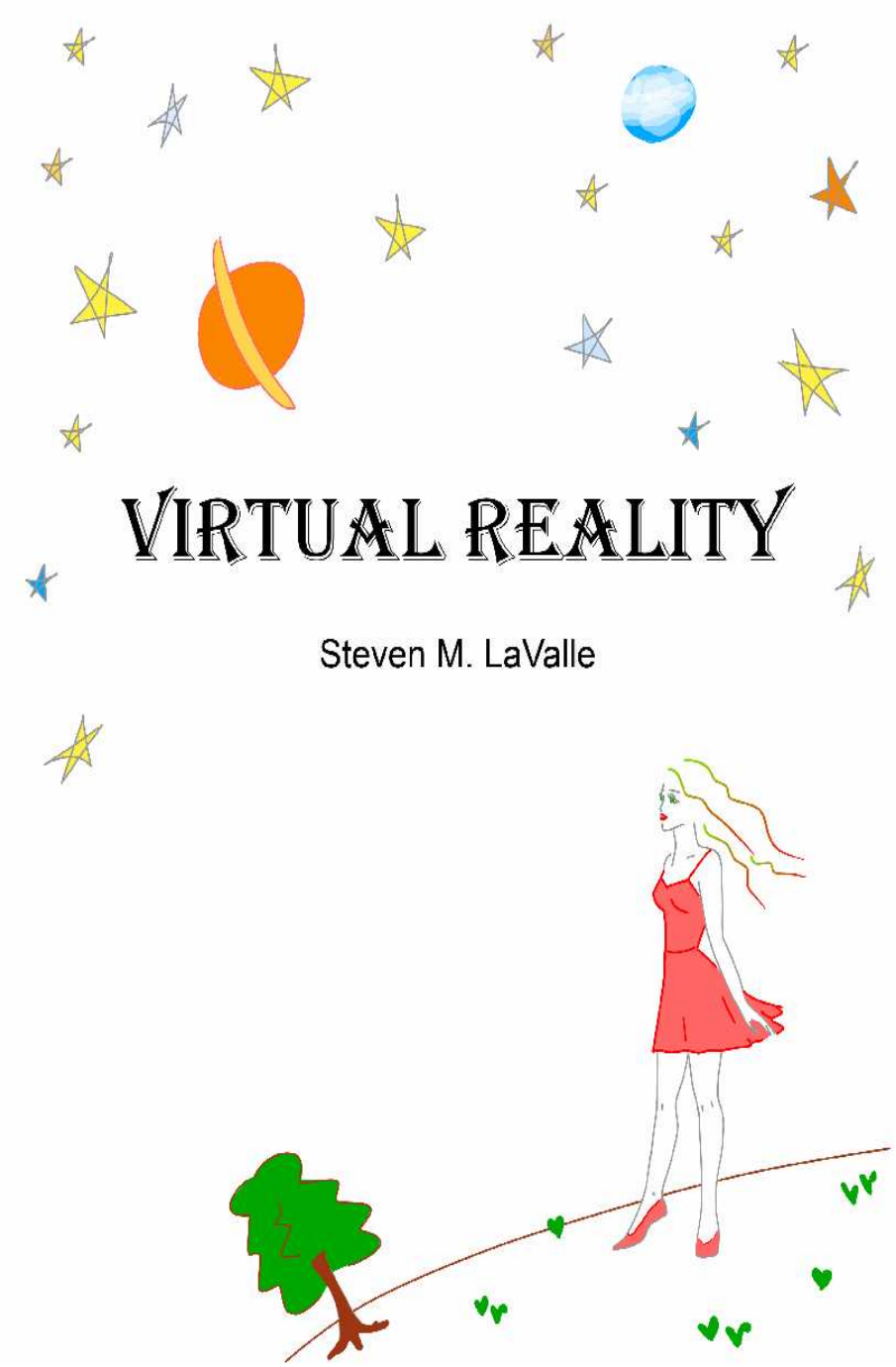


Chapter 6

Visual Perception

Steven M. LaValle

University of Oulu

Available for downloading at **http://lavalle.pl/vr/**

Chapter 6

Visual Perception

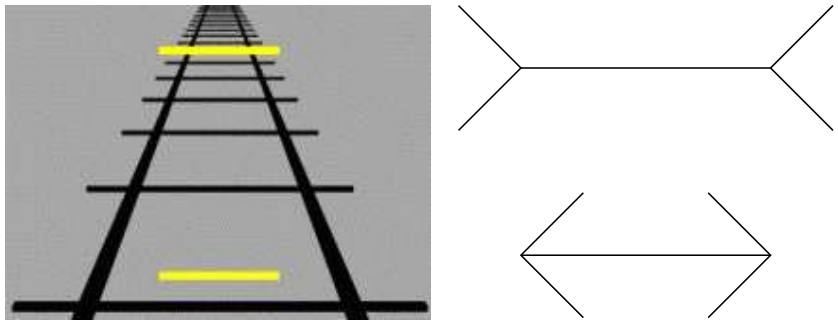
This chapter continues where Chapter 5 left off by transitioning from the *physiology* of human vision to *perception*. If we were computers, then this transition might seem like going from low-level hardware to higher-level software and algorithms. How do our brains interpret the world around us so effectively in spite of our limited biological hardware? To understand how we may be fooled by visual stimuli presented by a display, you must first understand how our we perceive or interpret the real world under normal circumstances. It is not always clear what we will perceive. We have already seen several optical illusions. VR itself can be considered as a grand optical illusion. Under what conditions will it succeed or fail?

Section 6.1 covers perception of the *distance* of objects from our eyes, which is also related to the perception of object *scale*. Section 6.2 explains how we perceive motion. An important part of this is the illusion of motion that we perceive from videos, which are merely a sequence of pictures. Section 6.3 covers the perception of color, which may help explain why displays use only three colors (red, green, and blue) to simulate the entire spectral power distribution of light (recall from Section 4.1). Finally, Section 6.4 presents a statistically based model of how information is combined from multiple sources to produce a perceptual experience.

6.1 Perception of Depth

This section explains how humans judge the distance from their eyes to objects in the real world using vision. The perceived distance could be *metric*, which means that an estimate of the absolute distance is obtained. For example, a house may appear to be about 100 meters away. Alternatively, the distance information could be *ordinal*, which means that the relative arrangement of visible objects can be inferred. For example, one house appears to be closer than another if it is partially blocking the view of the further one.

Figure 6.1: This painting uses a monocular depth cue called a *texture gradient* to enhance depth perception: The bricks become smaller and thinner as the depth increases. Other cues arise from perspective projection, including height in the visual field and retinal image size. ("Paris Street, Rainy Day," Gustave Caillebotte, 1877. Art Institute of Chicago.)



(a) (b)

Figure 6.2: Even simple line drawings provide significant cues. (a) The Ponzo illusion: The upper yellow bar appears to be longer, but both are the same length. (b) The Müller-Lyer illusion: The lower horizontal segment appears to be shorter than the one above, but they are the same length.

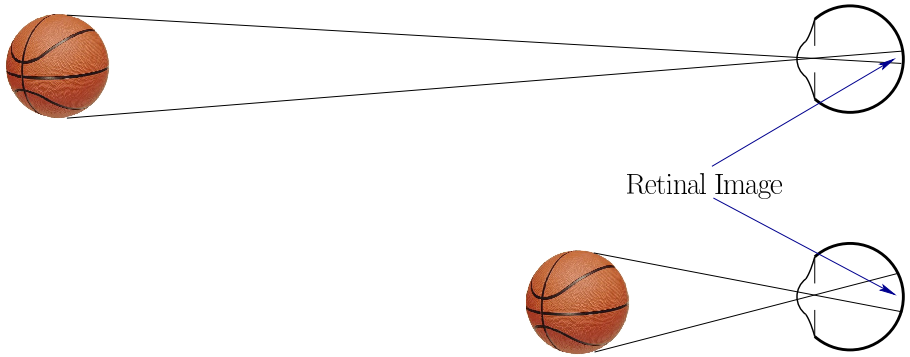


Figure 6.3: The retinal image size of a familiar object is a strong monocular depth cue. The closer object projects onto a larger number of photoreceptors, which cover a larger portion of the retina.

Monocular vs. stereo cues A piece of information that is derived from sensory stimulation and is relevant for perception is called a *sensory cue* or simply a *cue*. In this section, we consider only *depth cues*, which contribute toward depth perception. If a depth cue is derived from the photoreceptors or movements of a single eye, then it is called a *monocular depth cue*. If both eyes are required, then it is a *stereo depth cue*. There are many more monocular depth cues than stereo, which explains why we are able to infer so much depth information from a single photograph. Figure 6.1 shows an example. The illusions in Figure 6.2 show that even simple line drawings are enough to provide strong cues. Interestingly, the cues used by humans also work in computer vision algorithms to extract depth information from images [20].

6.1.1 Monocular depth cues

Retinal image size Many cues result from the geometric distortions caused by perspective projection; recall the "3D" appearance of Figure 1.23(c). For a familiar object, such as a human, coin, or basketball, we often judge its distance by how "large" is appears to be. Recalling the perspective projection math from Section 3.4, the size of the image on the retina is proportional to $1/z$, in which z is the distance from the eye (or the common convergence point for all projection lines). See Figure 6.3. The same thing happens when taking a picture with a camera: A picture of a basketball would occupy larger part of the image, covering more pixels, as it becomes closer to the camera. This cue is called *retinal image size*, and was studied in [4].

