Chapter 23

The Third Dimension: Distance

Contents

23.1 Definitions: Distance vs. depth ................................................. 580
23.2 Accommodation and convergence ............................................. 581
23.3 Pictorial cues ................................................................. 582
23.4 Motion cues ................................................................. 585
23.5 Binocular disparity: The most famous distance cue ................. 588
23.6 How do distance cues combine? ............................................... 593
23.7 Size constancy .............................................................. 596
23.8 Summary: Cues and heuristics for distance ............................ 599

In this chapter and the next we return to a problem we first encountered in Chapter 5. The physical world is three-dimensional, but the retinal image is two-dimensional. The First Transformation preserves the two-dimensional layout of the visual field — left and right, up and down — but collapses the third dimension. The facts of the First Transformation would seem to imply that information about the distance to an object, as well as its depth or three-dimensional shape, should be lost to perception.

Moreover, the size of the retinal image of an object, $s_{\text{retina}}$, depends both upon the size of the object $s_{\text{object}}$ and its distance from the eye $d$; that is,

$$s_{\text{retina}} = \frac{s_{\text{object}}}{d}.$$ 

If information about distance is lost at the level of the optics, then information about size should also be lost, and we should be unable to perceive the sizes of objects accurately.

Yet, bumblebees can fly. Clearly, we do perceive the distances, sizes, and shapes of objects, and we usually perceive them remarkably veridically. So somehow, we are getting the information to allow us to do these things. The question is, how? Part of the answer is that the retinal images contain many features that come about because of variations of distance. In fact, we use these features — distance cues — to generate veridical perceptual representations of the distances and sizes of objects. The main purpose of this chapter is to identify the many different distance cues, and to ask how they combine to determine our perceptions of distance.
In addition, the topic of distance perception provides an excellent illustrations of one of our most important themes. As it turns out, there are many different cues to depth and distance. But not one of the cues contains all of the information needed to use that cue as an information source about distance. That is, each of these different cues depends upon the use of its own special heuristics. A major goal of this chapter is to exercise your understanding of the concept of heuristics.

23.1 Definitions: Distance vs. depth

We begin with a definitional distinction between the terms distance and depth. Although these terms are often used interchangeably, we will use them in different ways: distance will be used to refer to the physical distance from the observer to some object. Depth will be used to refer to the three-dimensional shape of an object – for example, whether a given object is a three-dimensional sphere vs. a flat two-dimensional disk, or a three-dimensional human figure vs. a flat human-shaped cut-out. Of course, in a sense the depth of an object is just the difference in distance between its front and back surfaces, and the cues for distance and depth will clearly be closely related and overlapping. However, we believe that the definitional distinction is an important one.

In Chapter 19 we discussed the concept of figure-ground segregation – the perceptual task of determining which part of the retinal image is the figure and which is the ground. Similarly, in three dimensions, our perceptual systems need to work out the task of scene segmentation: Which blobs of space are occupied by objects, and which by empty spaces? It seems likely to DT that the eventual representations of objects will be very different from the representations of empty spaces. The analysis of the depths and three dimensional shapes of objects is a major task of object recognition, and it needs to be distinguished from the task of distance perception per se.

A second, related reason to emphasize the distinction comes from our simplest formulation of the functions of dorsal and ventral streams: Where vs. What. To know the distance of an object is to know where it is in space. But to know the three-dimensional shape is to begin the task of determining what it is; that is, the task of object recognition. If where and what are analyzed separately, then distance and depth should also be analyzed separately.

In this chapter, we will treat the logic and psychophysics of distance perception. We begin with a review of the cues that seem most likely to provide us with information about distance. We then present a brief discussion of the problem of how distance cues combine. Finally we relate the perception of distance to the perception of size. In the next chapter we will turn to the topic of depth, and to recent research on the quantitative combination of depth cues. We then turn to physiological studies of distance and depth cues together, and ask what is known about the physiological representation of distance and depth in both dorsal and ventral streams.

23.1.1 Cues to distance

A distance cue is a feature of the retinal image that comes about because of variations in distance. Its presence in the retinal image is therefore a potential source of information about distance.

What are the various distance cues? As a matter of logic and speculation, this question fascinated both the artist Leonardo da Vinci and the philosopher George Berkeley. The many cues they discovered, along with one they missed, still form the heart of our understanding of distance perception. In the next few sections we will explore several different kinds of cues to distance: accommodation and convergence, pictorial cues, cues based on motion, and binocular disparity.
23.2. ACCOMMODATION AND CONVERGENCE

Figure 23.1: Convergence. A. Top view of schematic eyes when looking at a distant object. When the eyes are rotated to place the two retinal images of an object on the two foveas, your two lines of sight are nearly or exactly parallel. B. Top view of schematic eyes when looking at a close object. The eyes must be converged to place the two corresponding lines of sight on the two foveas. C. The convergence angle $a$ needed to bring the two images into register on the two foveas, as a function of the distance $d$ of the object from the observer assuming a typical value for the space between the two eyes, $c$. When a near object changes in distance, there is a relatively large change in the convergence angle. But when a far-away object changes in distance by the same amount, the change in convergence is small. Hence, variations of convergence are more useful as distance cues for near than for distant objects. [A and B by DT; C From Palmer (1999, p. 206, Fig. 5.2.3)].

