UNIVERSITEIT AMSTERDAM

VU Research Portal

Depth Perception

Brenner, E.; Smeets, Jeroen B.

Link to publication in VU Research Portal

# CHAPTER 9

Depth Perception

ELI BRENNER AND JEROEN B. J. SMEETS

People use sensory information to guide their interactions with the environment. One thing that people, like most animals, almost always want to know about objects in their environment is where those objects are. If someone is touching an object, he or she can feel where it is by combining where the object touches their skin with knowledge about their own posture. In such haptic localization there is no fundamental difference between judgments of distance and judgments of azimuth or elevation (Figure 9.1). Similarly, there is no fundamental difference between haptic judgments of an object's dimension in depth and of its lateral or vertical dimension. However, people often want to know an object's position and dimensions before they touch it. In that case they usually rely on visual information. Occasionally they might rely on auditory information, such as when they try to find a friend who they hear talking. In auditory localization, azimuth is judged from differences between signals' arrival times and between their intensities in the two ears, elevation is judged from how the shape of the outer ear affects the sounds' spectral content and echoes sounds, and distance is primarily judged from the intensity of the sound (although reverberation and spectral content probably also provide some information). In visual localization, azimuth and elevation (relative to the head) are judged

by combining the position of the object's retinal image with information about the orientation of the eyes (Figure 9.2). Judging the distance is much more complicated. This chapter deals with how people visually judge distances in depth.

We consider three different aspects of judging distances in depth, which we refer to as judging *distance*, *depth*, and *depth ordering*. We use distance to refer to the distance from the observer expressed in some metric such as meters, number of steps required to reach it, number of eye-heights away, or any other measure that completely specifies the position. We use depth to refer to distances between structures. This could refer to distances between objects, but also to distances within an object, as when referring to a single object's extent in the viewing direction. Depth could be estimated by comparing two judged distances, but there are also ways to directly judge depth. Direct measures of depth are generally more precise than judgments of distance, but they do not directly provide metric information (as is explained). They need to be scaled by judgments of distance before they can provide information about the actual separation in depth, as one might need for judging whether an object is too big to be grasped. We use depth to refer to the scaled judgments. In some cases, such as when judging whether
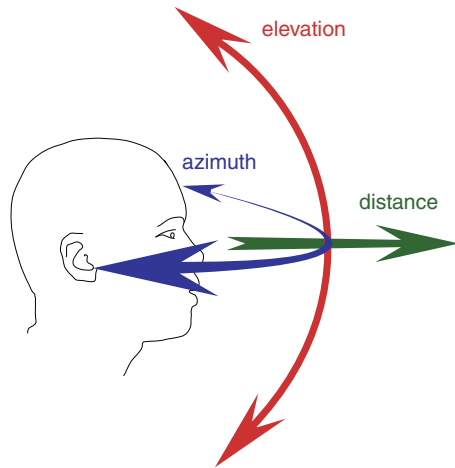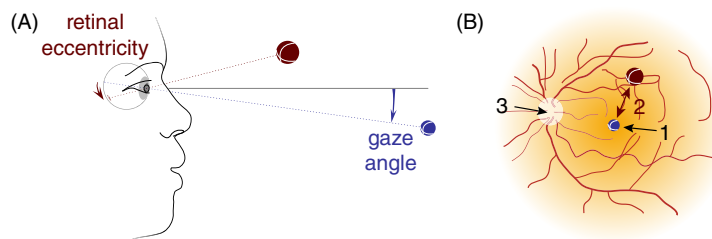
Figure 9.1   Positions relative to the head can be expressed in terms of a distance and direction from the head. The direction can be expressed as an azimuth (left or right of straight ahead) and an elevation (up or down with respect to straight ahead).

do provide information about such depth ordering.

Vision is based on two-dimensional retinal images of the three-dimensional world. These two-dimensional images represent what you can see at a certain moment in any direction. The azimuth and elevation of all visible structures in the environment are represented by the positions of their images in each of the two eyes. Differences between structures' azimuths and elevations are directly evident from the positions of their images on the retina. Distance would appear to be lost in the transformation from the three-dimensional world to the two-dimensional retinal images.

The fact that humans have two eyes with largely overlapping fields of view gives them the possibility to recover distance: Given a structure's direction with respect to both eyes, one can theoretically determine its distance through triangulation. Theoretically, such triangulation is enough to determine the distance to all objects that are simultaneously visible to both eyes, but the presence of this chapter specifically devoted to depth perception is an indication that in practice, judging distances is more complicated.

a structure is drawn on a surface or is an object lying on the surface, or when judging whether a wasp is in your room or safely behind the glass of the window, it is enough to know whether a structure is closer to you than another structure, without knowing how much closer. The direct measures of depth

Figure 9.2   Two balls' positions relative to the head (A) and their images' positions on the retina (B). The position on the retina that is stimulated by light reflected by an object (such as the dark red ball) depends on the object's azimuth and elevation (only the elevation is visible in (A) and on where one is looking (here, at the blue ball). Conversely, where one is looking (gaze angle in A) and the retinal eccentricity of the object in question (arrow 2 in B) can be combined to retrieve the object's azimuth and elevation. When looking directly at an object of interest, the retinal eccentricity is negligible (the object's image falls on the fovea, indicated by arrow 1 in B, which is the part of the retina with the highest density of photoreceptors), so the azimuth and elevation correspond with the direction of gaze. Arrow 3 in B indicates the optic nerve (blind spot).

The main reason for this complication is that the precision with which one can estimate the visual direction with respect to each eye is limited. Therefore, humans do not *only* rely on differences between the visual directions with respect to the two eyes, but also on other sources of visual information about distance and depth. The different sources of information are known as depth cues.

A fundamental difference between recovering distances by triangulation and doing so with the aid of other depth cues is that triangulation does not require any prior knowledge or assumptions about the world. As long as one can identify a structure's image in both eyes, one can judge the structure's distance. Other cues depend critically on regularities in the world. After presenting a short overview of the cues that are available for judging distances and depths, subsequent sections discuss these cues and the corresponding assumptions in greater depth. We then discuss how the cues' resolutions and the assumptions on which they rest influence the way in which they are combined for various kinds of judgments. When we mention *assumptions*, we conceive of them as being used unconsciously and of being the result of regularities in our everyday experience. However, people are certainly sometimes aware of the assumptions being violated, although this does not necessarily decrease the extent to which they rely on them (Muller, Brenner, & Smeets, 2008). Moreover, because reliable depth perception is not only advantageous for humans, but undoubtedly enhances the chances of survival in many species, there might be an innate component to the use of some assumptions.

Roughly speaking, there are three kinds of cues. The first kind consists of cues that are present in a single retinal image. These are sometimes also known as pictorial cues, because they can be captured and reproduced in pictures. The only direct information about distance that is present in a single image is that if you can see something in a certain direction, this thing must be the nearest object in that direction (except when looking through transparent surfaces). Thus, you do not know its distance, but you know that it is closer in depth than any other structure in that direction. If the object hides part of another object from view, it must be nearer than that object. If the other object occludes part of the object in question, the other object must be nearer. Thus, the fact that something is visible in a single image provides some information about the depth order: It tells you that it is the nearest object in that direction.

Other pictorial cues include various properties that are related to perspective, including image size, texture gradients, and height in the visual field, as well as contrast and blur (reviewed extensively in Sedgwick, 1986). These cues rely on regularities that normally exist in the world around us. Examples of regularities are that textures are isotropic, that shapes are symmetrical, that objects rest on surfaces rather than hovering in mid-air, and so on. Cues that rely on such regularities provide incorrect information when the regularities are violated. An obvious example of such a violation is when we take a photograph of a slanted textured surface, in which case the structure of the texture in the picture is consistent with a surface that is slanted with respect to the actual surface of the photograph.

The second kind of cues consists of cues that rely on using two eyes (reviewed extensively in Howard & Rogers, 1995). There are a number of ways to obtain information about distance by making use of the fact that we have two eyes, including extracting such information from the differences between the images in the two eyes (binocular disparities) and from the differences between the directions in which the two eyes are oriented (ocular convergence). People who are

unable to use binocular depth cues are often considered to be unable to see depth, but, of course, this is not true. If you are not such a person, and you close one eye, the world does not suddenly look flat. However, with one eye closed you may become less precise in judging distances, and therefore have more trouble pouring yourself a cup of tea.

The third kind of cues consists of active cues (Gibson, Gibson, Smith, & Flock, 1959; Rogers & Graham, 1979). The most prominent active cue arises when an observer moves in a static environment. When he or she does so, the directions to structures around him or her change in a manner that depends on the structures' distances. Individual changes in direction and relative direction are known as motion parallax. The combination of all such changes is known as the optic flow. When moving in a static environment, combining the changes in the directions to surrounding structures with information about how one is moving can provide information about the object's distance. Even if one does not know how much one has moved, as might be the case when looking out of a train window, one can still obtain information about the depth order. Another active cue for distance is changing the curvature of the lens

of the eye (accommodation), and detecting how changes in such curvature influences blur in the image.