Two important factors exist. First, the viewer must be familiar with the object to the point of comfortably knowing its true size. For familiar objects, such as people or cars, our brains performance *size constancy scaling* by assuming that

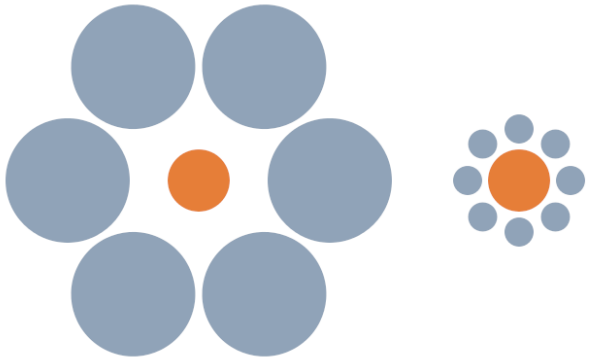


Figure 6.4: For the Ebbinghaus illusion, the inner disc appears larger when surrounded by smaller discs. The inner disc is the same size in either case. This may be evidence of discrepancy between the true visual angle (or retinal image size) and the *perceived* visual angle.

the distance, rather than the size, of the person is changing if they come closer. Size constancy falls of the general heading of *subjective constancy*, which appears through many aspects of perception, including shape, size, and color. The second factor is that, the object must be appear naturally so that it does not conflict with other depth cues.

If there is significant uncertainty about the size of an object, then knowledge of its distance should contribute to estimating its size. This falls under *size perception*, which is closely coupled to depth perception. Cues for each influence the other, in a way discussed in Section 6.4.

One controversial theory is that our *perceived visual angle* differs from the actual visual angle. The visual angle is proportional to the retinal image size. This theory is used to explain the illusion that the moon appears to be larger when it is near the horizon. For another example, see Figure 6.4.

Height in the visual field Figure 6.5(a) illustrates another important cue, which is the height of the object in the visual field. The Ponzo illusion in Figure 6.2(a) exploits this cue. Suppose that we can see over a long distance without obstructions. Due to perspective projection, the horizon is a line that divides the view in half. The upper half is perceived as the sky, and the lower half is the ground. The distance of objects from the horizon line corresponds directly to their distance due to perspective projection: The closer to the horizon, the further the perceived distance. Size constancy scaling, if available, combines with the height in the visual field, as shown in Figure 6.5(b).

(a) (b)

Figure 6.5: Height in visual field. (a) Trees closer to the horizon appear to be further away, even though all yield the same retinal image size. (b) Incorrect placement of people (the author and his son, Ethan) in the visual field illustrates *size constancy scaling*, which is closely coupled with depth cues. (Photo printed by permission of Nadia Inturias, Uyuni, Bolivia.)

Accommodation Recall from Section 4.4 that the human eye lens can change its optical power through the process of accommodation. For young adults, the amount of change is around 10D (diopters), but it decreases to less than 1D for adults over 50 years old. The ciliary muscles control the lens and their tension level is reported to the brain through efference copies of the motor control signal. This is the first depth cue that does not depend on signals generated by the photoreceptors.

Motion parallax Up until now, the depth cues have not exploited motions. If you have ever looked out of the side window of a fast-moving vehicle, you might have noticed that the nearby objects race by much faster than further objects. The relative difference in speeds is called *parallax* and is an important depth cue; see Figure 6.6. Even two images, from varying viewpoints within a short amount of time, provide strong depth information. Imagine trying to simulate a *stereo rig* of cameras by snapping one photo and quickly moving the camera sideways to snap another. If the rest of the world is stationary, then the result is roughly equivalent to having two side-by-side cameras. Pigeons frequently bob their heads back and forth to obtain stronger depth information than is provided by their pair of eyes. Finally, closely related to motion parallax is *optical flow*, which is a characterization of the rates at which features move across the retina. This will be revisited in Sections 6.2 and 8.4.

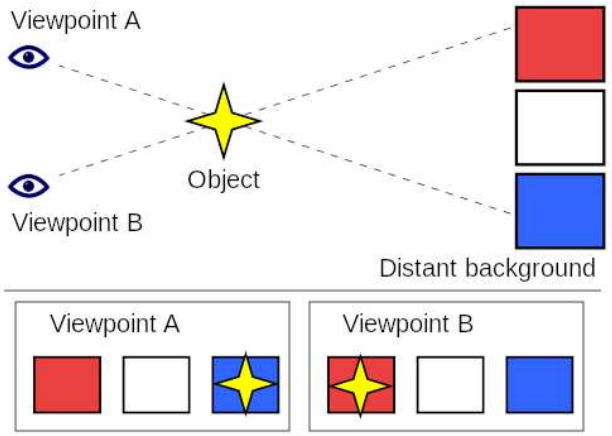
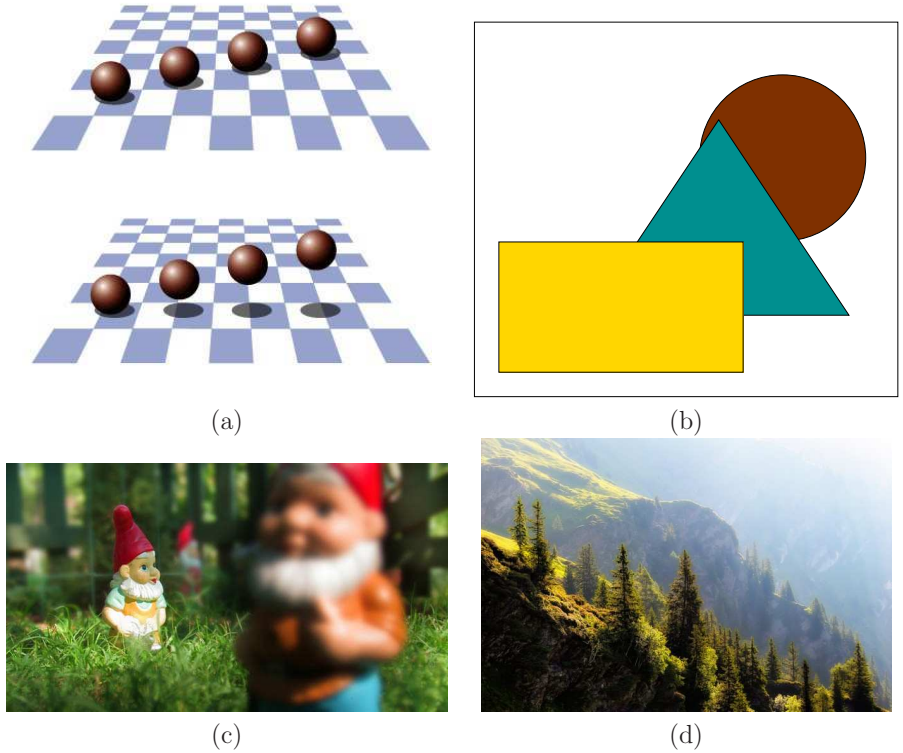


Figure 6.6: Motion parallax: As the perspective changes laterally, closer objects have larger image displacements than further objects. (Figure from Wikipedia.)

Other monocular cues Figure 6.7 shows several other monocular cues. As shown in Figure 6.7(a), shadows that are cast by a light source encountering an object provide an important cue. Figure 6.7(b) shows a simple drawing that provides an ordinal depth cue called *interposition* by indicating which objects are in front of others. Figure 6.7(c) illustrates the *image blur* cue, where levels are depth are inferred from the varying sharpness of focus. Figure 6.7(d) shows an *atmospheric cue* in which air humidity causes far away scenery to have lower contrast, thereby appearing to be further away.

6.1.2 Stereo depth cues

As you may expect, focusing both eyes on the same object enhances depth perception. Humans perceive a single focused image over a surface in space called the *horopter*; see Figure 6.8. Recall the vergence motions from Section 5.3. Similar to the accommodation cue case, motor control of the eye muscles for vergence motions provides information to the brain about the amount of convergence, thereby providing a direct estimate of distance. Each eye provides a different viewpoint, which results in different images on the retina. This phenomenon is called *binocular disparity*. Recall from (3.50) in Section 3.5 that the viewpoint is shifted to the right or left to provide a lateral offset for each of the eyes. The transform essentially shifts the virtual world to either side. The same shift would happen for a stereo rig of side-by-side cameras in the real world. However, the binocular disparity for humans is different because the eyes can rotate to converge, in addition to having a lateral offset. Thus, when fixating on an object, the retinal images between the left and right eyes may vary only slightly, but this nevertheless



(a) (b) (c) (d)

Figure 6.7: Several more monocular depth cues: (a) *Shadows* resolve ambiguous depth in the *ball and shadow illusion*. (b) The *interposition* of objects provides an ordinal depth cue. (c) Due to *image blur*, one gnome appears to be much closer than the others. (d) This scene provides an *atmospheric cue*: Some scenery is perceived to be further away because it has lower contrast.

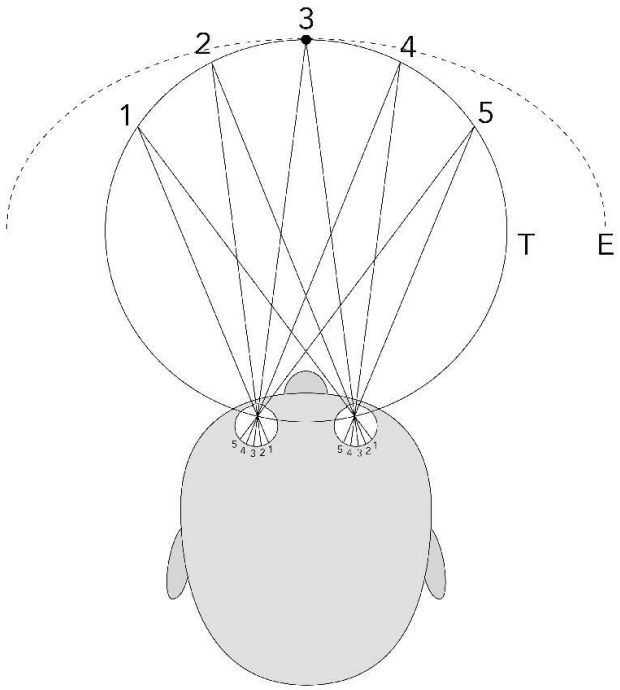


Figure 6.8: The *horopter* is the loci of points over which the eyes can converge and focus on a single depth. The T curve shows the theoretical horopter based on simple geometry. The E curve shows the empirical horopter, which is much larger and correspond to the region over which a single focused image is perceived. (Figure by Rainer Zenz.)