23.2 Accommodation and convergence

The first two distance cues are created by the motor responses of our visual systems. These cues are sometimes called ocular or physiological cues.

The first of these two cues is accommodation (Figure 4.9). As discussed in Chapter 4, the lens of the eye changes its thickness in order to bring objects at different distances to focus on the retina. In fact, the thickness of the lens required for best focus of a given object varies inversely with the distance of the object. Thus, if your perceptual systems received information about the state of accommodation of your lens, you would have some potentially useful information about the distance of the object – that is, accommodation is a potential distance cue.

The second of these cues, shown in Figure 23.1, is called convergence. When you fixate on a distant object, your two eyes point straight ahead, so that your two lines of sight are parallel or nearly so. When you shift your fixation to a nearby object, you rotate your eyeballs inward to converge your two lines of sight, so as to bring the two retinal images of the object onto the two foveas. The closer the object, the more the needed convergence. And, as in the case of accommodation, your state of convergence is a potential distance cue. (You can demonstrate convergence by asking a friend to first look at a distant object, and then look at his nose. You can also feel the changes your extraocular muscles make when you do the same thing.)

There are several points to be made about accommodation and convergence as distance cues.
First, both accommodation and convergence are cues to absolute distance – the state of accommodation or convergence varies in correspondence with the actual distance from your eyes to the viewed object. As we will see, this is important because most other cues are cues to relative rather than absolute distance.

Second, neither the accommodative state nor the state of convergence per se is useful as a distance cue – both need augmentation by other information. To use accommodation as a cue to object distance, the visual system must determine the state of the lens required for the sharpest retinal image. Similarly, to use the convergence angle, the visual system must analyze the two incoming retinal images, determine which sets of contours in the two eyes arise from the same object, and put these two sets of contours in register by convergence. That is to say, these physiological cues are not simple, and they require extensive image analysis before they can be used for distance perception.

But third, once the state of accommodation needed for good focus, or the convergence needed for binocular alignment, have been determined, only minimal heuristics are required. All that is needed is a lookup table – a list – identifying each state of accommodation or convergence with the corresponding distance value. (And again as you will see, this simplicity sets off the ocular cues from most other distance and depth cues, for which much more complex heuristics will be needed.)

And finally, over what range of distance are these two cues useful? Suppose you are looking at two objects that differ in distance by (say) five cm. As shown in Figure 23.1C, at near distances such a small change of distance brings about a substantial change of convergence. But at larger distances, the same change of distance brings about a smaller or even negligible change of convergence. The same argument holds for accommodation. Thus, both of these cues will be most sensitive for signalling differences in distance for nearby objects, and less and less sensitive at greater and greater distances. As a rule of thumb, it is often said that accommodation and convergence are useful distance cues only over the distance range from one’s nose to about 2 meters, and not useful beyond that distance. To gain information about greater distances, other cues will be needed.

23.3 Pictorial cues

The second category of cues are called pictorial cues (sometimes called painters’ cues or monocular cues). Pictorial cues are the static cues available within each retinal image. In contrast to accommodation and convergence, it turns out that the pictorial cues depend heavily on the use of heuristics.

Two very strong pictorial cues, interposition (or occlusion) and height in the visual field, are shown schematically in Figure 23.2. Interposition (Figure 23.2A) arises from the physical fact that if there are two or more objects in front of you on overlapping lines of sight, the nearest one will be interposed between you and the farther ones. In the retinal image, the shape of the nearer object will be complete (and typically regular), but the shapes of the farther objects will be incomplete (and typically irregular) because they are occluded by the nearer objects.

The problem, of course, comes with the converse – the heuristic. What kind of heuristic would one need to use, in order to assign distances correctly on the basis of interposition cues? The heuristic must include some kind of assumption about the likely shapes of objects – for example, that objects have simple shapes. The interpretational rule would be something like: if two parts of the image share a common contour, and one of the parts is simpler in shape than the other, the object that creates the simpler retinal shape is nearer than the one that creates the irregular
23.3. PICTORIAL CUES

Figure 23.2: Two pictorial cues: Interposition and height in the visual field. A. Interposition. Which object appears closest, and which farthest away? Notice that you get an impression of the order in distance, but no sense of the absolute distances of the objects. B. Fooling the interposition cue. The pattern in Panel A could be created from the physical arrangement in Panel B. In this case the interposition cue would lead to a false perception of distance. C. Height in the visual field. On the left, which square is farthest away? Which bird? Notice also, on the right, that the distance impressions become more certain when interposition and relative size cues are combined. (A and B from Rock (1975, p. 84, Fig. 3-3b); C by DT.).

retinal shape, and the nearer object is occluding part of the farther object. Notice that this heuristic forces your visual system to have a concept of "simple shape", which would need further definition. But even more importantly, if the nearer object has a peculiar shape, and there is a coincidental alignment between the contours of the two objects, this heuristic will give rise to a false perception of relative distances, as shown in Figure 23.2B.