PICTORIAL DEPTH CUES

Occlusion

Occlusion provides very reliable information as to which of two opaque surfaces that both partly occupy the same direction from the observer are closer (Figure 9.3A). It tells us nothing about how much closer, but the difference can be very small without any reduction in our certainty as to which is closer. Occlusion is therefore a very reliable cue, but it only provides information about which of two surfaces is closer, not how close they are. Moreover, it only provides information about surfaces that occupy positions in space that overlap in terms of their direction from us. Occlusion also only provides information about the depth order if one knows how to segment the image into objects. In Figure 9.3A the observer sees an image that would probably be interpreted as three overlapping rectangular surfaces. It could be that the central surface is a rectangle with a section removed. However, if so, one would have to assume that the
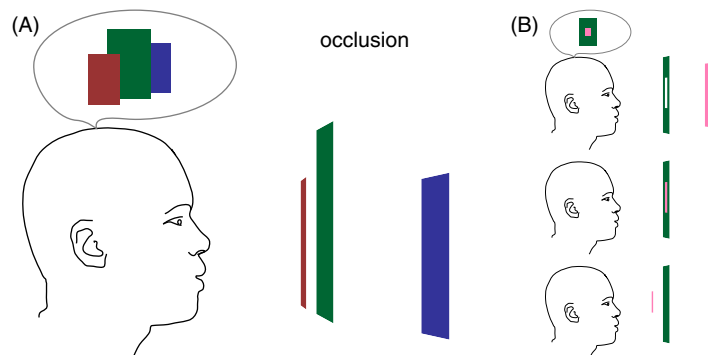
Figure 9.3   Nearby objects occlude ones that are farther away if they are in the same direction. If the image can be segmented into occluding and occluded surfaces in a straightforward manner, for instance by assuming that the surfaces have certain shapes, the depth order is evident (A). If not, even the depth order is ambiguous (B).

section that was removed from the central rectangle is exactly aligned with the edges of the leftmost rectangle. Similarly, the edges of the central and rightmost surfaces might be aligned rather than the central one occluding part of the rightmost one. If the surfaces were at different distances these alignments would only hold for a specially selected viewpoint, so it would be very unlikely. Thus, the interpretation in terms of overlapping rectangles is reasonable. It is less evident how the image observed in Figure 3B should be interpreted. The small rectangle within the larger one could be a distant surface seen through a hole in the nearer surface, a part of the same surface that was painted a different color, or a small rectangle in front of the larger one. Thus, in this case, although occlusion still tells you that the small rectangle is the nearest surface in that particular direction, occlusion does not tell you anything about even the depth order with respect to the large rectangle.

Height in the Visual Field

In general, there is a correlation between an object's height in the visual field and its distance. Looking downward we usually see things that are close to us, whereas looking up toward the horizon we usually see things that are farther away. This is because most things in our environment rest on surfaces. The relationship between depth order and height in the visual field is quite straightforward for small objects on a single horizontal surface. It is not as straightforward for large objects or when objects are not resting on the same horizontal surface or are not resting on surfaces at all. For large objects, it is important to realize that it is not the height of the center of the object that is relevant, but the height of its base, where the object makes contact with the surface, because it is the position on the surface that matters. Thus, for

Figure 9.4   Trees are taller than flowers, so they will often occupy higher positions in our visual field, as they do in this photograph. However, despite the indicated tree extending much higher in the image, the indicated tree and the indicated flower are judged to be at about the same distance because they rise from the ground at about the same height in the image.

instance, a tree and a flower next to the tree can be considered to have the same height in the visual field for the purpose of judging distance (Figure 9.4). Unlike for occlusion, the objects do not need to overlap in terms of their direction from the eyes for their depth order to be determined.

There is ample evidence that people make use of height in the visual field to judge objects' distances (Ooi, Wu, & He, 2001). When considering patterns on a surface, or when considering where objects make contact with a surface, the height in the visual field provides direct information about the depth order. If you can be certain that the surface is horizontal, and you know your eye height with respect to this surface, height in the visual field can provide estimates of
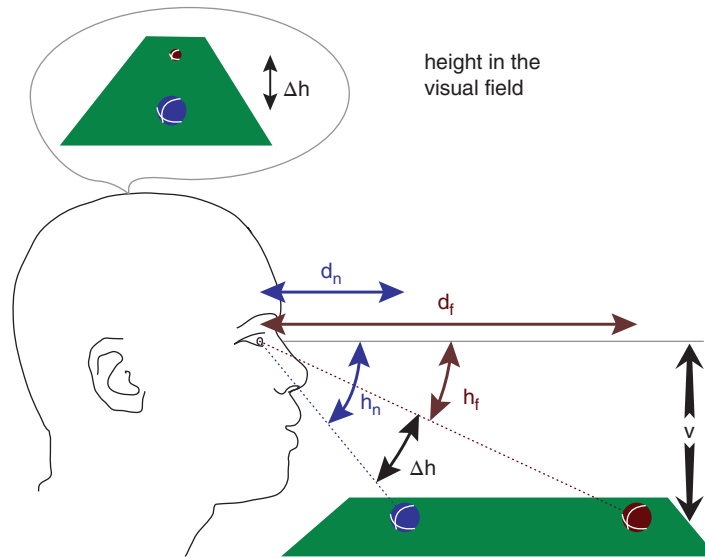
**Figure 9.5**   If the surface is horizontal, your own eye height is v, and your gaze angle with respect to the horizontal is h, the distance to a small object on the surface is $d = v/\tan(h)$. If there are two objects on the surface, and the distance to the farther object is known (d_f), the distance to the nearer object (d_n) can be judged from the difference between their heights in the visual field (Δh): $d_n = v/\tan\left(\Delta h + \text{atan}\,\frac{v}{d_f}\right)$. If the farther object is very far away ($d_f \approx \infty$; $h_f \approx 0$; $\Delta h \approx h_n$), doing so is equivalent to relying on the gaze angle with respect to the horizontal: $d_n = v/\tan(h_n)$.

the actual distances of objects located on the surface (Figure 9.5). This might for instance be the case for a person standing in an office. He or she can be assumed to be familiar with his or her eye height (for support for the idea of relying on eye height see Bridgeman & Cooke, 2015; Daum & Hecht, 2009). Moreover, the floor of the room can be assumed to be horizontal. Besides knowing his or her own eye height, the person would have to be able to judge the vertical position of the horizon. Judging this visually could be a problem in an enclosed space, in which case one might have to rely on a vestibular estimate of the horizontal eye level (Li, Dallal, & Matin, 2001), or on visual estimates based on objects in the scene (such as other people's eye heights) or on the optic flow if one is moving (the vertical position of the focus of expansion if one is moving through the office).

Considering a retinal resolution of about 1 minute of arc (1/60th of a degree), the resolution of judging two structures' depth order from their heights in the visual field is very good. It varies with distance, with separations of less than 1 mm being discernable for objects about 1 m away, and separations of about 1 cm being discernable for objects that are several meters away (*horizontal surface* curves in Figure 9.6). Of course, this is an estimate of the resolution for detecting that there is a depth difference between the objects. It is difficult to estimate the resolution for judging the actual distance of either of the objects, or of the separation between them, because this requires knowledge of the gaze angle or of the visual angle with respect to the horizontal. Although the former may or may not depend on the magnitude of the angle, the latter probably increases with the vertical separation in the image, possibly
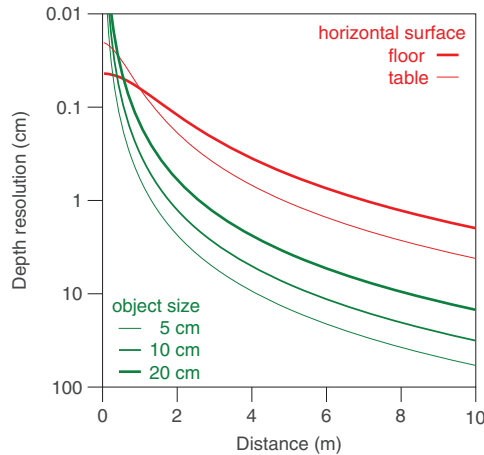
Figure 9.6 Resolution of *height in the visual field* and *image size* as depth cues. The figure shows how the resolution for detecting a separation in depth depends on the distance. The resolution is based on a threshold for detecting retinal separations of 1′ arc. The separation corresponding with this threshold is presented on an inverted logarithmic scale so that a high resolution is high in the figure. Red curves: differences in distance corresponding with a vertical retinal separation of 1′ arc for a horizontal surface 1.5 m below eye height (floor) or 70 cm below eye height (table). Green curves: differences in distance corresponding with a 1′ arc difference in image size for objects of three different sizes (5, 10, and 20 cm).

counteracting the increase in resolution for nearby objects to some extent. In either case, it is evident that the resolution for judging an actual distance is lower than what is shown in Figure 9.6.

If objects are not resting on the same surface, one can follow how objects rest on each other to use height in the visual field to judge their depth order to some extent (Figure 9.7; Meng & Sedgwick, 2001). A somewhat related way to judge objects' relationships with a ground plane when it is not evident that the objects are lying on the ground plane is to consider cast shadows. An object's shadow can give an indication of its distance

from a surface: A shadow close to the object suggests that the object is close to the surface, whereas a larger separation suggests a larger distance from the surface. A change in the assumed distance to the surface will modify the relation between height in the visual field and perceived distance: If an object is far above a surface, it will be perceived to be nearer than a similar object for which the shadow suggests that it is lying on the same surface (Figure 9.7; Allen, 1999; Kersten, Mamassian, & Knill, 1997).