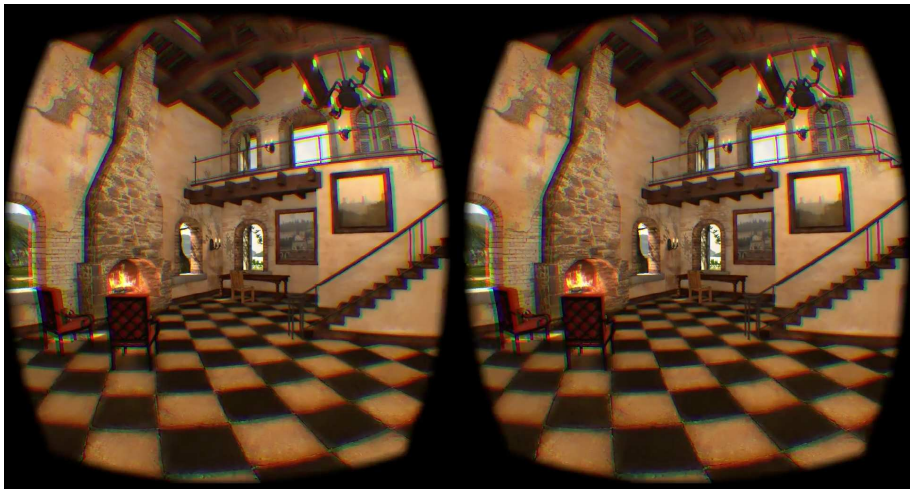


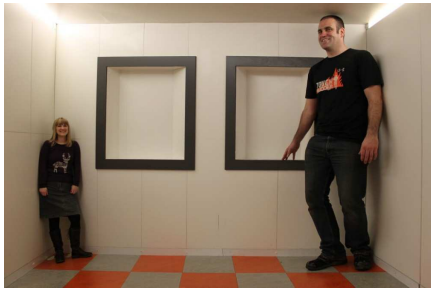
Figure 6.9: In the Tuscany demo from Oculus VR, there are not enough familiar objects to precisely resolve depth and size. Have you ever been to a villa like this? Are the floor tiles a familiar size? Is the desk too low?

provides a powerful cue used by the brain.

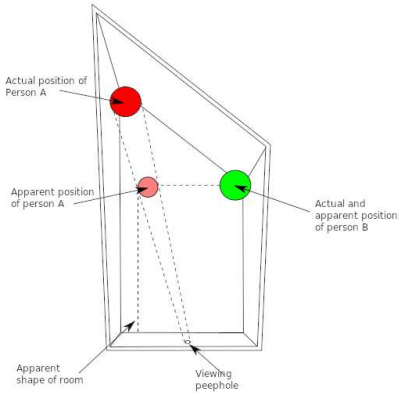
Furthermore, when converging on an object at one depth, we perceive double images of objects at other depths (although we usually pay no attention to it). This double-image effect is called *diplopia*. You can perceive it by placing your finger about 20cm in front of your face and converging on it. While fixating on your finger, you should perceive double images of other objects around the periphery. You can also stare into the distance while keeping your finger in the same place. You should see a double image of your finger. If you additionally roll your head back and forth, it should appear as if the left and right versions of your finger are moving up and down with respect to each other. These correspond to dramatic differences in the retinal image, but we are usually not aware of them because we perceive both retinal images as a single image.

6.1.3 Implications for VR

Incorrect scale perception A virtual world may be filled with objects that are not familiar to us in the real world. In many cases, they might resemble familiar objects, but their precise scale might be difficult to determine. Consider the Tuscany demo world from Oculus VR, shown in Figure 6.9. The virtual villa is designed to be inhabited with humans, but it is difficult to judge the relative sizes and distances of objects because there are not enough familiar objects. Further complicating the problem is that the user's height in VR might not match his height in the virtual world. Is the user too short, or is the world too big? A



(a)



(b)

Figure 6.10: The Ames room: (a) Due to incorrect depth cues, incorrect size perception results. (b) The room is designed so that it only appears to be rectangular after perspective projection is applied. One person is actually much further away than the other. (Figure by Alex Valavanis.)

common and confusing occurrence is that the user might be sitting down in the real world, but standing in the virtual world. An additional complication occurs if the interpupillary distance (recall from Section 4.4) is not matched with the real world. For example, if the user's pupils are 64mm apart in the real world but only 50mm apart in the virtual world, then the virtual world will seem much larger, which dramatically affects depth perception. Likewise, if the pupils are very far apart, the user could either feel enormous or the virtual world might seem small. Imagine simulating a Godzilla experience, where the user is 200 meters tall and the entire city appears to be a model. It is fine to experiment with such scale and depth distortions in VR, but it is important to understand their implications on the user's perception.

Mismatches In the real world, all of the depth cues work together in harmony. We are sometimes fooled by optical illusions that are designed to intentionally cause inconsistencies among cues. Sometimes a simple drawing is sufficient. Figure 6.10 shows an elaborate illusion that requires building a distorted room in the real world. It is perfectly designed so that when viewed under perspective projection from one location, it appears to be a rectangular box. Once our brains accept this, we unexpectedly perceive the size of people changing as they walk across the room! This is because all of the cues based on perspective appear to be functioning correctly. Section 6.4 may help you to understand how multiple cues are resolved, even in the case of inconsistencies.

In a VR system, it is easy to cause mismatches and in many cases they are unavoidable. Recall from Section 5.4 that vergence-accommodation mismatch occurs in VR headsets. Another source of mismatch may occur from imperfect head tracking. If there is significant latency, then the visual stimuli will not appear in the correct place at the expected time. Furthermore, many tracking systems track the head orientation only. This makes it impossible to use motion parallax as a depth cue if the user moves from side to side without any rotation. To preserve most depth cues based on motion, it is important to track head *position*, in addition to orientation; see Section 9.3. Optical distortions may cause even more mismatch.

Monocular cues are powerful! A common misunderstanding among the general public is that depth perception enabled by stereo cues alone. We are bombarded with marketing of "3D" movies and *stereo displays*. The most common instance today is the use of circularly polarized *3D glasses* in movie theaters so that each eye receives a different image when looking at the screen. VR is no exception to this common misunderstanding. CAVE systems provided 3D glasses with an active shutter inside so that alternating left and right frames can be presented to the eyes. Note that this cuts the frame rate in half. Now that we have comfortable headsets, presenting separate visual stimuli to each eye is much simpler. One drawback is that the rendering effort (the subject of Chapter 7) is doubled, although this can be improved through some context-specific tricks.

As you have seen in this section, there are many more monocular depth cues than stereo cues. Therefore, it is wrong to assume that the world is perceived as "3D" *only if* there are stereo images. This insight is particularly valuable for leveraging captured data from the real world. Recall from Section 1.1 that the virtual world may be synthetic or captured. It is generally more costly to create synthetic worlds, but it is then simple to generate stereo viewpoints (at a higher rendering cost). On the other hand, capturing panoramic, monoscopic images and movies is fast and inexpensive (examples were shown in Figure 1.9). There are already smartphone apps that stitch pictures together to make a panoramic photo, and direct capture of panoramic video is likely to be a standard feature on smartphones within a few years. By recognizing that this content is sufficiently "3D" due to the wide field of view and monocular depth cues, it becomes a powerful way to create VR experiences. There are already hundreds of millions of images in Google Street View, shown in Figure 6.11, which can be easily viewed using Google Cardboard or other headsets. They provide a highly immersive experience with substantial depth perception, even though there is no stereo. There is even strong evidence that stereo displays cause significant fatigue and discomfort, especially for objects at a close depth [13, 14]. Therefore, one should think very carefully about the use of stereo. In many cases, it might be more time, cost, and trouble than it is worth to obtain the stereo cues when there may already be sufficient monocular cues for the VR task or experience.

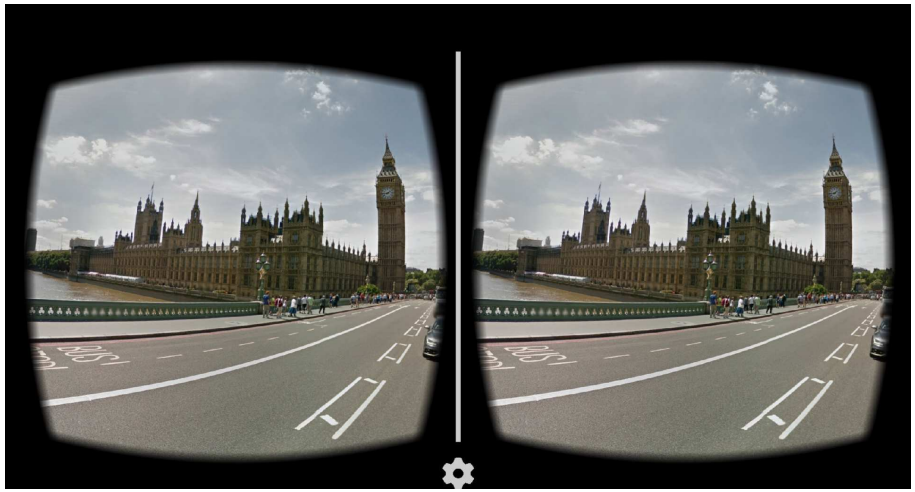


Figure 6.11: In Google Cardboard and other VR headsets, hundreds of millions of panoramic Street View images can be viewed. There is significant depth perception, even when the same image is presented to both eyes, because of monoscopic depth cues.

6.2 Perception of Motion

We rely on our vision to perceive motion for many crucial activities. One use is to separate a moving figure from a stationary background. For example, a camouflaged animal in the forest might only become noticeable when moving. This is clearly useful whether humans are the hunter or the hunted. Motion also helps people to assess the 3D structure of an object. Imagine assessing the value of a piece of fruit in the market by rotating it around. Another use is to visually guide actions, such as walking down the street or hammering a nail. VR systems have the tall order of replicating these uses in a virtual world in spite of limited technology. Just as important as the perception of motion is the perception of non-motion, which we called *perception of stationarity* in Section 2.3. For example, if we apply the VOR by turning our heads, then do the virtual world objects move correctly on the display so that they appear to be stationary? Slight errors in time or image position might inadvertently trigger the perception of motion.