Another strong pictorial cue is that of height in the visual field. This cue is shown in Figure 23.2C. This cue arises from the geometrical fact that if a set of objects is resting on the ground, the images of nearer objects intersect the ground lower in the retinal image than do the images of farther objects. Similarly, for a set of objects floating at equal altitudes above the ground, the image of nearer objects will be higher in the retinal image than the images of farther objects. Combining these two rules, we can say that far away objects will be closer to the horizon in the retinal image, whereas nearby objects will be either farther below or farther above the horizon in the retinal image.

The heuristic involved in the use of height in the visual field as a distance cue would be: order objects in distance according to their closeness to the horizon in the retinal image. This heuristic could, however, give false perceptions for objects that are near the ground but at unequal heights above it, or for objects floating at different altitudes. For example, the nearer of two birds can be imaged lower in the retinal image than the farther one if it is flying at a lower altitude. And think of a large group of balloons floating at different heights and different distances – the height-in-the-visual-field cue would be totally useless in this case.

The next set of pictorial cues can be called cues of perspective, or relative size. Perspective cues are shown in Figure 23.3A. In this photo, you are viewing a rows of plants extending into distance. These rows are parallel. Thus, as objects they form parallel lines. Other examples of parallel lines formed by objects are the two sides of a street or a set of railroad tracks. In cases like these, the
Figure 23.3: Perspective and atmospheric perspective. A. Perspective cues. The rows of plants converge in the distance is a typical perspective cue. B. Atmospheric perspective. The fog has larger effects on the image at greater distances. [A. Rock (1984, p. 19); B. “Relative Depth” by Dennis Markley from Blake and Sekuler (2006, Fig. 8.29, p 305).]

retinal image will contain two lines that converge at the horizon.

The plants in this example are all about the same physical size, so their retinal image sizes will be smaller the farther away they are. This variation in retinal image size of a set of similar objects is called the relative size cue. To make use this cue perceptually, our perceptual systems could use a heuristic like: if the retinal image contains a set of objects of similar shapes but diminishing sizes, represent them perceptually as a set of objects of a constant size but increasing in distance. (You also saw relative size cues at work in the previous Figure 23.2C.)

Another pictorial cue similar to relative size is the use of texture density. Consider a field of stones, or a lakefull of waves, or a lawfult of grass blades. At any given distance the element sizes vary – the stones come in a range of sizes. But in many cases the average size of the elements is about the same over variations in distance. One can see an example of this in the row of stalks shown in Figure 23.3A. The stalks become smaller and more densely placed as the distance increases. The distance cue arises because if the world contains repeated elements of a constant average size, the average size of the texture elements in the retinal image gets systematically smaller with increasing distance. The required heuristic is: if the retinal image contains a set of similar elements of diminishing average retinal image size, represent them perceptually as a set of objects of a constant average size but varying in distance.

A final pictorial cue is atmospheric (or aerial) perspective. This cue applies to the perception of large distances; for example, the relative distances of two or more mountain ranges. Since the atmosphere scatters light, the farther away a mountain range is, the bluer and lower in contrast it will be in the retinal image, and the fuzzier its contours – that is, the more the high spatial frequencies in its contours will be lost. Thus, color, contrast, and contour sharpness can function as distance cues. An example of atmospheric perspective is shown in Figure 23.2B. (Again, we leave the corresponding perceptual heuristic for you to figure out).
23.4. MOTION CUES

23.3.1 Absolute or relative distance?

Unlike accommodation and convergence, the various pictorial cues do not give information about absolute distances, but rather about distance relationships. For example, interposition is a qualitative cue to relative distance. It tells you which object is nearer, but gives you no information about how much nearer (think of the moon eclipsing the sun.). Perspective cues, on the other hand, are quantitative cues to relative distance. If the image of one tree is twice the size of another in the retinal image, the heuristics dictate that the second is twice as far away; if the distance between two assumedly parallel lines diminishes to 1/2, the distance has doubled, and so on.

23.3.2 The painter’s art: fool the eye

The pictorial cues are also sometimes called painters’ cues, because these are the cues that are available to a painter who wishes to capture the three-dimensional nature of the world on a two-dimensional canvas. Of course no matter what the painter does with his paint, the canvas will remain physically two-dimensional; that is, if he succeeds in creating a picture that we perceive as three-dimensional, he has fooled our eyes (as we have tried to do in Figures 23.2 and 23.3). A painting that uses pictorial cues (especially interposition) to yield a strong impression of distance variations is shown in Figure 23.4.

The painter’s trick is to tap into the complex set of pictorial cues and interpretational heuristics that we have just reviewed. If he paints two converging lines on his canvas, we may see two parallel lines converging into depth; a row of trees of decreasing size can become a row of trees receding into depth; a regularly shaped object that has a common contour with an irregularly shaped one may be perceived as occluding it; and an object that intersects the earth higher in the visual field may be seen as farther away.