What if the surface is not horizontal or not flat? Height in the visual field obviously becomes much less reliable if the surface is tilted to the side and the objects of interest are separated laterally (i.e., in azimuth). For surfaces that are slanted upward or downward, height in the visual field still provides reliable information about the depth order. For large surfaces with small slants, even judgments of distances and depths could be quite reliable as long as the judgments are based on the vertical separation with respect to the visible horizon rather than on the gaze angle with respect to the horizontal (for some indication that this is what people use see Gardner, Austerweil, & Palmer, 2010). The error that arises from the distance of the observer's eyes from the surface no longer being the height in the direction of gravity is negligible for modest slopes. Judgments based on gaze angle relative to gravity would obviously provide quite wrong estimates of the distance: When walking up a slope one would overestimate the distance and when walking down a slope one would underestimate the distance. Even relying on height in the visual field to judge depth order is obviously unreliable if the objects are on a surface that is not flat, if they are on different surfaces that are not connected by simple visible supports, if the positions at which the objects make contact with the surface are not visible, or if the objects are not on surfaces at all.

support and
shadow
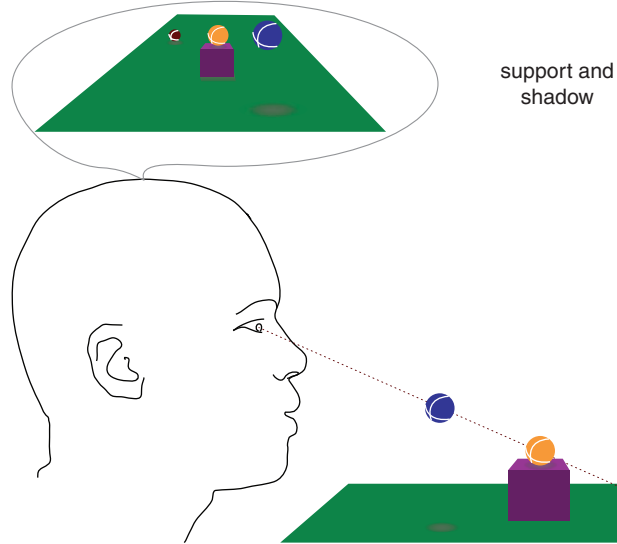
Figure 9.7   Objects' shadows and the way they rest on other objects can help determine how to interpret the height in the visual field. Whereas the red ball's shadow indicates that it is resting on the green surface, the blue ball's shadow indicates that it is hovering in the air above the green surface, and is therefore nearer than its height in the visual field might suggest. Similarly, the fact that the yellow ball is evidently resting on the purple cube suggests that its distance corresponds with the height in the visual field of the center of the cube's bottom surface, rather than with the height of the ball itself, making it appear to be closer than the red ball although they are at the same height in the visual field.
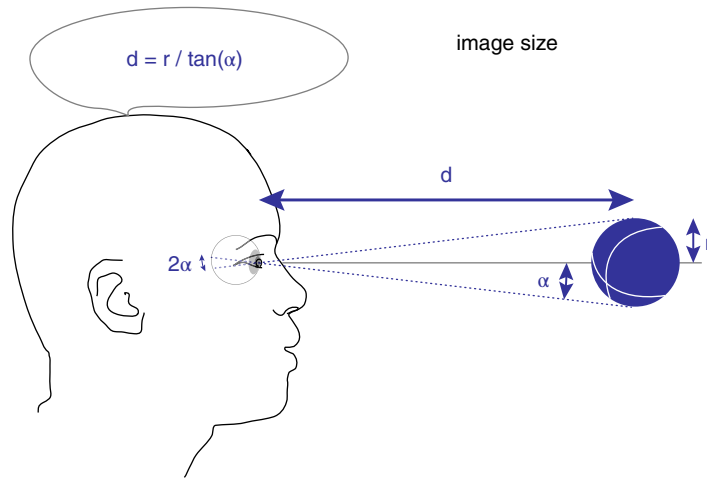
Image Size

If an observer is certain about an object's size, its retina image size can reveal its distance (Figure 9.8). Indeed, image size is used in the perception of distance (Gillam, 1995; McIntosh & Lashley, 2008). Although one might expect image size to only influence judgments of distance when the true object size is known, this is not the case. When judging the distance of an unknown object, people judge a large object to be closer than a smaller object that is presented at the same location (Collett, Schwarz, & Sobel, 1991; Lugtigheid & Welchman, 2010; Sousa, Brenner, & Smeets, 2011, 2012). People apparently use image size as a cue for distance even when they have no direct information about the actual object size. An explanation could be that people consider certain sizes to be more likely than others,

probably based on experience with similar-looking objects. The resolution for judgments of distance from retinal image size, assuming that the true object size is known extremely precisely, and again given a retinal resolution of 1′ arc, is shown by the *object size* curves in Figure 9.6. The resolution depends on the object's size and decreases rapidly with distance.

Texture

Besides considering the sizes of individual objects' retinal images, an obvious cue for determining relative distances is to consider the gradient in the sizes of similar objects' retinal images, such as the changes in the image sizes of stones or tiles across a surface that extends in depth. In this case the assumption is not that you know the actual size, but that the size is constant (isotropic) across

Figure 9.8 If an object is far away, its image on the retina is smaller than if it is nearby. An object's retinal image size (represented here by the diameter of the ball's image, 2α) is determined by the ratio between the size of the ball (represented by its radius, r) and its distance (d). For any retinal image size, knowing the true object size could tell you its distance.
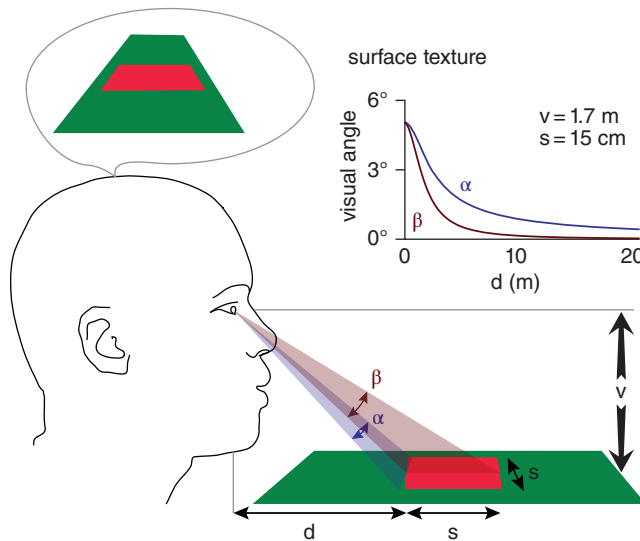
Figure 9.9 For surface texture such as floor tiles, the image size changes with the distance from the observer (d). The graph shows how the lateral angle (α) and the angle along the depth direction (β) depend on the distance when a 15 cm tile is examined from an eye-height of 1.7 m. The two curves show $\alpha = 2\tan^{-1}\left(\frac{\frac{1}{2}s}{\sqrt{d^2+v^2}}\right)$ and $\beta = \tan^{-1}\left(\frac{d+s}{v}\right) - \tan^{-1}\left(\frac{d}{v}\right)$.

space. Along a surface there are various components to the change in image size with distance. The angle filled by structures that are oriented orthogonal to the line of sight changes almost linearly with the inverse of the distance (angle α in Figure 9.9). What is true for such image sizes is also true for the density of regularly or randomly distributed texture elements: the density of the texture in the retinal image increases in accordance

with the decreases in single objects' sizes. If the actual texture elements are all identical or if their sizes vary at random across the surface, the distribution of the texture elements' retinal image sizes provides equivalent information. Thus, the depth order and even some indication of relative depths along the surface could be judged from texture gradients in the retinal image. However, the resolution for detecting a difference in distance on the basis of the local texture alone is quite poor, irrespective of the distance (see *lateral texture density* curves in Figure 9.10).