6.2.1 Detection mechanisms

Reichardt detector Figure 6.12 shows a neural circuitry model, called a *Reichardt detector*, which responds to directional motion in the human vision system. Neurons in the ganglion layer and LGN detect simple features in different spots in the retinal image. At higher levels, motion detection neurons exist that respond

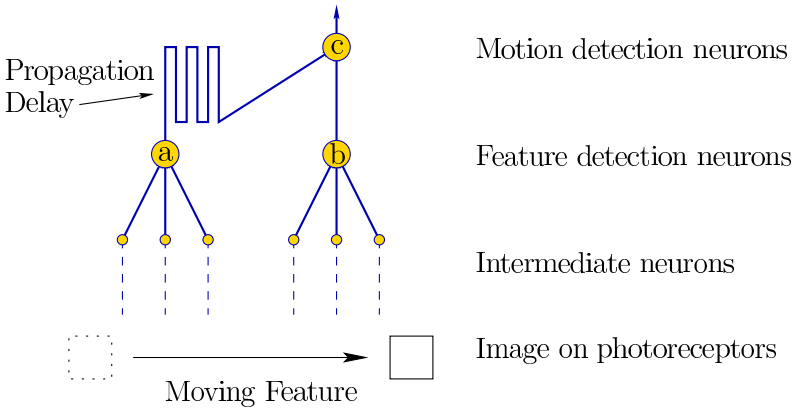


Figure 6.12: The neural circuitry directly supports motion detection. As the image feature moves across the retina, nearby feature detection neurons (labeled a and b) activate in succession. Their outputs connect to motion detection neurons (labeled c). Due to different path lengths from a and b to c, the activation signal arrives at different times. Thus, c activates when the feature was detected by a slightly before being detected by b.

when the feature moves from one spot on the retina to another nearby spot. The motion detection neuron activates for a feature speed that depends on the difference in path lengths from its input neurons. It is also sensitive to a particular direction of motion based on the relative locations of the receptive fields of the input neurons. Due to the simplicity of the motion detector, it can be easily fooled. Figure 6.12 shows a feature moving from left to right. Suppose that a train of features moves from right to left. Based on the speed of the features and the spacing between them, the detector may inadvertently fire, causing motion to be perceived in the opposite direction. This is the basis of the *wagon-wheel effect*, for which a wheel with spokes or a propeller may appear to be rotating in the opposite direction, depending on the speed. The process can be further disrupted by causing eye vibrations from humming [17]. This simulates stroboscopic conditions, which discussed in Section 6.2.2. Another point is that the motion detectors are subject to adaptation. Therefore, several illusions exist, such as the *waterfall illusion* [1] and the *spiral aftereffect*, in which incorrect motions are perceived due to aftereffects from sustained fixation [1, 9].

From local data to global conclusions Motion detectors are *local* in the sense that a tiny portion of the visual field causes each to activate. In most cases, data from detectors across large patches of the visual field are *integrated* to indicate coherent motions of rigid bodies. (An exception would be staring at pure analog TV static.) All pieces of a rigid body move through space according

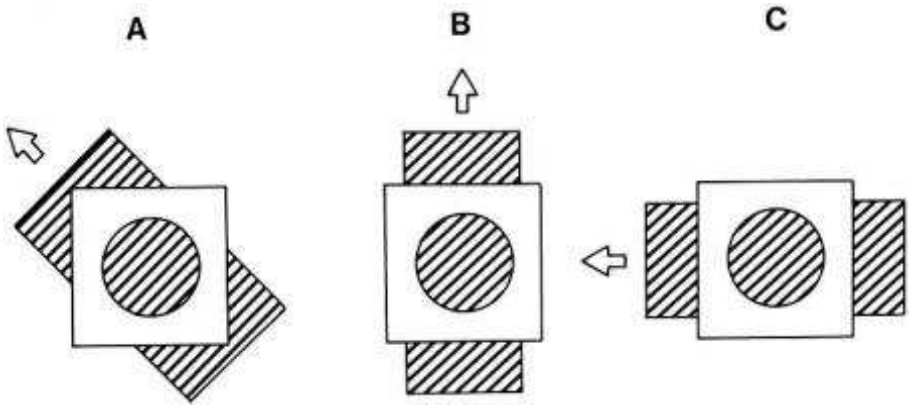


Figure 6.13: Due to local nature of motion detectors, the *aperture problem* results. The motion of the larger body is ambiguous when perceived through a small hole because a wide range of possible body motions could produce the same effect inside of the hole. An incorrect motion inference usually results.

to the equations from Section 3.2. This coordinated motion is anticipated by our visual system to match common expectations. If too much of the moving body is blocked, then the *aperture problem* results, which is shown in Figure 6.13. A clean mathematical way to describe the global motions across the retina is by a *vector field*, which assigns a velocity vector at every position. The global result is called the *optical flow*, which provides powerful cues for both object motion and self motion. The latter case results in *vection*, which is a leading cause of VR sickness; see Sections 8.4 and 10.2 for details.

Distinguishing object motion from observer motion Figure 6.14 shows two cases that produce the same images across the retina over time. In Figure 6.14(a), the eye is fixed while the object moves by. In Figure 6.14(b), the situation is reversed: The object is fixed, but the eye moves. The brain uses several cues to differentiate between these cases. Saccadic suppression, which was mentioned in Section 5.3, hides vision signals during movements; this may suppress motion detectors in the second case. Another cue is provided by proprioception, which is the body's ability to estimate its own motions due to motor commands. This includes the use of eye muscles in the second case. Finally, information is provided by large-scale motion. If it appears that the entire scene is moving, then the brain assumes the most likely interpretation, which is that the user must be moving. This is why the haunted swing illusion, shown in Figure 2.20, is so effective.

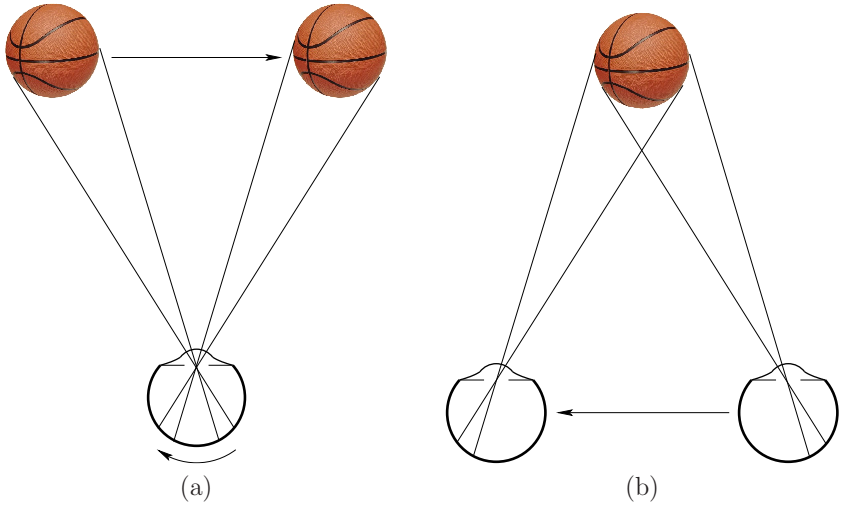


Figure 6.14: Two motions that cause equivalent movement of the image on the retina: (a) The eye is fixed and the object moves; (b) the eye moves while the object is fixed. Both of these are hard to achieve in practice due to eye rotations (smooth pursuit and VOR).

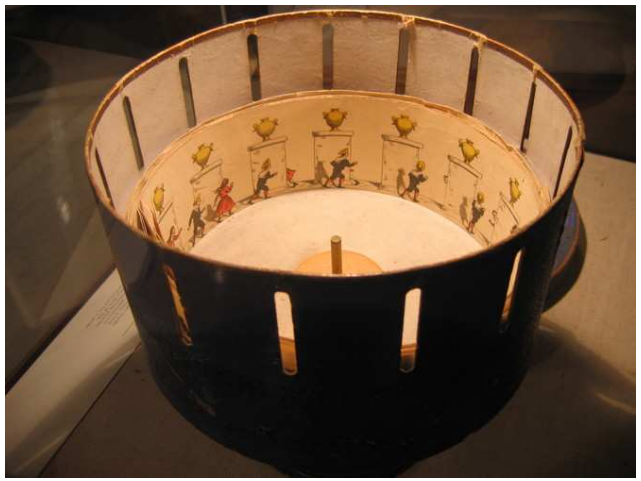


Figure 6.15: The *zoetrope* was developed in the 1830s and provided stroboscopic apparent motion as images became visible through slits in a rotating disc.

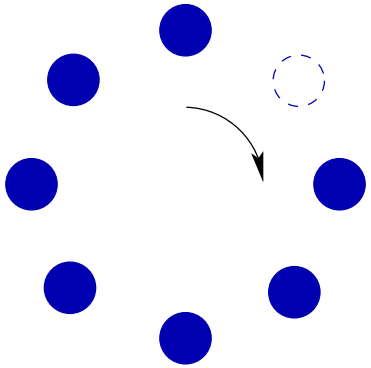


Figure 6.16: The *phi phenomenon* and *beta movement* are physiologically distinct effects in which motion is perceived [22, 19]. In the sequence of dots, one is turned *off* at any give time. A different dot is turned *off* in each frame, following a clockwise pattern. At a very low speed (2 FPS), beta movement triggers a motion perception of each *on* dot directly behind the *off* dot. The *on* dot appears to jump to the position of the *off* dot. At a higher rate, such as 15 FPS, there instead appears to be a moving hole; this corresponds to the phi phenomenon.

6.2.2 Stroboscopic apparent motion

Nearly everyone on Earth has seen a motion picture, whether through a TV, smartphone, or movie screen. The motions we see are an illusion because a sequence of still pictures is being flashed onto the screen. This phenomenon is called *stroboscopic apparent motion*; it was discovered and refined across the 19th century. The *zoetrope*, shown in Figure 6.15 was developed around 1834. It consists of a rotating drum with slits that allow each frame to be visible for an instant while the drum rotates. In Section 1.3, Figure 1.24 showed the Horse in Motion film from 1878.