Thus, the painter’s art – the art of creating illusory distance – provides a particularly striking example of the use of heuristics. The pictorial cues to distance are of the logical form: X when occurs in the physical world, Y occurs in the retinal image. The relevant heuristics – Y is in the retinal image, so X must be in the physical world – are converse statements, and not always true. The eye can be fooled by creating a distance cue on the flat canvas instead of by varying distance itself. It is the same cue, but created by subterfuge. Insofar as it fools the perceptual system into applying a heuristic that is usually worthwhile but false in this instance, it emphasizes the statistical and therefore risky nature of heuristics in perception. But notice also that it is the observer’s false application of heuristics that allows painters and photographers to represent three-dimensional space so realistically on two-dimensional surfaces, and delight us with the results.

23.4 Motion cues

A third class of distance cues arises from the motion of either the physical object or the observer with respect to the other. A particular form of physical motion will yield a particular pattern of temporal change in the retinal image. These temporal patterns of change are often called optic flow patterns or motion gradients. If the perceiver has and uses the appropriate heuristic, she can use the optic flow pattern to set up a (usually) accurate representation of the three-dimensional structure of the world.

One of the simplest examples of an optic flow pattern, a looming pattern, arises as an object of
CHAPTER 23. THE THIRD DIMENSION: DISTANCE

Figure 23.4: The painter’s art – fool the eye. Interposition alone can create an effective impression of depth.[“Boating” by Gabriele Munter from Rock (1984, p. 72).]
a fixed size approaches the observer. In this case, the retinal image of that object will expand in a particular pattern over time, against the fixed images of the surrounding stationary objects. If the object’s trajectory will miss the observer’s head, the expansion pattern in the retinal image will be asymmetrical. But if the object is aimed right at the observer’s nose, the expansion pattern will be symmetrical, and the observer had better duck!

Optic flow patterns can also arise from motion of the observer. Two of the simplest optic flow patterns are diagrammed in Figure 23.5. The first, optical expansion, is similar to looming, but involves an expansion of the whole field of view (Figure 23.5A). As you move forward, the images of the objects in front of you will expand and move toward your visual periphery in the retinal image. The image of the object you are fixating will remain centered on the fovea, and the images of the rest of the objects will expand in a complex pattern that depends on the distances and sizes of each of the objects, and the speed and direction of your motion. Optical expansion patterns are easiest to demonstrate when you are moving forward rapidly, and looking forward, for example when driving a car.

Similarly, as you move sideways through the world (e.g. as you look out the side window of a car or train), the images of objects at different distances will move across the retina at different speeds. This motion pattern is often called motion parallax (Figure 23.5B). In general, images of nearby objects will move rapidly, whereas far-away objects will move more slowly. The image motion for each object will also depend on the point of fixation – the fixated object, of course, will remain fixed on the fovea, and the images of all of the non-fixated objects will move along particular retinal trajectories. Objects nearer than the fixation point will move in the direction opposite to the direction of the observer, whereas objects farther away than the fixation point will move in the
same direction as the observer. [You can easily demonstrate motion parallax by closing one eye and moving your head back and forth from left to right. Vary your fixation distance, and notice the direction of motion of objects closer and farther than your point of fixation.]

Again, as with the painters’ cues, motion cues must come paired with heuristics. Our visual systems presume that particular patterns of motion in the retinal image are caused by whatever physical motion most typically causes them. (These probable heuristics for motion cues to distance are left for you to work out.)

And of course, just as the pictorial cues can be called painters’ cues, the motion cues could be called movie-makers’ cues: if the cinematographer creates the usual motion cues to distance on a flat screen, our perceptual system can be fooled into perceiving three-dimensional motion. A particularly interesting example comes from the cinematographer’s trick of moving the camera around the scene, in order to simulate a moving observer and thus create motion-based distance cues.

23.5 Binocular disparity: The most famous distance cue

Finally, we come to the most well-known distance cue: binocular disparity. The binocular disparity cue arises from the fact that you have two eyes, located in two different positions in your head. Your two eyes view the world from slightly different vantage points, with the result that there is a slight difference – a disparity – in the configuration of the retinal images of objects in the left vs. right eye. The perception of distance (or depth) resulting from using the binocular disparity cue is called stereopsis, or less formally stereo vision.

The geometrical principles involved in binocular disparity are shown in Figure 23.6. In Figure 23.6A, suppose there are two objects (perhaps the masts of two sailboats) in front of you. The first one, F, is straight ahead of you at a given distance, and a second one, E, is farther away and slightly to the right. Suppose you fixate F. As shown in the figure, the two retinal images will be slightly different – geometrically, the retinal images of F and E will be closer together in the left eye and farther apart in the right eye.

Binocular disparity can be demonstrated readily using your two index fingers. Hold the two fingers out in front of you, with the left index finger at (say) 20 cm straight in front of you and the right index finger at (say) 30 cm and displaced slightly to the right. Now fixate your left index finger, and alternately close each eye. With your left eye open, the two fingers should look closer together, while with the right eye open, they should look farther apart.