For structures that recede in depth along a ground surface, the angle between the surface
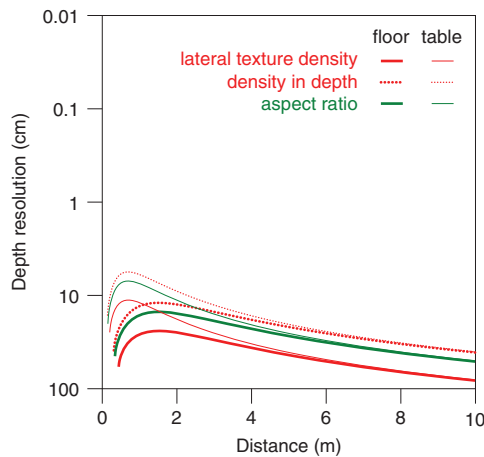
Figure 9.10 Changes in texture density and aspect ratio with distance. Red curves: differences in distance corresponding with a change in texture density of 8% (lowest value reported in Anobile, Cicchini, & Burr, 2014). Solid: lateral density (horizontal on the retina). Dashed: density in depth (vertical on the retina). Green curves: differences in distance corresponding with a 5% change in aspect ratio (Nachmias, 2008). Both measures are shown for a surface 1.5 m below eye height (floor) and 70 cm below eye height (table). Other details as in Figure 9.6.

and the line of sight depends on the distance, so the structure's angular extent changes in a more complicated manner with distance (angle β in Figure 9.9). A comparison of the horizontal and vertical extents of objects' retinal images can provide information about the objects' slants if the objects' shapes are known. Again, what is true for the image size of a regularly shaped object is also true for the density and sizes of regularly or randomly distributed texture elements. For judging slant, one could also rely on other gradients than the density or size of texture elements in the retinal image. For instance, if the texture consists of oriented elements, one could rely on gradients in the distribution of orientations in the retinal image (Warren & Mamassian, 2010). For judging slant it is important to realize that for texture that is not flat on the surface, such as pebbles, matters may be more complicated than we have sketched above, because the slant will also determine to what extent parts of the surface, such as pebbles of many sizes, occlude each other. Slant on its own does not provide any information about distance or depth, but knowing the slant could be important for interpreting other cues such as height in the visual field or texture density. Moreover, if one is certain about the true slant, one could use the gradient along the ground surface (*density in depth* curves in Figure 9.10) or the difference between the gradients in the two directions (*aspect ratio* curves in Figure 9.10) to determine the depth order. Although the latter measures provide a slightly better resolution than relying on gradients orthogonal to the line of sight, they also depend on more assumptions, so it is not evident that any of these measures could play a major direct role in judging distances or depths.

The use of texture gradients relies on assumptions about the elements on the surface, such as that the elements are identical

or similar, and that they are regularly or randomly distributed. Any systematic ordering could be misinterpreted as depth. An extreme case of assuming a certain ordering is when lines converge toward a single point. This is usually interpreted by the visual system as the lines being parallel but receding in depth, so that the constant separation between them leads to a smaller separation in the image as the distance increases (see top left of Figure 9.9). Similarly, horizontally elongated ellipses are readily seen as slanted circles. Such cues can be very strong, especially when judging slant (Muller et al., 2009). As surface slant needs to be considered when height in the visual field is interpreted in terms of distance, the most important influence of texture cues on judged distance may be mediated by estimates of slant in combination with height in the visual field, rather than through direct judgments of distance. This illustrates that the depth cues that we have at our disposal might not be independent. We return to this issue in the section about combining cues.

Image Quality

Image quality can be informative about distance. Light reflected by objects is slightly diffused by particulates in the air on its way to our eyes. Consequently, contrast decreases with distance. Reduced contrast is therefore indicative of a large distance (*aerial perspective*). To use this cue to estimate distance, one must make assumptions about the local atmosphere and about the objects in question. Except under very foggy or rainy conditions, the changes in contrast with distance are so small that this cue is only effective for detecting large separations in depth. This cue can therefore be useful for judging which of two distant buildings is farther away, but it will seldom be useful for nearby objects. Its resolution is obviously very poor.

At small distances, blur can provide some information about distance. If you are looking at a surface at a certain distance, and accommodation is adjusted to that viewing distance, the images of things (edges or changes in surface reflectance) that are at that distance will be sharp, whereas the images of things at other distances will be blurred. Consequently, if an edge between the object that you are looking at and another object is sharp, the edge belongs to the object that you are looking at, so this object is probably in front of the other object. On the other hand, if the edge is blurred, the edge belongs to the other object, so the other object is probably occluding the object that you are looking at (Marshall, Burbeck, Ariely, Rolland, & Martin, 1996). This use of blur to judge the depth order assumes that the border of the object itself is sharp.

Both contrast and blur appear to contribute to judgments of distance (Held, Cooper, & Banks, 2012; O'Shea, Govan, & Sekuler, 1997). In order for contrast to provide information about more than the depth order one would have to consider the weather conditions. Similarly, in order for the instantaneous blur to provide information about more than the depth order one would have to consider the size of the pupil and the state of accommodation of the eye. An alternative mechanism for using blur to obtain information about distances or depths is by minimizing blur at the position of interest through accommodation, and using the required accommodation to judge the distance. People probably use this mechanism to some extent because accommodation has been shown to contribute to judgments of depth (Watt, Akeley, Ernst, & Banks, 2005), and distance judgments are slightly better when looking normally than when looking through a pinhole (in which case the image is sharp irrespective of the accommodation of the lens; Frisby, Buckley, & Horsman, 1995).

BINOCULAR DEPTH CUES

When thinking about judging distance or depth, the first cues that come to mind are usually the binocular cues. It is therefore not surprising that the many contributions of binocular vision to depth perception have been studied very extensively (for reviews see Foley, 1980; Howard & Rogers, 1995). In principle, the distance to any structure that is being fixated could be determined on the basis of the viewing directions of the two eyes, through triangulation. Before discussing how people make use of the small differences between the images in the two eyes (retinal disparities) to judge distances and depths, we therefore consider judgments of the orientations of the eyes.

Eye Orientation

The precision with which we know the orientation of each eye is obviously limited. Determining how well we normally know the orientation of our eyes is difficult, because when doing so subjects are necessarily kept in the dark to remove any other cues about distance or direction of gaze. As the eyes drift when one is placed in the dark, determining the eye orientation in this manner will lead to it appearing to be poorer than it actually is. To circumvent this, one can examine how well distances and directions can be judged when comparing them across a single saccade (Brenner & van Damme, 1998; Enright, 1991, 1996). The standard deviation in judgments of the orientation of each eye, when estimated in this manner, is slightly more than 6 minutes of arc (Brenner & Smeets, 2000). Converting this to 95% confidence intervals of the perceived location shows that fixated structures' positions, including their distances, could be judged quite reliably on the basis of estimates of the eyes' orientations for nearby structures, but that for distances that are beyond reach the judgments of distance
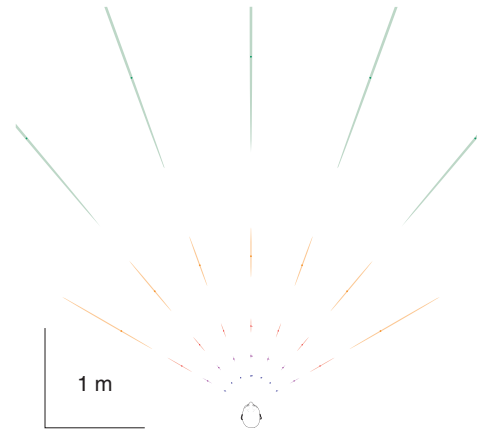
Figure 9.11 Estimated precision of judging positions on the basis of eye orientation information alone. The shaded regions are 95% confidence intervals around the positions indicated by the dots (at distances of 30, 50, 80, 150, and 350 cm from the observer). These values are based on independent errors with standard deviations of about 6 minutes of arc per eye.

become quite poor (Figure 9.11). Because these estimates are based on comparisons across a single saccade, they should probably be considered to represent the best possible precision of judging positions on the basis of eye orientation alone. The *vergence* curves in Figure 9.12 show how the resolution for judging depth order from information about the orientation of the eyes rapidly decreases with distance.

A Choice of Coordinates

To describe binocular vision conveniently, we assume that the two eyes are both oriented toward the same structure, and consider the plane including the two eyes and the structure that is being fixated as the plane within which the azimuth and distance of gaze is determined. We consider the rotation of the plane around the axis through the two eyes as the elevation of gaze. When considering two structures, if the angle at the eye between the directions to the two structures is the same
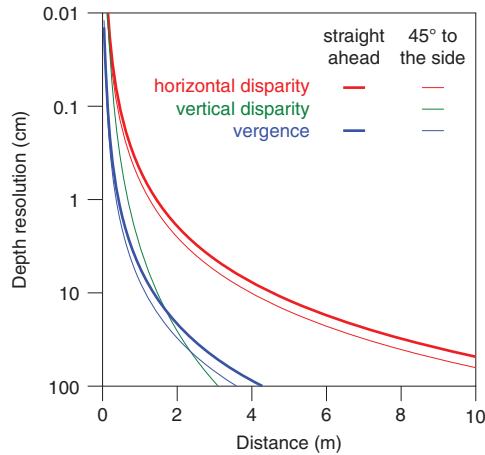
Figure 9.12   Resolution of binocular depth cues for structures that are straight in front of the observer or 45° to the side. Values are based on a 6.5 cm separation between the eyes, and either a retinal resolution of 1′ arc (horizontal and vertical disparity) or an oculomotor resolution of 12′ arc (vergence). The vertical disparity does not depend on the distance for objects that are straight ahead. Details as in Figure 9.6.

for both eyes, the structures are considered to have no relative binocular disparity. If the angles at the eye differ in magnitude along the direction of the axis through the eyes, the structures are considered to have a different *horizontal disparity*. If the angles at the eye differ in magnitude in the orthogonal direction, the structures are considered to have a different *vertical disparity*. The distinction between these directions also applies to the special case of a structure's disparity with respect to the structure that is fixated, in which case we refer to the differences between the angles as horizontal or vertical *retinal disparity*.