Why does this illusion of motion work? An early theory, which has largely been refuted in recent years, is called *persistence of vision*. The theory states that images persist in the vision system during the intervals in between frames, thereby causing them to be perceived as continuous. One piece of evidence against this theory is that images persist in the visual cortex for around 100ms, which implies that the 10 FPS (Frames Per Second) is the slowest speed that stroboscopic apparent motion would work; however, it is also perceived down to 2 FPS [19]. Another piece of evidence against the persistence of vision is the existence of stroboscopic apparent motions that cannot be accounted for by it. The *phi phenomenon* and *beta movement* are examples of motion perceived in a sequence of blinking lights, rather than flashing frames (see Figure 6.16). The most likely reason that stroboscopic apparent motion works is that it triggers the neural motion detection circuitry illustrated in Figure 6.12 [8, 11].

FPS	Occurrence
2	Stroboscopic apparent motion starts
10	Ability to distinguish individual frames is lost
16	Old home movies; early silent films
24	Hollywood classic standard
25	PAL television before interlacing
30	NTSC television before interlacing
48	Two-blade shutter; proposed new Hollywood standard
50	Interlaced PAL television
60	Interlaced NTSC television; perceived flicker in some displays
72	Three-blade shutter; minimum CRT refresh rate for comfort
90	Modern VR headsets; no more discomfort from flicker
1000	Ability to see zipper effect for fast, blinking LED
5000	Cannot perceive zipper effect

Figure 6.17: Various frame rates and comments on the corresponding stroboscopic apparent motion. Units are in Frames Per Second (FPS).

Frame rates How many frames per second are appropriate for a motion picture? The answer depends on the intended use. Figure 6.17 shows a table of significant frame rates from 2 to 5000. Stroboscopic apparent motion begins at 2 FPS. Imagine watching a security video at this rate. It is easy to distinguish individual frames, but the motion of a person would also be perceived. Once 10 FPS is reached, the motion is obviously more smooth and we start to lose the ability to distinguish individual frames. Early silent films ranged from 16 to 24 FPS. The frame rates were often fluctuating and were played at a faster speed than they were filmed. Once sound was added to film, incorrect speeds and fluctuations in the speed were no longer tolerated because both sound and video needed to be synchronized. This motivated playback at the fixed rate of 24 FPS, which is still used today by the movie industry. Personal video cameras remained at 16 or 18 FPS into the 1970s. The famous Zapruder film of the Kennedy assassination in 1963 was taken at 18.3 FPS. Although 24 FPS may be enough to perceive motions smoothly, a large part of cinematography is devoted to ensuring that motions are not so fast that jumps are visible due to the low frame rate.

Such low frame rates unfortunately lead to perceptible *flicker* as the images rapidly flash on the screen with black in between. This motivated several workarounds. In the case of movie projectors, two-blade and three-blade shutters were invented so that they would show each frame two or three times, respectively. This enabled movies to be shown at 48 FPS and 72 FPS, thereby reducing discomfort from flickering. Analog television broadcasts in the 20th century were at 25 (*PAL standard*) or 30 FPS (*NTSC standard*), depending on the country. To double the frame rate and reduce perceived flicker, they used *interlacing* to draw half the image in one frame time, and then half in the other. Every other

horizontal line is drawn in the first half, and the remaining lines are drawn in the second. This increased the frames rates on television screens to 50 and 60 FPS. The game industry has used 60 FPS standard target for smooth game play.

As people started sitting close to giant CRT monitors in the early 1990s, the flicker problem became problematic again because sensitivity to flicker is stronger at the periphery. Furthermore, even when flicker cannot be directly perceived, it may still contribute to fatigue or headaches. Therefore, frame rates were increased to even higher levels. A minimum acceptable ergonomic standard for large CRT monitors was 72 FPS, with 85 to 90 FPS being widely considered as sufficiently high to eliminate most flicker problems. The problem has been carefully studied by psychologists under the heading of *flicker fusion threshold*; the precise rates at which flicker is perceptible or causes fatigue depends on many factors in addition to FPS, such as position on retina, age, color, and light intensity. Thus, the actual limit depends on the kind of display, its size, specifications, how it is used, and who is using it. Modern LCD and LED displays, used as televisions, computer screens, and smartphone screens, have 60, 120, and even 240 FPS.

The story does not end there. If you connect an LED to a pulse generator (put a resistor in series), then flicker can be perceived at much higher rates. Set the pulse generator to produce a square wave at several hundred Hz. Go to a dark room and hold the LED in your hand. If you wave it around so fast that your eyes cannot track it, then the flicker becomes perceptible as a zipper pattern. Let this be called the *zipper effect*. This happens because each time the LED pulses on, it is imaged in a different place on the retina. Without image stabilization, it appears as an array of lights. The faster the motion, the further apart the images will appear. The higher the pulse rate (or FPS), the closer together the images will appear. Therefore, to see the zipper effect at very high speeds, you need to move the LED very quickly. It is possible to see the effect for a few thousand FPS.

6.2.3 Implications for VR

Unfortunately, VR systems require much higher display performance than usual. We have already seen in Section 5.4 that much higher resolution is needed so that pixels and aliasing artifacts are not visible. The next problem is that higher frame rates are needed in comparison to ordinary television or movie standards of 24 FPS or even 60 FPS. To understand why, see Figure 6.18. The problem is easiest to understand in terms of the *perception of stationarity*, which was mentioned in Section 2.3. Fixate on a nearby object and yaw your head to the left. Your eyes should then rotate to the right to maintain the object in a fixed location on the retina, due to the VOR (Section 5.3). If you do the same while wearing a VR headset and fixating on an object in the virtual world, then the image of the object needs to shift across the screen while you turn your head. Assuming that the pixels instantaneously change at each new frame time, the image of the

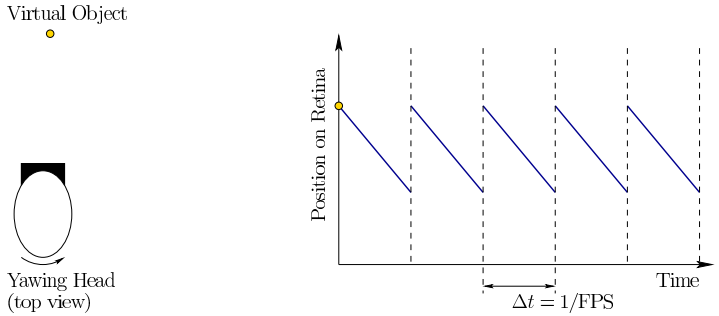


Figure 6.18: A problem with *perception of stationarity* under stroboscopic apparent motion: The image of a feature slips across the retina in a repeating pattern as the VOR is performed.

virtual object will slip across the retina as shown in Figure 6.18. The result is a kind of *judder* in which the object appears to be wobbling from side to side with high frequency but small amplitude.

The problem is that each feature is fixed on the screen for too long when ideally it should be moving continuously across the screen. At 60 FPS, it is fixed for 16.67ms during each frame (in an idealized setting, which ignores scanout issues from Section 5.4). If the screen is instead turned on for only one or two milliseconds for each frame, and then made black during the remaining times, then the amount of retinal image slip is greatly reduced. This display mode is called *low persistence*, and is shown in Figure 6.19(a). The short amount of time that the display is illuminated is sufficient for the photoreceptors to collect enough photons to cause the image to be perceived. The problem is that at 60 FPS in low-persistence mode, flicker is perceived, which can lead to fatigue or headaches. This can be easily perceived at the periphery in a bright scene in the Samsung Gear VR headset. If the frame rate is increased to 90 FPS or above, then the adverse side effects of flicker subside for nearly everyone. If the frame rate is increased to 500 FPS or beyond, then it would not even need to flicker, as depicted in Figure 6.19(b).

One final point is that fast pixel switching speed is implied in the Figure 6.19. In a modern OLED display panel, the pixels can reach their target intensity values in less than 0.1ms. However, many LCD displays change pixel values much more slowly. The delay to reach the target intensity may be as long as 20ms, depending on the amount and direction of intensity change. In this case, a fixed virtual object appears to smear or blur in the direction of motion. This was easily observable in the Oculus Rift DK1, which used an LCD display panel.

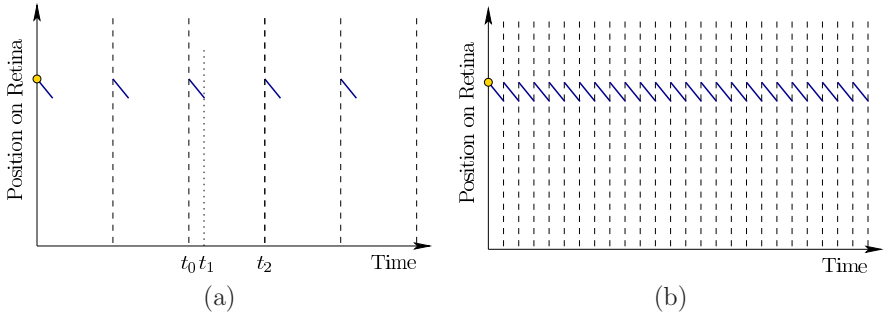


Figure 6.19: An engineering solution to reduce retinal image slip: (a) Using *low persistence*, the display is lit for a short enough time to trigger photoreceptors ($t_1 - t_0$) and then blanked for the remaining time ($t_2 - t_1$). Typically, $t_1 - t_0$ is around one to two milliseconds. (b) If the frame rate were extremely fast (at least 500 FPS), then the blank interval would not be needed.

Figure 6.20: In 2014, this dress photo became an Internet sensation as people were unable to agree upon whether it was "blue and black" or "white and gold", which are strikingly different perceptions of color.

6.3 Perception of Color

What makes an object "purple", "pink", or "gray"? Color perception is unusual because it is purely the result of our visual physiology and neural structures, rather than something that can be measured in the physical world. In other words, "It's all in your head." If two people have comparable color perception systems, then they can discuss colors using commonly agreed upon names while they perceive an object as having the same color. This contrasts other perception topics such as motion, depth, and scale, all of which correspond to measurable quantities in the surrounding world. The size of an object or the speed of its motion relative to some frame could be determined by instrumentation. Humans would be forced to agree on the numerical outcomes regardless of how their individual perceptual systems are functioning.