Two more examples, in which the two masts are at different distances along the same line of sight, are shown in Figure 23.6B and C. These cases can also be demonstrated with your fingers. Hold the two fingers out in front of you, one directly behind the other, and fixate the near finger. With your left eye open, the far finger should appear to the left of the near finger; with your right eye open, the far finger should appear to the right of the near finger. This pattern has been named uncrossed disparity, and it provides a cue that the fixated finger is the closer finger. Now fixate the far finger. The near finger seen with the left eye will lie to the right of the far finger, whereas the near finger seen with the right eye will be seen to the left of the far finger. This pattern has been named crossed disparity, and it provides a cue that the the fixated finger is the farther finger. These names arise because of the perceptual locations of the near and far objects as seen in demonstrations such as this one.
23.5. BINOCULAR DISPARITY: THE MOST FAMOUS DISTANCE CUE

Figure 23.6: The geometry of binocular disparity. The two points in front of the observer represent two vertical lines or objects, such as two sailboat masts. Since the two eyes see the scene from slightly different vantage points, the two retinal images will differ slightly. In each case, the observer fixates at F. A. Two masts in different directions at different distances. The distance EF will be smaller in the left eye than in the right eye. In technical terms, the difference in the two angles subtended by the images in the two eyes is the retinal disparity. For example, if the angle EOF in the left eye were 2°, and in the right eye 1°, then the disparity would be 1°. B. Fixating the nearer of two masts in the same direction. The two images of G will fall inside the two images of F (toward the nose in each eye). By convention this configuration is called uncrossed disparity. C. Fixating the further of two masts in the same direction. The two images of D will fall outside the two images of F (away from the nose in each eye). By convention this configuration is called crossed disparity.
In Figure 23.6B and C, of course, the rays “cross” within the eye in both cases, and the nomenclature is confusing. The names arise because in perception, when you fixate the near finger and attend to the far finger, the far finger is perceived as being to the left of the near finger in the left eye, and to the right of the near finger in the right eye, hence “uncrossed”. If you fixate the far finger and attend to the near finger, the near finger is to the right in the left eye and to the left in the right eye, hence “crossed”. If you remember that the retinal image is left-right reversed with respect to your perception, you can reconcile the “fingers” demonstration with Figure 23.6B and C.

In sum, binocular disparity is a distance cue. The magnitude and direction of the disparity are determined by the relative locations of objects. A complex scene with many objects at different distances yields a complex pair of images with many different disparities, determined by the locations and distances of the objects.

Now, what perceptual heuristics will allow us to make use of disparity as a distance cue? The basic heuristic must be that, relative to fixation, a disparity between the locations of figural elements in the two retinal images should be interpreted as a difference in their distances. Crossed disparities are interpreted as revealing objects nearer than the fixation point, uncrossed disparities as objects farther than the fixation point, and the quantitative variation in disparity is a function of distance to the object and the distance to the fixated point.

As a historical note – as you have just seen, the doubled and disparate images from the left and right eyes are easy to become aware of, and the geometry of binocular disparity is not difficult to work out. Yet surprisingly, early authors including both Da Vinci and Berkeley missed this distance cue. The first scientific report of an analysis of binocular disparity is that of Wheatstone in 1838. Wheatstone built a stereoscope – a device for presenting two different pictures separately to the left and right eyes – and showed that artificially produced disparity gives rise to the perception of differences in distance. Thus stereoscopes, like paintings and movies, provide another example of how our visual systems can be fooled when the artist or scientist mimics a distance cue artificially, and the visual system applies the usual heuristic and interprets the cue as having arisen from differences in distance.

Some of Wheatstone’s original stereograms are shown in Figure 23.7. The figure legend provides instructions for free viewing of stereo images to yield the perception of variations in distance and depth. The method is called crossed free fusion because it involves crossing the eyes. There is an alternative method in which one diverges the eyes rather than crosses them that is called parallel free fusion. Parallel free fusion gives the same depth relation as a steroscope while crossed free fusion reverses the depth relations.

### 23.5.1 Random dot stereograms

In early thinking about stereopsis, a particular order of sequential processing was often implicitly assumed: that analysis of the patterns in the two eyes must be performed before the two patterns could be “matched up” by the visual system to yield a stereo cue. But the scientist Bela Julesz (1971) created a new custom-designed stimulus called a random dot stereogram. Examples of random dot stereograms are shown in Figures 23.8.

To create the stereogram, Julesz first divided a region of visual space up into very small checks, and colored each check, or dot, randomly either black or white. He then duplicated the image. In a central region of the second image, he “picked up” a set of dots from, say, a T-shaped region
23.5. BINOCULAR DISPARITY: THE MOST FAMOUS DISTANCE CUE

Figure 23.7: Examples of Wheatstone’s original stereograms and instructions for free fusion. The four pictures on the right have been made by reversing the left and right eye views of Wheatstone’s figures 14 and 16, to change the direction of the disparity. With a little effort, most people can learn to “free fuse” these stereograms. Chose one of the figures (#16 works well). Hold the figure at normal reading distance, but vertically in front of your eyes. Now hold one finger about 2/3 of the way from your face to the page, with its tip aligned with the bottom of the stereogram. Look at the stereogram, and you should see two images of your finger. Move your finger back and forth in depth until one image of the finger is centered under the left half of the stereogram, the other under the right half. Now converge on your finger, but shift your attention back to the stereogram. After a while the two images should merge perceptually, and you should see the smaller circle standing out in front of the larger one. This method is called crossed free fusion [Modified from Gillam (1995, p. 43, Fig. 7)].
Figure 23.8: Random dot stereograms. A. A dense random dot stereograms, showing a T shaped figure in front of its surround. B. A sparse random dot stereograms, showing a square in front of its surround. Depending on how one free fuses these images, the depth order may be reversed. (From Julesz (1971, A: page 22, Fig 2.4-4. B: page 122, Fig. 4.5-5)].
near the center, moved this whole set of dots, say two dots worth to the left, creating a disparity between the two T-shaped areas. Finally, he filled in the space created at the left edge of the set with new random dots.