Horizontal Disparity

For understanding the relationship between distance and horizontal disparities, it is convenient to start with a description of locations

for which the aforementioned angles between the directions to two structures are the same for both eyes (i.e., points with no horizontal disparity). These locations fall on circles through the two eyes. One such circle is the circle for which the horizontal *retinal* disparity is zero: the circle through the two eyes and the fixation point (Figure 9.13A). It is easy to see that shifting one's gaze between structures with the same horizontal disparity, so that both eyes rotate by the same amount, will not change the angle between the lines of sights of the two eyes (known as the vergence angle; Figure 9.13B). That points for which the vergence angle is the same lie on Vieth-Müller circles is explained in Figure 9.13C. Of course, points with the same horizontal disparity lie on such circles, irrespective of the orientation of the eyes.

It is tempting to interpret Figure 9.13 as showing that structures on a Vieth-Müller circle have the same retinal eccentricities in both eyes. Such an interpretation would justify our choice of coordinate system because the retinal images are the basis of visual perception. However, the step from angles in the aforementioned coordinate system to positions on the retina assumes a certain orientation of the eyes around the line of sight (torsion). The orientation of the eyes around the line of sight is more or less fixed for each gaze direction (Donders' law), and is indeed more or less appropriate for aligning the images in the two eyes (Cooper, Burge, & Banks, 2011). Thus, the images of structures on the Vieth-Müller circle in Figure 9.13A can be considered to fall on corresponding retinal positions in the two eyes. The images of structures on other such circles, not passing through the point of regard, do not fall on corresponding retinal positions, but the retinal disparity is the same for all such structures because they have no relative disparity with respect to each other.

Horizontal disparity is an important source of depth information for everyone with
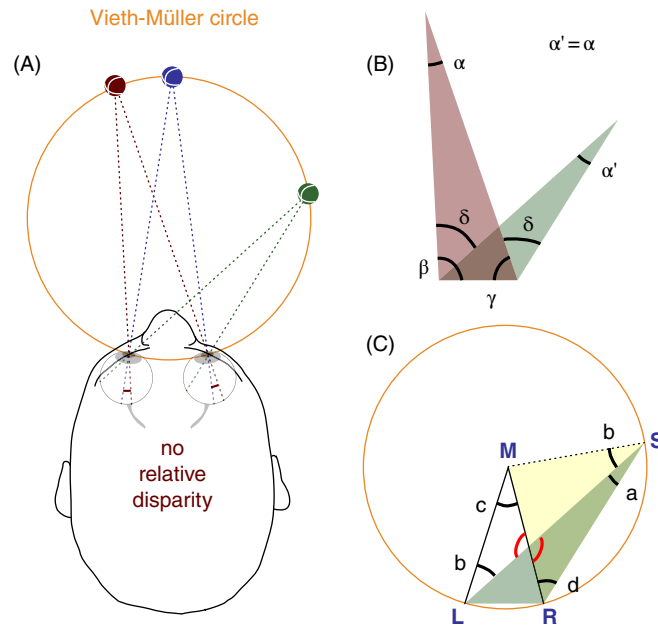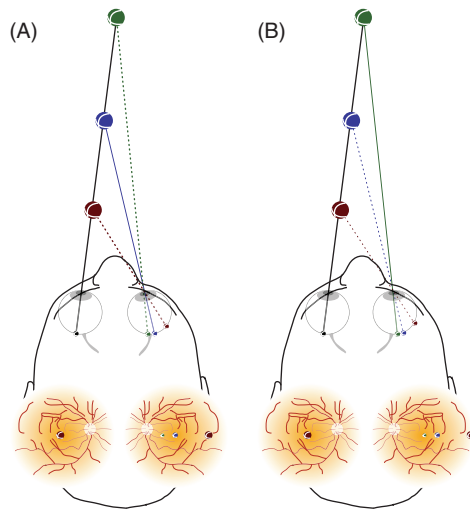
Figure 9.13   The Vieth-Müller circle: a circle through both eyes (A). If the angle between two structures is the same for both eyes (δ in B), the angle that the lines of sight would make if one were to fixate either of the two structures is also the same: For both triangles the angles sum to π, so $\alpha + \beta + \gamma = \pi$ (red triangle) and $\alpha' + (\beta - \delta) + (\gamma + \delta) = \pi$ (green triangle), which means that $\alpha = \alpha'$. To see that the points for which this holds fall on the Vieth-Müller circle consider any position on that circle (e.g., that of the structure S in C). Because the two angles indicated in red are equal, and the sum of the angles in a triangle is π, b + c = a + d. The lines MR and MS have the same length (the radius of the circle), so the angles that they make with the line RS are also equal d = a + b (yellowisosceles triangle RSM). Substituting a + b for d in b + c = a + d we can see that c = 2a, showing that for any position S on the circle the angle a will be the same.

normal binocular vision, but just knowing the disparity is not enough to determine distances or depths, because the horizontal retinal disparity only tells you the distance if you know the fixation distance. Differences between the retinal positions of the images of a structure in the two eyes provide information about the structure's relative distance (with respect to the structure that is fixated), but the magnitude of the difference that corresponds with a given difference in retinal disparity increases with the fixation distance (Figure 9.14). The sign of the difference globally indicates whether the structure is nearer or farther than the structure that is fixated, but the relationship between horizontal retinal disparities and positions in space is not simple. The same is true for relative disparities between two structures. First, positions with the same horizontal disparity are on circles through the eyes, so they are not at the same distance from the observer; neither in terms of radial distance (as defined in Figure 9.1) nor in terms of a Cartesian distance (forward-backward, as opposed to left-right and up-down). Secondly, equal changes in disparity do not correspond with equal separations in distance (see the 1° differences between consecutive circles in Figure 9.15). The same horizontal disparity corresponds with a much larger distance when fixating a more distant structure.

Figure 9.15 Circles of positions with the same horizontal disparity, in steps of 1° relative to positions at the horizon (distance = ∞). Note that the change in disparity with distance is approximately inversely proportional to the distance, as can be inferred from the fact that c = 2a in Figure 9.13C.

Figure 9.14 Retinal disparity. Example showing different positions of the retinal images in the right eye for objects that are aligned for the left eye, when fixating either the blue (A) or the green (B) ball. The fixated ball obviously has zero retinal disparity. The retinal separation for an equivalent separation in space decreases with distance. The retinal positions alone provide no information about distance because they depend on which object is fixated.

Because the angles between structures do not change when we rotate our eyes, but positions of the objects' retinal images do change (Figure 9.14), relying on relative disparities between visible structures rather than on each structure's retinal disparity to judge separations in depth means that we do not have to worry about the orientations of the eyes. Not using gaze as a reference means that the resolution is determined by the retinal resolution at two structures' images' positions rather than by the resolution of judging the vergence angle of the eyes and the retinal resolution at a single structure's images' positions. In order to accurately judge the relative disparity the eyes must be directed at about the same distance as the structures of interest, because if the retinal disparity is too large the two retinal images will not be attributed to the same structure (resulting in

double vision: diplopia). Moreover, although knowing the relative disparity provides information about the depth order, in order to interpret a horizontal relative disparity in terms of an actual depth one must also know the overall distance. For retinal disparities it would be logical to obtain such knowledge by judging where one is fixating, both in terms of distance and lateral position. Considering that we usually direct our head more or less toward where we are looking, one can see from Figure 9.15 that the adjustments to the lateral position are not very critical: the distance from the head does not change much with small angles from straight ahead when following Vieth-Müller circles rather than circles centred on a point between the eyes (the origin of our measure of distance). However, the changes in distance between the circles with identical changes in disparity are very different at different distances, so knowing the relative disparity between two structures only provides reliable information about their separation in depth if one knows the distance of one of them.

Relative binocular disparity is the most sensitive source of information about distance, with some people being able to reliably detect depth differences of as little as 5 seconds of arc for foveal targets under certain conditions (McKee, 1983). The precision with which people can judge relative disparity obviously decreases with distance from the fovea, both in depth and in the frontal plane (Schor & Badcock, 1985; Siderov & Harwerth, 1995; Siderov, Harwerth, & Bedell, 1999), but the decline in precision is not dramatic. The precision is hardly poorer if the target is moving (Ramamurthy, Bedell, & Patel, 2005; Westheimer & McKee, 1975, 1978). That it is the relative disparity that is critical is evident from studies showing how disparity is judged with respect to a slanted plane (Glennerster, McKee, & Birch, 2002; Mitchison & Westheimer, 1984). Comparing relative disparities is therefore very precise, but interpreting relative disparities in terms of actual distances is complicated.

Vertical Disparity

Until now, we have only considered binocular separations between objects that are on the plane through the eyes and the fixation point. The elevation of this plane is irrelevant for the cues that we are interested in, so looking up or down will not change anything. However, a large part of the image on the retina is obviously not concerned with objects that lie within this plane. In terms of horizontal disparity this does not matter (if our assumptions about the relevant coordinate system are correct), but for structures that are not on this plane, the angle with respect to the plane (and therefore the vertical retinal eccentricity of the images in the eyes) can differ between the eyes. This is easiest to understand by considering the difference in size between the images in the two eyes (Gillam & Lawergren, 1983): if a vertical rod's distance differs for the two eyes, the vertical image sizes will differ in accordance with the differences in distance. The sign of the difference depends on the position: if the rod is to the left, its image in the left eye will be larger than that in the right eye. A difference in vertical image position is referred to as *vertical disparity*. From the aforementioned, it should be clear that the vertical disparity increases with the azimuth and decreases with the distance (with respect to the head). Considering that the same uncertainty about the eyes' orientations leads to a much larger uncertainty about the distance than about the lateral position (Figure 9.11), vertical disparities might be useful for judging the viewing distance even though their dependence on both distance and azimuth means that this requires an estimate of the direction of gaze. Moreover, knowing the direction could be circumvented by relying on the gradient of vertical disparity throughout the fusible part of the scene (Brenner, Smeets, & Landy, 2001).