The dress Figure 6.20 illustrates this point with the *dress color illusion*. It was worn by Cecilia Bleasdale and became an Internet meme when millions of people quickly began to argue about the color of the dress. Based on the precise combination of colors and lighting conditions, its appearance fell on the boundary of what human color perceptual systems can handle. About 57% perceive it as blue and black (correct), 30% percent perceive it as white and gold, 10% perceive blue and brown, and 10% could switch between perceiving any of the color combinations [6].

Dimensionality reduction Recall from Section 4.1 that light energy is a jumble of wavelengths and magnitudes that form the spectral power distribution. Figure 4.6 provided an illustration. As we see objects, the light in the environment is reflected off of surfaces in a wavelength-dependent way according to the spectral distribution function (Figure 4.7). As the light passes through our eyes and is focused onto the retina, each photoreceptor receives a jumble of light energy that contains many wavelengths. Since the power distribution is a function of wavelength, the set of all possible distributions is a *function space*, which is generally infinite-dimensional. Our limited hardware cannot possibly sense the entire function. Instead, the rod and cone photoreceptors sample it with a bias toward certain target wavelengths, as was shown in Figure 5.3 of Section 5.1. The result is a well-studied principle in engineering called *dimensionality reduction*. Here, the infinite-dimensional space of power distributions collapses down to a 3D *color space*. It is no coincidence that human eyes have precisely three types of cones, and that our RGB displays target the same colors as the photoreceptors.

Yellow = Green + Red To help understand this reduction, consider the perception of "yellow". According to the visible light spectrum (Figure 4.5), yellow has a wavelength of about 580nm. Suppose we had a pure light source that shines light of exactly 580nm wavelength onto our retinas with no other wavelengths.

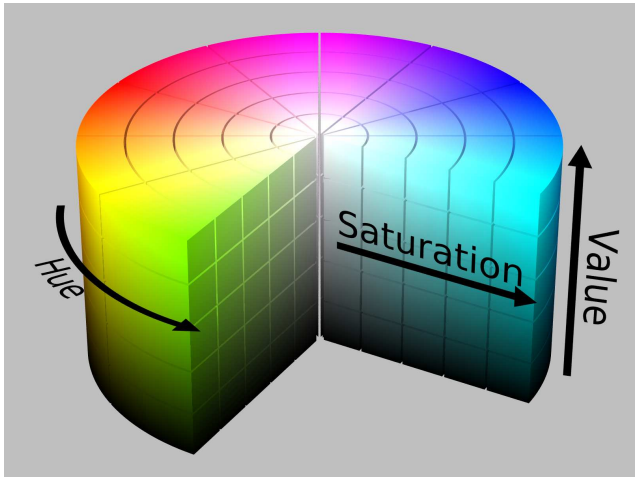


Figure 6.21: One representation of the HSV color space, which involves three parameters: hue, saturation, and value (brightness). (Figure by Wikipedia user SharkD.)

The spectral distribution function would have a spike at 580nm and be zero everywhere else. If we had a cone with peak detection at 580nm and no sensitivity to other wavelengths, then it would perfectly detect yellow. Instead, we perceive yellow by activation of both green and red cones because their sensitivity regions (Figure 5.3) include 580nm. It should then be possible to generate the same photoreceptor response by sending a jumble of light that contains precisely two wavelengths: 1) Some "green" at 533nm, and 2) some "red" at 564nm. If the magnitudes of green and red are tuned so that the green and red cones activate in the same way as they did for pure yellow, then it becomes impossible for our visual system to distinguish the green/red mixture from pure yellow. Both are perceived as "yellow". This matching of colors from red, green and blue components is called *metamerism*. Such a blending is precisely what is done on a RGB display to produce yellow. Suppose the intensity of each color ranges from 0 (dark) to 255 (bright). Red is produced by RGB= $(255, 0, 0)$, and green is RGB= $(0, 255, 0)$. These each activate one LED (or LCD) color, thereby producing a pure red or green. If both are turned on, then yellow is perceived. Thus, yellow is RGB= $(255, 255, 0)$.

Color spaces For convenience, a parameterized *color space* is often defined. One of the most common in computer graphics is called *HSV*, which has the following three components (Figure 6.21):

- The *hue*, which corresponds directly to the perceived color, such as "red" or "green".

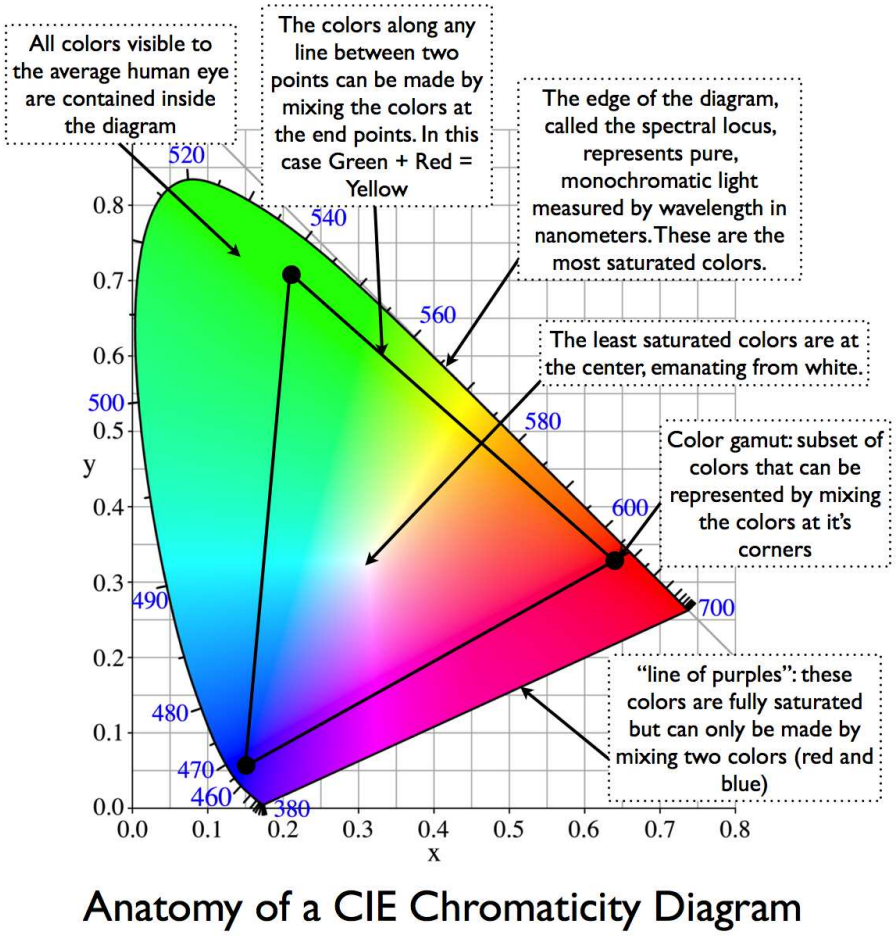


Figure 6.22: 1931 CIE color standard with RGB triangle. This representation is correct in terms of distances between perceived colors. (Figure by Jeff Yurek, Nanosys.)

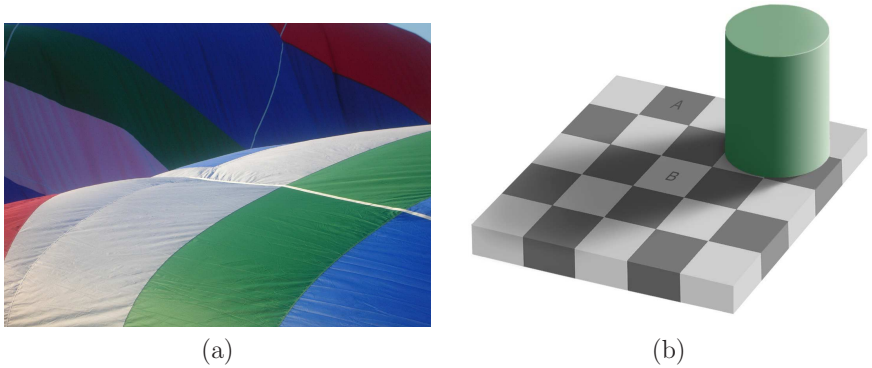
- The *saturation*, which is the purity of the color. In other words, how much energy is coming from wavelengths other than the wavelength of the hue?
- The *value*, which corresponds to the brightness.

There are many methods to scale the HSV coordinates, which distort the color space in various ways. The RGB values could alternatively be used, but are sometimes more difficult for people to interpret.

It would be ideal to have a representation in which the distance between two points corresponds to the amount of perceptual difference. In other words, as two points are further apart, our ability to distinguish them is increased. The distance should correspond directly to the amount of distinguishability. Vision scientists designed a representation to achieve this, resulting in the 1931 *CIE color standard* shown in Figure 6.22. Thus, the CIE is considered to be undistorted from a perceptual perspective. It is only two-dimensional because it disregards the brightness component, which is independent of color perception according to color matching experiments [8].

Mixing colors Suppose that we have three pure sources of light, as in that produced by an LED, in red, blue, and green colors. We have already discussed how to produce yellow by blending red and green. In general, most perceptible colors can be matched by a mixture of three. This is called *trichromatic theory* (or *Young-Helmholtz theory*). A set of colors that achieves this is called *primary colors*. Mixing all three evenly produces perceived *white light*, which on a display is achieved as RGB= $(255, 255, 255)$. Black is the opposite: RGB= $(0, 0, 0)$. Such light mixtures follow a linearity property. Suppose primary colors are used to perceptually match power distributions of two different light sources. If the light sources are combined, then their intensities of the primary colors need only to be added to obtain the perceptual match for the combination. Furthermore, the overall intensity can be scaled by multiplying the red, green, and blue components without affecting the perceived color. Only the perceived brightness may be changed.

