and depth cues assuming that we perceive objects in a three-dimensional world; we use shadows, occlusions and edges, to interpret the properties of objects; we build up global interpretations from many local properties; yet, we also use familiar poses of objects and typical locations of illuminations help us to interpret ambiguous images.

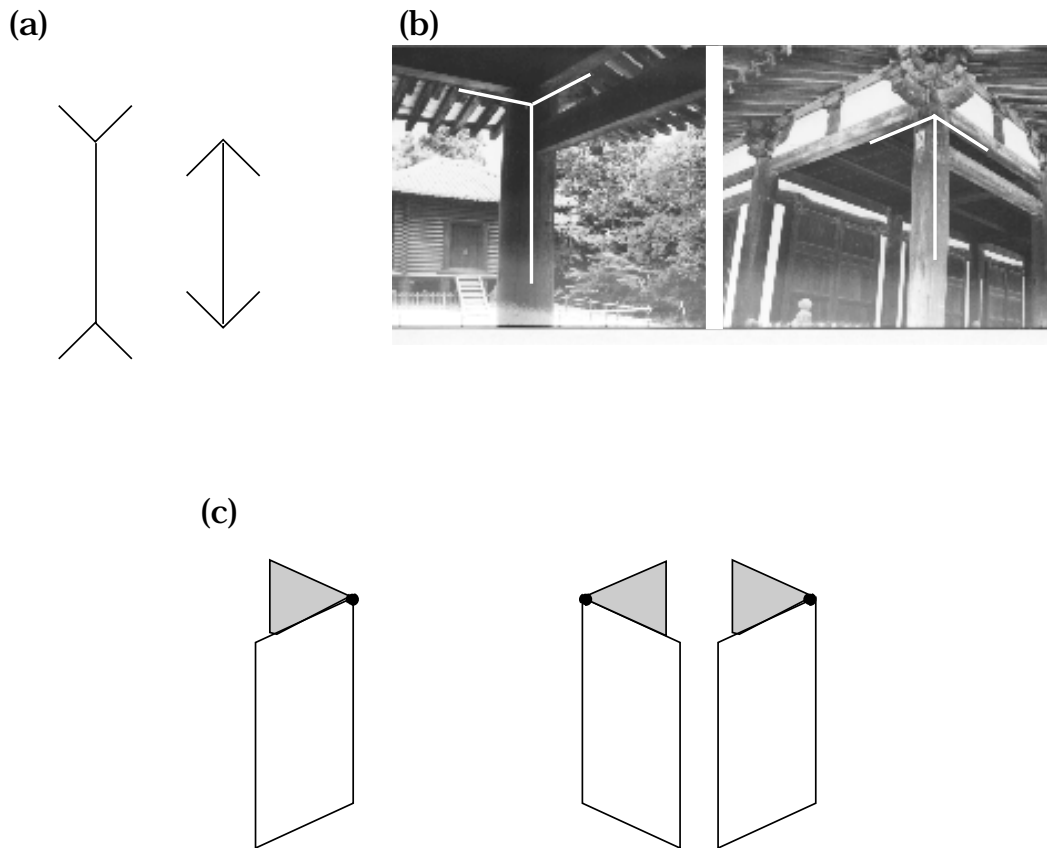


Figure 1.10: *The Muller-Lyer illusion.* (a) The classic Muller-Lyer illusion is shown. The line with the arrows pointed outward appears longer to most people, but the line lengths are the same. (b) The left and right parts of the image show two views of a corner of a building. From the outside, when the corner is relatively close to the viewer, the edges between the corner and the ceiling are oriented like closed arrows. From the inside, when the corner is relatively far from the viewer, the edges are oriented like open arrows. (c) A three-dimensional analogue of the Muller-Lyer is shown. The separations between the middle dot and the dots on either side are the same. In this rendering of the illusion, the open and closed arrow shapes are not a cue for depth. Yet, the Muller-Lyer illusion is quite powerful (Source for panel c: DeLucia and Hochberg, 1991).

The rules governing our sight reflect the physics of the world we see. These rules describe the interactions of objects at a physical scale within the domain of the psychologist and engineer, a human scale. This is not the physics of the sub-atomic or the physics of the galactic, but it is the physics of the world in which we live and interact. By studying the rules of this human-centered physics of perceived objects, we learn how we might see. By making neural and behavioral measurements of our brain and our perceptions, we learn how we do see.