Notice that in Figure 23.8A the central T – the figure created by the disparity – cannot be seen in either of the two images alone. Yet in fact, when these pictures are viewed by the two eyes separately, the observer can readily see the T, standing out in depth! Julesz thus made the important point that the analysis of the figure need not always precede the application of the stereo processing algorithm in our perceptual machinery – the stereo algorithm can itself analyze the incoming patterns and find the figure. Similarly, these demonstrations show that binocular disparity by itself, without any other cues, is sufficient to yield the perception of variations in distance. We will return to stereopsis and its limitations in our discussion of depth cues in the next chapter.

23.6 How do distance cues combine?

We started this chapter with the intuition that information about distance must be lost in the projection from the three-dimensional physical world to the two dimensional retinal image. Instead, we have found that the two-dimensional retinal images, and the motor responses of the eye to these images, provide a remarkably large number of partial sources of distance information – distance cues. None of these cues is perfect – each one depends on the use of heuristics, with the consequence that each one will fail under conditions in which its individual heuristic fails. Somehow the different cues, working together, manage to cover for each others’ weaknesses, and allow us to reconstruct a sufficient approximation of the third dimension to function in the physical world.

This idea brings us to the question of cue integration. How do the different distance cues work together? How do they combine to give us a single, unambiguous perception of the distance to an object in the visual field? How is each cue weighted in combining its distance estimate with that from other cues? Are some cues more important than others?

These questions are vital to pose, but they have no simple answers. Part of the problem concerns paradigms. For example, many early perception researchers approached the question by an cue-isolation paradigm: try to render all but one of the distance cues ineffective, and test distance perception with that single cue. The problem with cue isolation experiments is that cues set to zero are not neutralized – instead, they are probably signalling that the stimulus consists of a flat plane, so one tends to find that the isolated cue (unavoidably opposed by all the ruled out cues) is not very effective.

Alternatively, we could try an cue-opposition paradigm: use stimuli containing one cue that indicates one distance order and another cue that indicates the opposite distance order, and see which one “wins out”. A good example would be to take a complex pictorial stereogram, and reverse the pictures presented to the two eyes. This maneuver will reverse the stereo cue while maintaining the pictorial cues in their original state. Which will dominate? Most people tend to assume that the stereo cue will dominate. Remarkably, most studies report that the pictorial cues dominate in this situation.

Many theorists working in the field of distance perception, however, would argue that neither of these paradigms has much value. The argument is that in perception in real-world situations, most of the distance cues vary together, and most of the time all of the distance cues signal much
the same distance. That is, tests of the cooperation of cues may be a more useful paradigm than cue isolation or opposition.

In short, many studies have been carried out on the question of the combination of distance cues. In general, the more the distance cues the more accurate the perception of distance. However, there is no simple summary of the outcomes of these studies, and no simple combination rule has emerged. We will return to the question of cue integration when we discuss the topic of depth (three-dimensional form) in the next chapter.

23.6.1 Cutting and Vishton’s analysis

Instead, we turn to an interesting summary and synthesis of information provided by James Cutting and Peter Vishton (1995). Cutting and Vishton combined logical (geometrical) analyses of the various distance cues with the results of empirical studies, and came up with some interesting educated guesses about the values and roles of different distance cues.

In our review of distance cues, we emphasized that different distance cues have different properties. Some cues – particularly accommodation and convergence – provide information about absolute distances. Others, like height in the visual field and the various perspective cues, provide information about relative distances. And yet others, particularly interposition, provide us only with ordinal information – they tell us about the order of objects in depth, but not the distances between them.

In addition to the absolute vs. relative vs. ordinal distinction, Cutting and Vishton emphasize that different cues have different levels of sensitivity, and different distance ranges over which they yield sufficient information to be useful in the calculation of distance. Their estimates of the sensitivities and useful distance ranges of the various cues are shown in Figure 23.9. (Unfortunately we don’t have space to justify these estimates in detail, but you can derive at least some of them from the principles of geometry).

On the basis of their summary, Cutting and Vishton point out that there are basically three patterns of variation of sensitivity with distance. In the first category are cues that are most sensitive at short distances, and diminish in sensitivity with increasing distance. As shown earlier in Figure 23.1B, convergence provides an example of this pattern; so does accommodation. In fact the ocular cues are never highly sensitive, and they become ineffective at a distance of perhaps two meters. Motion perspective and binocular disparity also show this pattern. They are both highly sensitive distance cues at near distances. But like accommodation and convergence, the sensitivities of motion perspective and binocular disparity also fall with distance, to become ineffective beyond about 10 to 30 meters. Height in the visual field is also in this category.