The discussion so far has focused on *additive mixtures*. When mixing paints or printing books, colors mix subtractively because the spectral reflectance function is being altered. When starting with a white canvass or sheet of paper, virtually all wavelengths are reflected. Painting a green line on the page prevents all wavelengths other than green from being reflected at that spot. Removing all wavelengths results in black. Rather than using RGB components, printing presses are based on CMYK, which correspond to cyan, magenta, yellow, and black. The first three are pairwise mixes of the primary colors. A black component is included to reduce the amount of ink wasted by using the other three colors to subtractively produce black. Note that the targeted colors are observed only if the incoming light contains the targeted wavelengths. The green line would appear green under pure, matching green light, but might appear black under pure blue light.



(a) (b)

Figure 6.23: (a) The perceived hot air balloon colors are perceived the same regardless of the portions that are in direct sunlight or in a shadow. (Figure by Wikipedia user Shanta.) (b) The *checker shadow illusion* from Section 2.3 is explained by the *lightness constancy* principle as the shadows prompt compensation of the perceived lightness. (Figure by Adrian Pingstone; original by Edward H. Adelson.)

Constancy The dress in Figure 6.20 showed an extreme case that results in color confusion across people due to the strange lighting conditions. Ordinarily, human color perception is surprisingly robust to the source of color. A red shirt appears to be red whether illuminated under indoor lights at night or in direct sunlight. These correspond to vastly different cases in terms of the spectral power distribution that reaches the retina. Our ability to perceive an object as having the same color over a wide variety of lighting conditions is called *color constancy*. Several perceptual mechanisms allow this to happen. One of them is *chromatic adaptation*, which results in a shift in perceived colors due to prolonged exposure to specific colors. Another factor in the perceived color is the expectation from the colors of surrounding objects. Furthermore, memory about how objects are usually colored in the environment biases our interpretation.

The constancy principle also appears without regard to particular colors. Our perceptual system also maintains *lightness constancy* so that the overall brightness levels appear to be unchanged, even after lighting conditions are dramatically altered; see Figure 6.23(a). Under the *ratio principle* theory, only the ratio of reflectances between objects in a scene are perceptually maintained, whereas the overall amount of reflected intensity is not perceived. Further complicating matters, our perception of object lightness and color are maintained as the scene contains uneven illumination. A clear example is provided from shadows cast by one object onto another. Our perceptual system accounts for the shadow and adjusts our perception of the object shade or color. The checker shadow illusion shown in Figure 6.23 is caused by this compensation due to shadows.

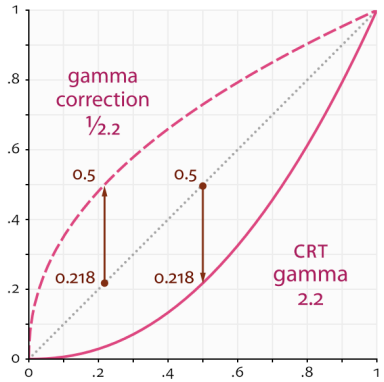


Figure 6.24: Gamma correction is used to span more orders of magnitude in spite of a limited number of bits. The transformation is $v' = cv^{\gamma}$, in which c is constant (usually $c = 1$) and γ controls the nonlinearity of the correction or distortion.

Display issues Displays generally use RGB lights to generate the palette of colors and brightness. Recall Figure 4.36, which showed the subpixel mosaic of individual component colors for some common displays. Usually, the intensity of each R, G, and B value is set by selecting an integer from 0 to 255. This is a severe limitation on the number of brightness levels, as stated in Section 5.4. One cannot hope to densely cover all seven orders of magnitude of perceptible light intensity. One way to enhance the amount of contrast over the entire range is to perform *gamma correction*. In most displays, images are encoded with a gamma of about 0.45 and decoded with a gamma of 2.2.

Another issue is that the set of all available colors lies inside of the triangle formed by R, G, and B vertices. This limitation is shown for the case of the *sRGB* standard in Figure 6.22. Most the CIE is covered, but many colors that humans are capable of perceiving cannot be generated on the display.

6.4 Combining Sources of Information

Throughout this chapter, we have seen perceptual processes that combine information from multiple sources. These could be cues from the same sense, as in the numerous monocular cues used to judge depth. Perception may also combine information from two or more senses. For example, people typically combine both visual and auditory cues when speaking face to face. Information from both sources makes it easier to understand someone, especially if there is significant background noise. We have also seen that information is integrated over time, as in the case of saccades being employed to fixate on several object features. Finally, our memories and general expectations about the behavior of the sur-

rounding world bias our conclusions. Thus, information is integrated from prior expectations and the reception of many cues, which may come from different senses at different times.

Statistical decision theory provides a useful and straightforward mathematical model for making choices that incorporate prior biases and sources of relevant, observed data. It has been applied in many fields, including economics, psychology, signal processing, and computer science. One key component is *Bayes' rule*, which specifies how the *prior* beliefs should be updated in light of new observations, to obtain *posterior* beliefs. More formally, the "beliefs" are referred as *probabilities*. If the probability takes into account information from previous information, it is called a *conditional probability*. There is no room to properly introduce *probability theory* here; only the basic ideas are given to provide some intuition without the rigor. For further study, find an online course or classic textbook (for example, [16]).

Let

$$H = \{h_1, h_2, \ldots, h_n\} \tag{6.1}$$

be a set of *hypotheses* (or interpretations). Similarly, let

$$C = \{c_1, c_2, \ldots, c_m\} \tag{6.2}$$

C be a set of possible outputs of a *cue detector*. For example, the cue detector might output the eye color of a face that is currently visible. In this case C is the set of possible colors:

$$C = \{\text{BROWN}, \text{BLUE}, \text{GREEN}, \text{HAZEL}\}. \tag{6.3}$$

Modeling a face recognizer, H would correspond to the set of people familiar to the person.

We want to calculate probability values for each of the hypotheses in H. Each probability value must lie between 0 to 1, and the sum of the probability values for every hypothesis in H must sum to one. Before any cues, we start with an assignment of values called the *prior distribution*, which is written as $P(h)$. The "P" denotes that it is a probability function or assignment; $P(h)$ means that an assignment has been applied to every h in H. The assignment must be made so that

$$P(h_1) + P(h_2) + \cdots + P(h_n) = 1, \tag{6.4}$$

and $0 \leq P(h_i) \leq 1$ for each i from 1 to n.

The prior probabilities are generally distributed across the hypotheses in a diffuse way; an example is shown in Figure 6.25(a). The likelihood of any hypothesis being true before any cues is proportional to its frequency of occurring naturally, based on evolution and the lifetime of experiences of the person. For example, if you open your eyes at a random time in your life, what is the likelihood of seeing a human being versus a wild boar?

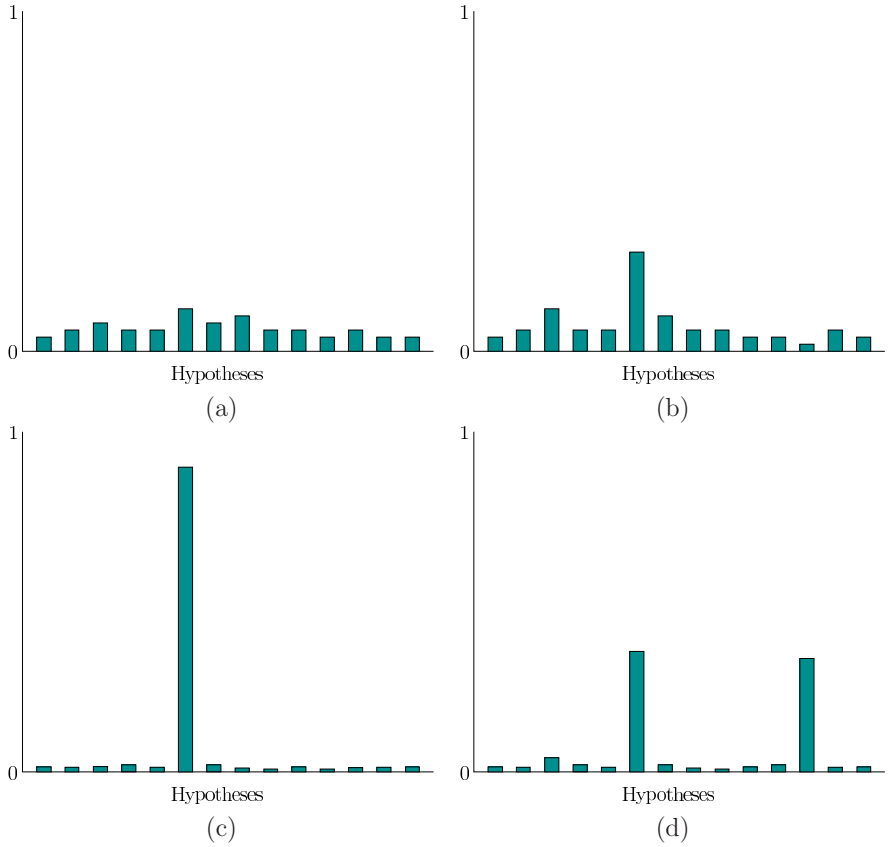


Figure 6.25: Example probability distributions: (a) A possible prior distribution. (b) Preference for one hypothesis starts to emerge after a cue. (c) A peaked distribution, which results from strong, consistent cues. (d) Ambiguity may result in two (or more) hypotheses that are strongly favored over others; this is the basis of multistable perception.

Under normal circumstances (not VR!), we expect that the probability for the correct interpretation will rise as cues arrive. The probability of the correct hypothesis should pull upward toward 1, effectively stealing probability mass from the other hypotheses, which pushes their values toward 0; see Figure 6.25(b). A "strong" cue should lift the correct hypothesis upward more quickly than a "weak" cue. If a single hypothesis has a probability value close to 1, then the distribution is considered *peaked*, which implies high confidence; see Figure 6.25(c). In the other direction, inconsistent or incorrect cues have the effect of diffusing the probability across two or more hypotheses. Thus, the probability of the correct hypothesis may be lowered as other hypotheses are considered plausible and receive higher values. It may also be possible that two alternative hypotheses remain strong due to ambiguity that cannot be solved from the given cues; see Figure 6.25(d).