The curves in Figure 9.16 show the azimuths and distances at which structures that are at three vertical retinal eccentricities have three different values of vertical disparity. Of course, if there is vertical disparity, the vertical retinal eccentricities are not the same in both eyes. We therefore consider the mean vertical retinal eccentricities of the two eyes to be the overall vertical retinal eccentricity. For nearby targets at large vertical eccentricities, vertical disparity can be quite large. The vertical disparity depends on the azimuth with respect to the head, the vertical retinal eccentricity, and the distance. Thus, if you know the direction of gaze, the vertical disparity at a given vertical retinal eccentricity could provide information about the distance (as long as the structure of interest is not straight in front of you). However, vertical disparities are quite small, unless the structure of interest is extremely nearby, so their resolution (see Figure 9.12) is probably
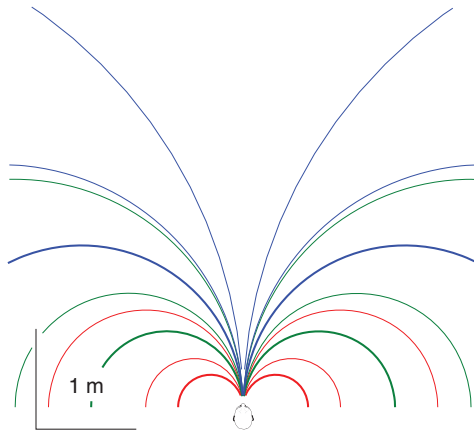
Figure 9.16 Vertical disparity. The curves show positions for which the difference in vertical retinal eccentricity (vertical disparity) is 10, 20, or 30 minutes of arc (indicated by increasing line width), while the mean vertical eccentricity is 5, 12, or 30°(red, green, and blue, respectively). On the left, the eccentricity with respect to the left eye is larger. On the right, the eccentricity with respect to the right eye is larger. Eccentricity is expressed as the vertical angle at the eye with respect to the plane through the eyes and the fixation point.
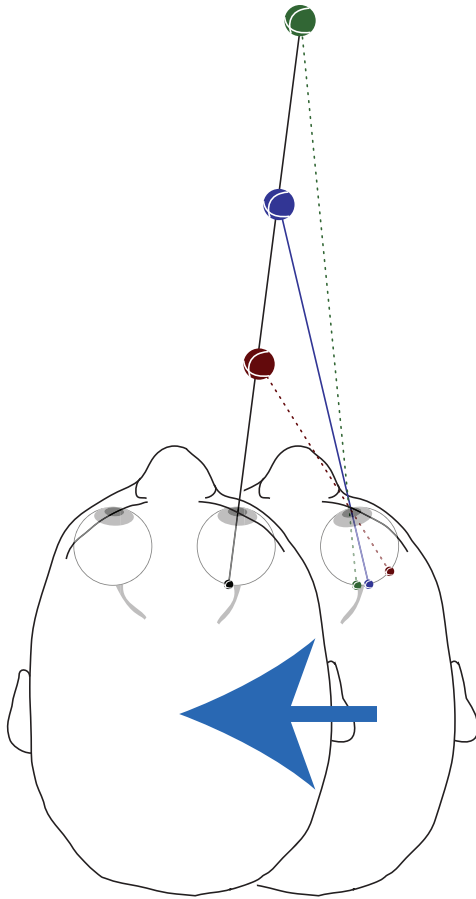
normally insufficient for judging the distance of individual objects. The vertical disparity increases with increasing vertical retinal eccentricity, so vertical disparities are larger at large retinal eccentricities, but this advantage is probably alleviated by the decrease in resolution with retinal eccentricity. Nevertheless, despite vertical disparities therefore probably not being useful for directly judging an object's distance (Cumming, Johnston, & Parker, 1991; Sobel & Collett, 1991), the overall pattern of vertical disparities might be used to obtain estimates of the viewing distance with which to judge distances or directly scale horizontal disparities (Adams et al., 1996; Backus, Banks, van Ee, & Crowell, 1999; Brenner et al., 2001; Duke, Oruç, Qi, Backus, 2006).

ACTIVE DEPTH CUES

We already mentioned that people might actively accommodate to remove blur, and use the amount of accommodation to judge the distance (Watt et al., 2005). A more important active depth cue is *motion parallax*. Motion parallax is similar to binocular disparity in that it is based on having different views of the same scene (Figure 9.17). In motion parallax the different views are obtained at different moments. As a result, interpreting the changing image of the scene as being caused by a change of viewpoint, and using this to derive structures' distances, is only straightforward if the scene is stationary and one knows one's own movement quite reliably. In binocular vision, the scene being stationary is not an issue as the two views are obtained simultaneously. Moreover, the distance between the eyes is fixed so we can consider it to be known with high accuracy. For motion parallax, if the scene is indeed stationary, the resolution for detecting a difference in depth, or for detecting the depth order, might depend on the extent of self-motion (Figure 9.18), although the velocity of self-motion is probably also important.

In order to interpret motion parallax in terms of actual distances or depths one has to also judge one's own movement or scale the motion information in some other manner. Thus, motion parallax requires scaling of the retinal motion just as binocular vision requires scaling of the horizontal disparity. There are more similarities between motion parallax and horizontal disparity. For instance, in both cases differences in depth could be judged from the directions with respect to the eye(s), but they could also be judged from changes in relative positions within the retinal image(s) of the scene: the three objects being aligned for the left eye but not for the right eye in Figure 9.14 correspond with the three objects being aligned

Figure 9.18   Resolution of motion parallax as a depth cue for various extents of lateral motion of the head. Values based on a retinal resolution of 1′ arc, assuming that the scene is static. Details as in Figure 9.6.

Figure 9.17   The analogy between motion parallax and horizontal disparity. Both rely on differences in the direction to the structure of interest from different positions. For horizontal binocular disparity the different positions are the positions of the two eyes. For motion parallax the different positions are positions of the same eye at different moments. In this example, the observer aligns the three objects with respect to the right eye by moving to the left, in analogy with the alignment in the left eye in Figure 9.14. The farther the object from the eye, the smaller the change in angle with respect to the eye when one moves. This is evident when looking out of a train window: Nearby objects pass quickly, while distant objects pass slowly.

after but not before the head moved to the left in Figure 9.17. Despite the similarities, there are also some fundamental differences,

which is probably why performance is not identical when based on matched versions of the two cues (Bradshaw, Hibbard, Parton, Rose, Langley, 2006).

When discussing motion parallax the emphasis is often on modest lateral self-motion (i.e., self-motion in a direction orthogonal to the distance). However, one can also obtain information about an object's distance from changes in its image size as one moves toward it (Peh, Panerai, Droulez, Cornilleau-Pérès, & Cheong, 2002). For extensive self-motion one must obviously consider that most structures' distances will constantly be changing. Although the observer's movements do not influence the actual separations between static objects, or their sizes, it does influence the extent to which the separations between structures are in depth, so when it is important to isolate the depth component, such as when judging a surface's slant, one also has to consider the continuous changes. If it is safe to assume that the whole scene is static, it is theoretically possible to judge the instantaneous relative

depths within the whole scene from the optic flow (Gibson, 1979; Koenderink, 1986). The depths could be scaled by information about one's own motion, or by any known distance, to obtain judgments of the actual distances and depths. Such interpretation of the optic flow is presumably responsible for some of the ways we consider distances during everyday tasks such as locomotion (Duchon & Warren, 2002).

To know whether it is safe to interpret specific image motion in terms of distance, one must verify that the image motion is not the result of the object in question itself moving relative to other objects in the scene. Due to the regularities in the retinal image motion that is caused by self-motion, it is usually possible to reliably determine the direction of self-motion (van den Berg, 1992; Warren & Hannon, 1988) and separate the influences of self-motion from ones of object motion (Brenner & van den Berg, 1996; Warren & Rushton, 2008, 2009). When there is limited visual information, not all motion parallax is interpreted as depth. In an extensive series

of experiments, Gogel and colleagues have shown that a static object appears to move in response to lateral self-motion when its distance is misjudged (Gogel, 1990; Gogel & Tietz, 1973), rather than the judgment of distance being adjusted to conform to the object being static. This is not the case when there is more visual information (Glennerster, Tcheang, Gilson, Fitzgibbon, & Parker, 2006), so apparently people only assume that the scene is static if there is support for this from within the image.