In the second category are distance cues that are equally effective at all distances. These cues are represented by horizontal lines in Cutting and Vishton’s graph. The cue of interposition is the best example. As we argued earlier, interposition is a unique distance cue in that it occurs any time one object lies in front of another. Hence it is a highly sensitive cue at all distance, as it can signal a miniscule difference in distance at any distance. (The downside is that it gives only ordinal information). Similarly, the cues of relative size and relative density (which we have called perspective cues) are also shown as horizontal lines. Cutting and Vishton have placed them at lower sensitivity levels than occlusion to indicate that perspective cues are not as sensitive to small differences in distance as occlusion is.

And in the third category, there is one cue whose efficacy actually increases with distance – the
Figure 23.9: Cutting and Vishton’s (1995) analysis. Summary of the sensitivities of the different distance cues, and how these sensitivities vary with the distance of an object. The horizontal axis in this graph is absolute distance – how far the object is from the observer. The vertical axis is sensitivity to changes in distance – $\Delta d$, the smallest detectable difference in distance, normalized to the absolute distance ($d$). By this measure, a difference threshold of 0.1 meter (10 cm) at 10 meters, and 1 meter at 100 meters, would both be plotted as sensitivities of 0.01. Cutting and Vishton also introduce the assumption, shown by the dashed horizontal line, that below a sensitivity of 0.1 a cue no longer provides useful information about distance. [Modified from Cutting and Vishton (1995, p. 80, Fig. 1).]
cue of aerial perspective. The effect of haze might distinguish two hills at 1 versus 2 kilometers but not two people at 10 versus 20 meters.

Cutting and Vishton’s graph has the advantage that it summarizes and represents information about many distance cues in a form that makes them commensurate, and should be helpful in leading to more sophisticated theorizing in the area of distance perception and cue integration. In their terms, the assembled information provides a network of constraints on models of distance perception.

At the end of their analysis, Cutting and Vishton introduce the interesting argument that different ranges of distance are of different functional importance to the perceiving organism. As shown at the top of Figure 23.9, they divide distance into three ranges. The first is personal space (out to perhaps 2 meters) – the space that immediately surrounds each person. The second is action space, the space within which we move and within which we direct our actions toward objects. And the third is vista space, the space at larger distance, beyond the range at which evolving humans could expect their actions to have any immediate impact. Cutting and Vishton suggest that environmental demands are different in these different distance ranges, and that different constellations of cues have been selected to combine optimally to serve perception and action in these different distance ranges. They suggest that studies of cue integration might profitably take the different distance ranges into account. Perhaps if studies of distance cue integration were sorted out by distance range we would find simpler rules of combination.

DT is also curious as to whether one could make a similar analysis in the time domain. That is, there are some situations in which rapid response to a motion cue would be critical to survival – the expansion pattern caused by the shadow of a descending hawk again comes to mind. It seems possible that different distance cues might be weighted differently depending on the immediacy of the task the observer is asked to perform. Taking off from Cutting and Vishton and from Milner and Goodale (1995), perhaps different distance cues are used in different weights for action than for perception.

Finally, we note that, despite all of the available distance cues, human distance perception is not always as good as modern man might wish, and we provide ourselves with many crutches to augment our perception of distance. For example, cameras are equipped with range finders precisely because our perception of distance is not accurate enough to guide optimal focusing of our cameras. And golfers have distance scopes that use a graded scale to judge the visual angle of the flag at the next hole, and in that way judge its distance.

### 23.7 Size constancy

#### 23.7.1 The size/distance invariance hypothesis

We turn now to the perception of size. Under ordinary viewing conditions, our perception of the sizes of objects is usually quite veridical: as stated earlier, we have a good degree of size constancy. The question is, how can this be? Since the size of the retinal image of an object varies inversely with its distance, the retinal image cannot supply us with information about the object’s physical size. The question of how we know about object size has been a classic problem in visual perception.

The usual answer has been to suppose that the relatively veridical perception of size comes about because of the presence of distance cues and the ensuing relatively veridical perception of distance. That is, in physical terms, the size of an object can be calculated if its retinal image size
and its distance are known:

\[ s_{\text{object}} = s_{\text{retina}} d. \]

A modified version of this equation, using perceived rather than physical variables, has commonly also been assumed to hold:

\[ s_{\text{perceived}} = s_{\text{retina}} d_{\text{perceived}}. \]

This equation is often called the size-distance invariance hypothesis. An implicit assumption in the size-distance invariance hypothesis is that the visual system calculates perceived size by first calculating a unified estimate of perceived distance, and then using the estimate of perceived distance in the calculation of perceived size.

### 23.7.2 A size constancy experiment

Within the context of the size/distance invariance hypothesis, many classical studies of size constancy take the approach of asking what distance cues are necessary for the veridical perception of size. A particularly nice experiment on distance cues and size constancy was carried out by Alfred H. Holway and Edwin G. Boring (1941). The setup they used is shown in Figure 23.10A.