To take into account information from a cue, a *conditional distribution* is defined, which is written as $P(h \mid c)$. This is spoken as "the probability of h given c." This corresponds to a probability assignment for all possible combinations of hypotheses and cues. For example, it would include $P(h_2 \mid c_5)$, if there are at least two hypotheses and five cues. Continuing our face recognizer, this would look like $P(\text{BARACK OBAMA} \mid \text{BROWN})$, which should be larger than $P(\text{BARACK OBAMA} \mid \text{BLUE})$ (he has brown eyes).

We now arrive at the fundamental problem, which is to calculate $P(h \mid c)$ after the cue arrives. This is accomplished by *Bayes' rule*:

$$P(h \mid c) = \frac{P(c \mid h)P(h)}{P(c)}. \tag{6.5}$$

The denominator can be expressed as

$$P(c) = P(c \mid h_1)P(h_1) + P(c \mid h_2)P(h_2) + \cdots + P(c \mid h_n)P(h_n), \tag{6.6}$$

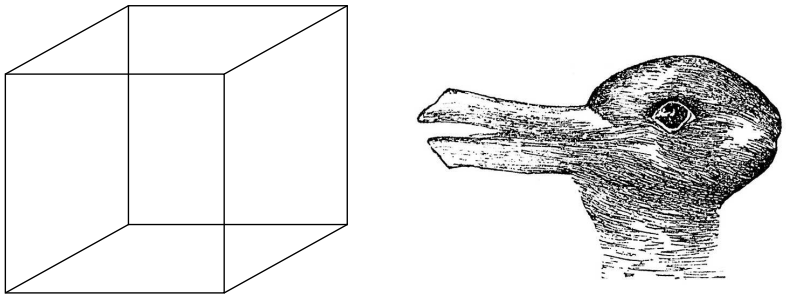
or it can be ignored it as a normalization constant, at which point only relative likelihoods are calculated instead of proper probabilities.

The only thing accomplished by Bayes' rule was to express $P(h \mid c)$ in terms of the prior distribution $P(h)$ and a new conditional distribution $P(c \mid h)$. The new conditional distribution is easy to work with in terms of modeling. It characterizes the likelihood that each specific cue will appear given that the hypothesis is true.

What if information arrives from a second cue detector? In this case, (6.5) is applied again, but $P(h \mid c)$ is now considered the prior distribution with respect to the new information. Let $D = \{d_1, d_2, \ldots, d_k\}$ represent the possible outputs of the new cue detector. Bayes' rule becomes

$$P(h \mid c, d) = \frac{P(d \mid h)P(h \mid c)}{P(d|c)}. \tag{6.7}$$

Above, $P(d \mid h)$ makes what is called a *conditional independence* assumption: $P(d \mid h) = P(d \mid h, c)$. This is simpler from a modeling perspective. More



(a) (b)

Figure 6.26: (a) The Necker cube, studied in 1832 by Swiss crystallographer Louis Albert Necker. (b) The *rabbit duck illusion*, from the 23 October 1892 issue of Fliegende Blätter.

generally, all four conditional parts of (6.7) should contain c because it is given before d arrives. As information from even more cues becomes available, Bayes' rule is applied again as many times as needed. One difficulty that occurs in practice and modeled here is *cognitive bias*, which corresponds to numerous ways in which humans make irrational judgments in spite of the probabilistic implications of the data.

Multistable perception In some cases, our perceptual system may alternate between two or more conclusions. This is called *multistable perception*, for which the special case of two conclusions is called *bistable perception*. Figure 6.26(a) shows two well-known examples. For the *Necker cube*, it is ambiguous which cube face that is parallel to the viewing plane is in the foreground. It is possible to switch between both interpretations, resulting in bistable perception. Figure 6.26(b) shows another example, in which people may see a rabbit or a duck at various times. Another well-known example is called the *spinning dancer illusion* by Nobuyuki Kayahara. In that case, the silhouette of a rotating dancer is shown and it is possible to interpret the motion as clockwise or counterclockwise.

McGurk effect The *McGurk effect* is an experiment that clearly indicates the power of integration by mixing visual and auditory cues [10]. A video of a person speaking is shown with the audio track dubbed so that the spoken sounds do not match the video. Two types of illusions were then observed. If "ba" is heard and "ga" is shown, then most subjects perceive "da" being said. This corresponds to a plausible fusion of sounds that explains the mismatch, but does not correspond to either original cue. Alternatively, the sounds may combine to produce a perceived "bga" in the case of "ga" on the sound track and "ba" on the visual track.

Implications for VR Not all senses are taken over by VR. Thus, conflict will arise because of mismatch between the real and virtual worlds. As stated several times, the most problematic case of this is vection, which is a sickness-causing conflict between visual and vestibular cues arising from apparent self motion in VR while remaining stationary in the real world; see Section 8.4. As another example of mismatch, the user's body may sense that it is sitting in a chair, but the VR experience may involve walking. There would then be a height mismatch between the real and virtual worlds, in addition to mismatches based on proprioception and touch. In addition to mismatches among the senses, imperfections in the VR hardware, software, content, and interfaces cause inconsistencies in comparison with real-world experiences. The result is that incorrect or untended interpretations may arise. Even worse, such inconsistencies may increase fatigue as human neural structures use more energy to interpret the confusing combination. In light of the McGurk effect, it is easy to believe that many unintended interpretations or perceptions may arise from a VR system that does not provide perfectly consistent cues.

VR is also quite capable of generating new multistable perceptions. One example, which actually occurred in the VR industry, involved designing a popup menu. Suppose that users are placed into a dark environment and a large menu comes rushing up to them. A user may perceive one of two cases: 1) the menu approaches the user, or 2) the user is rushing up to the menu. The vestibular sense should be enough to resolve whether the user is moving, but the visual sense is overpowering. Prior knowledge about which is happening helps yield the correct perception. Unfortunately, if the wrong interpretation is made, then VR sickness in increased due to the sensory conflict. This, our perceptual system could by tricked into an interpretation that is worse for our health! Knowledge is one of many VR sickness factors discussed in Section 12.3.

Further Reading

As with Chapter 5, much of the material from this chapter appears in textbooks on sensation and perception [5, 8, 23] For a collection of optical illusions and their explanations, see [12]. For more on motion detection, see Chapter 7 of [8]. Related to this is the history of motion pictures [3, 2].

To better understand the mathematical foundations of combining cues from multiple sources, look for books on Bayesian analysis and statistical decision theory. For example, see [15] and Chapter 9 of [7]. An important issue is adaptation to VR system flaws through repeated use [18, 21]. This dramatically effects the perceptual results and fatigue from mismatches, and is a form of perceptual learning, which will be discussed in Section 12.1.

Bibliography

[1] H. B. Barlow and R. M. Hill. Evidence for a physiological explanation of the waterfall illusion. *Nature*, 200:1345–1347, 1963.

[2] D. Bordwell and K. Thompson. *Film History: An Introduction, 3rd Ed.* McGraw-Hill, New York, NY, 2010.

[3] K. Brown. Silent films: What was the right speed? *Sight and Sound*, 49(3):164–167, 1980.

[4] W. C. Gogel. An analysis of perceptions from changes in optical size. *Perception and Psychophysics*, 60(5):805–820, 1998.

[5] E. B. Goldstein. *Sensation and Perception, 9th Ed.* Wadsworth, Belmont, CA, 2014.

[6] R. Lafer-Sousa, K. L. Hermann, and B. R. Conway. Striking individual differences in color perception uncovered by the dress photograph. *Current Biology*, 25(13):R545–R546, 2015.

[7] S. M. LaValle. *Planning Algorithms*. Cambridge University Press, Cambridge, U.K., 2006. Available at http://planning.cs.uiuc.edu/.

[8] G. Mather. *Foundations of Sensation and Perception*. Psychology Press, Hove, UK, 2008.

[9] G. Mather, F. Verstraten, and S. Anstis. *The motion aftereffect: A modern perspective*. MIT Press, Boston, MA, 1998.

[10] H. McGurk and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264:746–748, 1976.

[11] A. Mikami, W. T. Newsome, and R. H. Wurtz. Motion selectivity in macaque visual cortex. II. Spatiotemporal range of directional interactions in MT and V1. *Journal of Neurophysiology*, 55:1328–1339, 1986.

[12] J. Ninio. *The Science of Illusions*. Cornell University Press, Ithaca, NY, 2001.

[13] E. Peli. The visual effects of head-mounted display (HMD) are not distinguishable from those of desk-top computer display. *Vision Research*, 38(13):2053–2066, 1998.

[14] E. Peli. Optometric and perceptual issues with head-mounted displays. In P. Mouroulis, editor, *Visual instrumentation : optical design and engineering principles*. McGraw-Hill, New York, NY, 1999.

[15] C. P. Robert. *The Bayesian Choice, 2nd. Ed.* Springer-Verlag, Berlin, 2001.

[16] S. Ross. *A First Course in Probability, 9th Ed.* Pearson, New York, NY, 2012.

[17] W. Rushton. Effect of humming on vision. *Nature*, 216:1173–1175, 2009.

[18] A. R. Seitz, J. E. Nanez, S. R. Halloway, and T. Watanabe. Perceptual learning of motion leads to faster-flicker perception. *Journal of Vision*, 6(6):158, 2015.

[19] R. M. Steinman, Z. Pizlo, and F. J. Pizlo. Phi is not beta, and why Wertheimer's discovery launched the Gestalt revolution. *Vision Research*, 40(17):2257–2264, 2000.

[20] R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag, Berlin, 2010.

[21] R. B. Welch and B. J. Mohler. Adapting to virtual environments. In K. S. Hale and K. M. Stanney, editors, *Handbook of Virtual Environments, 2nd Edition*. CRC Press, Boca Raton, FL, 2015.

[22] M. Wertheimer. Experimentelle Studien über das Sehen von Bewegung (Experimental Studies on the Perception of Motion). *Zeitschrift für Psychologie*, 61:161–265, 1912.

[23] J. M. Wolfe, K. R. Kluender, and D. M. Levi. *Sensation and Perception, 4th Ed.* Sinauer, Sunderland, MA, 2015.