COMBINING DEPTH CUES

Figure 9.19 looks strange because the banana must be closer to us than the apple, because it occludes part of the apple, but a number of cues suggest that it is not. On the left side of the picture, the image of the banana is a bit small in relation to the apple. Of course, any object could give rise to an image of any size, because image size scales with the object's distance (the larger the distance,

Figure 9.19   Why does this picture look strange? The banana occludes the apple, but its position on the table indicates that it is behind the apple. The image sizes also suggest that the banana is slightly farther away. The shadows confirm that the apple and the right side of the banana are resting on the table.

the smaller the image size), but the fact that the banana occludes the apple constrains the possible distances. This must therefore be an exceptionally large apple, or an exceptionally small banana. Another cue that suggests that the banana is farther away than the apple is that its base is higher in the visual field. That could just mean that the banana is suspended in mid-air, but besides that being unlikely for a banana, the banana's shadow on the right side of the picture confirms that it is lying on the table. The conflict between the cues makes the picture look strange.

Until now, we have considered all the different depth cues in isolation. We saw that only knowing the orientations of the eyes could directly provide an estimate of the distance, and this cue's resolution is quite poor, except perhaps at very short distances. Most other cues can provide information about the depth order, but require scaling to provide information about actual distances or depths. Most of them are also based on assumptions that may or may not be correct. Some can only be used in certain circumstances (if the surface is textured; if one is free to move). In all cases the resolution varies with distance, usually decreasing monotonically as the distance increases. Since the assumptions underlying different cues are not the same, and the required scaling and change in sensitivity with distance and other parameters are also different for different cues, it should not surprise us to see conflicts between the estimates of distance or depth provided by different cues, although the conflicts will normally not be as evident as in Figure 9.19.

The abundance of depth cues means that either one cue has to be selected, or they have to be combined in some manner (Cutting & Vishton, 1995). Because some cues provide information faster than others (van Mierlo, Louw, Smeets, & Brenner, 2009), one may even have to switch between cues or adjust the way they are combined as time passes.
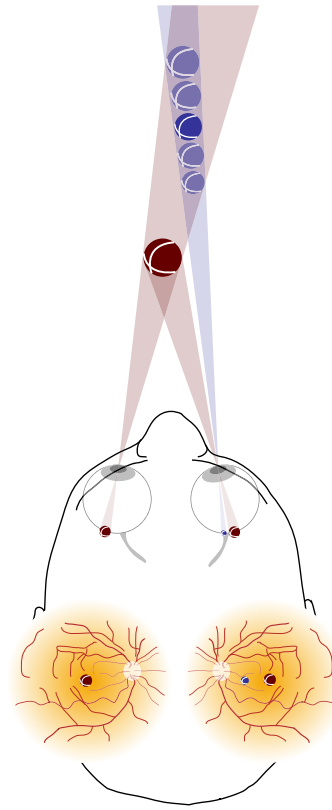
Figure 9.20 The fact that the blue ball is only visible to the right eye indicates that it must be within the range indicated by the five blue balls.

In a few cases, combining information from what we have been considering different cues could even provide additional information. For instance, the fact that a visible structure occludes the image of a second structure in one eye limits the possible positions of the partly occluded structure in terms of relative disparities (Figure 9.20; Harris & Wilcox, 2009). The most obvious example of combining cues is combining the orientation of the eyes (ocular convergence) with horizontal disparities to judge depths. Another straightforward example is directly combining horizontal and vertical disparities (Read, 2010; Rogers & Bradshaw, 1995), for instance to determine the most likely

structure and distance given the combination of disparities (Bülthoff, 1991).

It seems obvious that one must somehow consider how reliable the diverse distance cues are when combining or selecting between them. One must also consider what one is trying to judge. For instance, occlusion can provide very reliable information about which of two objects is closer than the other, but not about how much closer it is. Similarly, some cues may be more suitable for judging distances, whereas others may be more suitable for judging depths. Which cues are most suitable also depends on the circumstances. In all cases, it would make sense to combine the cues to obtain the most likely value of the judgment of interest, rather than only relying on the "best" cue. A relatively simple way to achieve this is by averaging them in the way that maximizes the overall precision (Ernst & Banks, 2002; Hillis, Watt, Landy, & Banks, 2004; Jacobs, 1999; Muller et al., 2008; van Beers, Sittig, & Denier van der Gon, 1999). If the cues all provide estimates with independent normally distributed precisions (which is not necessarily always the case; Bradshaw & Rogers, 1996; Oruç, Maloney, & Landy, 2003), the combination that gives the best overall estimate of distance is a weighted average, where the weights are inversely proportional to the cues' precisions. Presumably, the estimates are converted into common units before being combined in this manner (Landy, Maloney, Johnston, & Young, 1995). This kind of weighted averaging is often referred to as optimal cue combination.

Optimizing precision is only the best way to combine cues if the differences between the estimates are really due to measurement errors. For depth perception, most cues are based on assumptions, so one or more of the assumptions being violated could also cause discrepancies between the cues. If people consider the likelihood of assumptions being violated, they should reduce the weight given

to a cue when faced with evidence that an assumption that is required for using that cue is probably not justified (Knill, 2007; Mamassian & Landy, 2001; Muller et al., 2009). In general, large cue conflicts could indicate that an assumption must be violated. Binocular cues do not depend on assumptions that can be violated (although one may have failed to correctly match the corresponding structures in the two eyes). Pictorial cues do rely on assumptions that can be violated. Nevertheless, when judging slant, people did not increase the weight given to binocular cues with respect to pictorial cues as the cue conflict increased, except when the cue conflict was extremely large (van Ee, Adams, & Mamassian, 2003; van Ee, van Dam, & Erkelens, 2002). People did rely less on retinal image size for judging distance when object size varied more on previous trials (Seydell, Knill, & Trommershäuser, 2010; Sousa et al., 2013), indicating that in some cases people do consider whether the assumption underlying the use of the cue in question is likely to be correct (here, that the approximate object size is known). That there is some flexibility in assigning weights to various cues is evident from a study in which feedback was provided about the accuracy of the judgment. In that case, more weight was given to the more reliable cue than it would get based on its precision alone (van Beers et al., 2011).

One may be surprised to notice when comparing the resolution of the cues described in Figures 9.7, 9.10, 9.12, and 9.18 that height in the visual field is often the cue with the highest resolution. In particular, one may be surprised that its resolution is better than that of binocular vision. The reported resolution for height in the visual field is the resolution for determining the depth order of structures that are both directly on a horizontal surface. Binocular vision is more flexible in terms of the layout of the items in the scene. The main

reason for height in the visual field having a higher resolution is that it is based on vertical retinal separations in each eye rather than on differences between the horizontal separations in the two eyes (or possibly in the elevation of gaze rather than the convergence of the eyes). We recently found (Brenner, Driesen, & Smeets, 2014) that an equivalent difference between the resolution of changing elevation (whereby the *surface* is vertical rather than horizontal) and changing binocular information results in hitting a falling ball primarily being determined by the changing elevation, rather than by the changing binocular information or changing image size (that potentially provides direct information about the time to contact; Tresilian, 1993). This suggests that in daily life we may often also rely quite strongly on height in the visual field for judging distance.

CONSISTENCY

If people optimize the way they combine the available cues to obtain the best possible estimate of the attribute of interest, both the cues and their weights will differ for different judgments. This could give rise to inconsistencies between judgments. Besides interpreting the available information differently, people may even gather information differently when making different judgments by scanning the scene differently with their eyes. Thus, we should not be too surprised by inconsistencies between errors in, for instance, judging size and distance (Kilpatrick & Ittelson, 1953). Such inconsistencies are very clear in studies of perceived motion in depth, where the perceived displacement can be quite inconsistent with the perceived speed (Brenner, van den Berg, & van Damme, 1996). It should be noted that if inconsistencies arise from combining cues in different ways for different judgments, there should still be a reasonable correlation between such judgments. Indeed, such

correlations have been found, even when the judgments themselves are far from veridical (Brenner & van Damme, 1999).

People make systematic errors when asked to compare distances in depth with lateral or frontal separations (Kudoh, 2005; Loomis, Da Silva, Fujita, & Fukusima, 1992) and when performing exocentric pointing tasks (Cuijpers, Kappers, & Koenderink, 2000; Kelly, Loomis, & Beall, 2004). Such inconsistencies between judgments from different positions and in different directions have been contrasted with the ability to reliably walk to previously seen targets with one's eyes closed (Kudoh, 2005), and to the path that one takes when walking to a previously seen target with one's eyes closed not influencing where one ends up, even under conditions in which one does make considerable errors (Philbeck, Loomis, & Beall, 1997). Combining different cues or even the same cues with different weights for different judgments could be responsible for the lack of consistency between the judgments. The cue combinations could therefore be influenced by subtle details of the way in which the response is measured or the comparison made. For instance, asking for equal distances might encourage people to partly rely on directly comparing retinal image sizes, which, of course, is not a veridical cue for judging separations in the world if one of the lines is receding in depth. This cue cannot be used to judge an object's distance from oneself, so it will not influence blind walking.

PERCEIVED MOTION IN DEPTH

A clear example of inconsistencies in depth perception is the comparison of perceived motion in depth with the perceived changes in distance due to the same motion stimulus (Brenner et al., 1996). Retinal image size is not the main cue for judging distance, but *changing* image size plays a very prominent role when judging motion in depth (Brenner

et al., 1996; Gray & Regan, 1996; Regan, 1997; Regan, Kaufman, & Lincoln, 1986). A constant image size can completely overrule changing binocular information about motion in depth if the image is large (Erkelens & Collewijn, 1985; Glennerster et al., 2006). The reason that image size plays a so much more prominent role in judgments of motion in depth than in judgments of distance is probably that rather than having to assume that the structure of interest has a certain size, of which one cannot usually be very certain, one only has to assume that the size is not changing.