Holway and Boring’s subjects were seated at the intersection of two corridors, and asked to compare the sizes of disks of light that appeared at various distances down the two corridors. Down the right corridor at a series of different distances, from 10 to 120 feet, were a series of different test stimuli with a fixed standard retinal image size, \( S_s \) in degrees of visual angle. In other words, the physical sizes of the test stimuli varied with their distances, \( D_s \), so that each one subtended the same visual angle of 1°, and hence each one produced the same retinal image size. Ten feet down the left corridor was a comparison disk, whose size, \( S_c \) (quantified in degrees of visual angle) was adjusted by the subject. On each trial, the experimenter presented one of the test disks. The subject’s task was to vary the size of the comparison disk until he judged that test and comparison disks were matched in physical size.

In the first condition of the experiment, Holway and Boring allowed the subjects to view the disks with natural binocular viewing with the lights on. In subsequent conditions different distance cues were sequentially eliminated. In the second condition, subjects viewed the disks monocularly, eliminating both retinal disparity and convergence cues. In the third condition, the accommodation cue was removed by having subjects view the disks through a small artificial pupil. And in the fourth condition, a “reduction tunnel” – baffles – were used to eliminate reflections from the walls (perspective cues).

The results of the experiment are shown in Figure 23.10B. With all distance cues available, the subjects matched the physical sizes of the disks very closely at all test distances – they showed good size constancy. Interestingly, constancy was preserved under monocular viewing, suggesting that neither binocular disparity nor convergence is necessary for size constancy. However, when accommodation was also eliminated, size constancy broke down rather dramatically, and the subjects’ matches began to approach matches based on retinal image size. And in the final condition, in which reflections from the walls were eliminated, the subjects’ matches conformed to matches of retinal image size. This experiment thus shows very clearly the dependence of size constancy on the particular combination of distance cues.

There are three more points to be made about size constancy. First, we need to be careful with our conclusions. It might be tempting to conclude that accommodation and perspective cues
Figure 23.10: Holway and Boring’s experiment. A. The experimental set-up. B. The results are shown by the matched comparison disk size ($S_c$ in degrees of visual angle) as a function of the distance to the standard disk ($D_s$ in feet). The diagonal dashed line shows the predicted size matches if the subject has perfect size constancy. The horizontal dashed line shows the prediction if the subject matches the visual angles subtended by the two stimuli. [From Holway and Boring (1941, A from Fig. 2, p. 24; B from Fig. 22, p. 34).]
are the distance cues most important to size constancy. Instead, the more valid conclusion is that these cues are sufficient to allow size constancy. In other words, had Holway and Boring done the experiment in the opposite direction, eliminating accommodation and perspective cues first, they probably would have found that the combination of disparity and convergence was also sufficient to yield good size constancy. In all probability size perception is opportunistic, using whatever distance cues are available in any given situation.

The second point is a subtle one about causality. It is tempting to conclude that since size constancy depends upon the presence of distance cues, it must be dependent upon a neural calculation of distance per se. This theoretical perspective is implicit in the equations at the beginning of this section. But in fact, Holway and Boring’s experiment has shown only that the same cues that contribute to the perception of distance also contribute to the perception of size. Whether distance is computed first and then enters into the calculation of size, or whether size and distance are calculated independently from the same base of distance cue information, is a difficult problem to sort out. DT suggests an open mind on this issue.

The third point concerns the cues we use to perceive the sizes of very distant objects, for which the only available distance cues are interposition, perspective and atmospheric perspective. Most people report that size constancy is non-existent at large distances: that everything looks tiny. Many authors suggest that we can see distant objects in either of two ways: as very small (that is, size constancy breaks down), or as having their normal sizes (that is, size constancy still holds). In the latter case, we probably use the more cognitively based cue of familiar size. That is, to have size constancy, we only need to perceive a distant house to be about the size that houses usually are, a distant person to be the size that people usually are, and so on. Such a cue necessarily comes with a heuristic: Things are their normal sizes. (Think about it. When would size constancy governed by this heuristic break down?)

Fourth, the size constancy experiment described here, like many, minimize pictorial cues to size such as retinal size, relative texture and familiarity. To fully understand size constancy, one has to also ask how does variation in pictorial cues combine with variation in distance cues. We will discuss some examples of such cue combinations in the following chapter on depth.

### 23.8 Summary: Cues and heuristics for distance

In summary: we started with the puzzle of how we can know about the distances of objects. We have found that the visual system has available perhaps a dozen different distance cues, each with its own peculiar set of limitations, and many with complex and fallible heuristics required for their use. Yet we can unite the various distance cues to judge the distances and sizes of objects with reasonable accuracy.

At the meta-level, this topic has given us a chance to work with the concept of cues and heuristics, and more generally the necessity for believing in heuristics. The long and short of distance perception is: there are cues contained within the incoming signal, but there is insufficient information to allow veridical distance estimates to be made. To yield up its secrets, each distance cue must be paired with the appropriate heuristic. Logically, these heuristics must be stored within our visual brains, and called upon when needed.

What a remarkable system! The entire scheme seems to DT so complex that it’s hard to imagine it being instantiated in an ensemble of neurons. The number and variety of cues, and the necessary
heuristics, are mind-boggling. Yet bumblebees can fly, and DT at least can only marvel at this jury-rigged but effective analysis of the visual scene.

In the next chapter we turn to the question of depth perception. We then review the available evidence for the analysis of depth and distance cues within the visual cortex.