The motion of a small dot in a scene consisting of static dots is easier to detect if the small dot is moving in the same direction in both eyes than if it is moving in opposite directions in the two eyes (Sumnall & Harris, 2000, 2002). Perhaps the binocular cue is combined with other depth cues, such as the not-changing size and required accommodation, despite the very small size of the target. This would be consistent with judgments of other attributes being determined by the resolution of the combined cues, rather than by the resolution of the individual cues (Hillis et al., 2004; Lugtigheid, Brenner, & Welchman, 2011; Sousa et al., 2009). However, it is also plausible that specialized motion detectors detect lateral motion, while motion in depth is detected on the basis of changes in disparity rather than on the basis of differences between the motion in the two eyes (Harris & Rushton, 2003).

THE SPECIAL ROLE OF DISTANT STRUCTURES

When a single object is presented in the dark, people tend to misjudge its distance (e.g., Gogel & Tietz, 1973). Different studies report different systematic errors, but all agree that the range of distances is underestimated: Close objects seem to be farther away than they are, and far objects seem to be nearer than they are. This underestimation of the range of distances can be interpreted as subjects considering certain distances to be more likely than others (a distance prior). For instance, they might assume that the farthest an object can be within the room in which the experiment is conducted is about 2 m, or that the farthest they can see when looking downward is about 1.5 m below their eyes (the normal distance to the ground beneath their feet). They may even assume both of the above, so that the prior about an object's distance depends on circumstances such as the direction in which one is looking, possibly contributing to the systematic tendency to overestimate vertical with respect to horizontal distances (Higashiyama & Ueyama, 1988). However, for understanding the underestimation of the range of distances of isolated objects, the absence of a structured background might also be relevant, because there are numerous examples of distant structures playing a role that they obviously cannot play if they are not visible.

The most obvious example of distant structures playing a special role is the role of the horizon when judging height in the visual field with respect to the horizon (Gardner et al., 2010), but distant structures also appear to play a role in interpreting binocular disparities and optic flow. In binocular vision there is a minimal angle of zero degrees between the directions with respect to the two eyes when a structure is very far away so that the lines of sight of the two eyes would be almost parallel if one were to try to fixate it. This minimal angle can be used to limit the range of possible distances for an object of interest, because no structure in the scene could give rise to a horizontal disparity corresponding with a vergence angle of less than 0° (Sousa et al., 2010). In a similar manner, because the direction to distant structures hardly changes when we move, it is reasonable to relate the changes in the directions to objects of interest to the direction to distant structures

(Brenner & van den Berg, 1996; van den Berg & Brenner, 1994). Relating distances to the most distant structure will obviously result in an underestimation of the range of distances if the farthest structure is the only structure. Thus, some of the systematic errors that are found when isolating cues might be artifacts of removing the distant structures that are normally used to help scale the cues themselves.

SIZE AND SHAPE

We have to judge distances from ourselves for knowing where things are. We also need judgments of distance to interpret retinal depth cues such as horizontal disparity in terms of separations between structures in depth, to determine how far objects are from each other, or how far they extend along the line of sight. Without any scaling, retinal depth cues can only tell us the depth order. Judgments of distance are therefore not only essential for judging distance, but also for judging size and shape. Because this chapter is about depth perception, we have emphasized the use of retinal image size to judge distance, assuming or knowing that the object has a certain size. The retinal image size is obviously also needed to judge an object's size (dimensions in the directions of azimuth and elevation), given an estimate of the distance ($r = d / \tan \alpha$ in Figure 9.8). Using object size to estimate distance and judged distance to estimate size provides an obvious problem if they are estimated sequentially. It would therefore make sense to estimate both together rather than sequentially by finding the combination of both that is most consistent with all the available information (the Bayesian approach proposed in Bülthoff, 1991). In that case, it may appear that only misjudging the retinal image size could lead to the inconsistencies between the two judgments that have often been reported (e.g., Kilpatrick & Ittelson, 1953). However, this is not necessarily true, because if the reported judgments were not made completely simultaneously, both the active acquisition of information (for instance by eye movements) and the way the cues were combined could have been optimized for the instantaneous judgment of interest, even if the other judgment was also estimated at that time. If so, the inconsistency would be across moments in time, rather than between the attributes (size and distance), with the judgment that one is going to use (to perform an action or report about) at each moment determining how the cue combination is optimized at that moment.

Size and distance have a very straightforward relationship. The relationship between shape and distance is more complex. Shape could be derived from separate judgments of size and depth, but judgments of shape could be made without estimating either the actual size or the actual depth, because only the relationship between the extents in different directions needs to be known. For instance, if an object is rotating, the motion in its image provides information about the shape without first requiring any scaling by distance (structure from motion; Figure 9.21A; Todd, 1985). Similarly, texture cues to slant can inform you about surface orientation even when the distance remains unknown (Figure 9.21B; Rosenholtz & Malik, 1997). Such information could be used to recognize objects by their shape. It is less evident that it could be useful for knowing where an object is and whether one can grasp it, but knowing the shape could theoretically contribute to judging the distance and size because the extents in different directions (what we have been calling size and depth) scale differently with distance, so knowing the shape could influence what is considered to be the most likely distance given the combination of disparities and retinal extents.
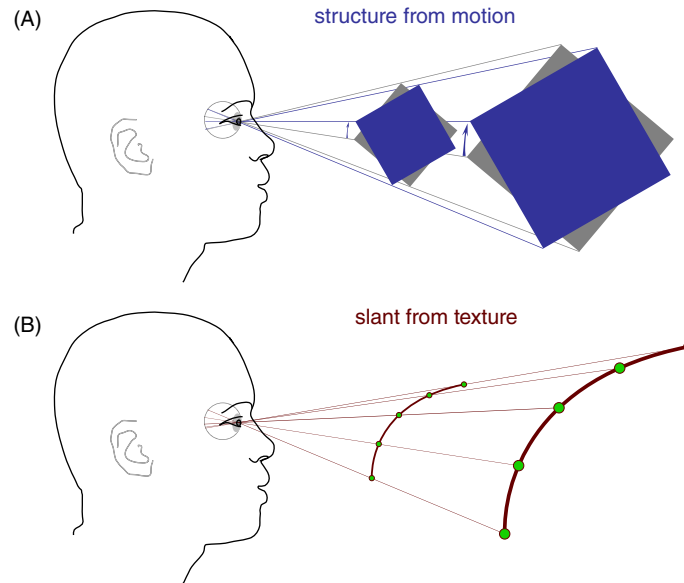
Figure 9.21   Shape and slant cues that are independent of distance. (A) Rotating a cube leads to the same image motion if the cube is both twice as far and twice as large. (B) Judgments of local slant from the separation of texture elements in the retinal image are also independent of the viewing distance.

As previously mentioned, people misjudge the distance of isolated objects in the dark. It is therefore not surprising that they also misjudge the shape of simulated isolated objects: the horizontal disparities and lateral extents are scaled by incorrect estimates of distance (Johnston, 1991). Perhaps surprisingly, rotating such simulated objects to make people see the shape correctly does not affect the judged distance (as measured by pointing distance and judged size; Brenner & van Damme, 1999). Thus, people tolerate inconsistencies between the size and depth that are derived from scaled retinal cues and the shape that is derived from unscaled retinal cues when judging shape and distance, rather than searching for the combination of distance and shape that is most consistent with all the cues. Note that this is a different case than when misjudging distance gives rise to perceived motion, and vice versa (Ono & Ujike, 2005), because in that case the conflict is attributed to another percept. In the case of rotating an object providing reliable information about its depth, the conflict is not attributed to the judged size or perceived distance. A difficult task for future research will be to explain why conflicts between the ways in which cues are interpreted are tolerated for some combinations of attributes, but not for others.

CONCLUSION

We have seen that many cues contribute to judgments of distance. They do so to varying extents depending on the circumstances. For instance, height in the visual field is a very useful cue as long as you are interested in objects resting on flat surfaces at a known height. Similarly, binocular cues are very reliable as long as the object of interest is nearby. Our description of the many cues and their limitations is obviously far from complete. Although the main cues have been

known for many decades, research on the ways in which they are combined and on how they influence each other and are influenced by specific aspects of the surrounding is relatively new.

REFERENCES

Adams, W., Frisby, J. P., Buckley, D., Gårding, J., Hippisley-Cox, S. D., & Porrill, J. (1996). Pooling of vertical disparities by the human visual system. *Perception*, *25*, 165–176.

Allen, B. P. (1999). Shadows as sources of cues for distance of shadow-casting objects. *Perceptual and Motor Skills*, *89*, 571–584.

Anobile, G., Cicchini, G. M., & Burr, D. C. (2014). Separate mechanisms for perception of numerosity and density. *Psychological Science*, *25*, 265–270.

Backus, B. T., Banks, M. S., van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, *39*, 1143–1170.

Bradshaw, M. F., Hibbard, P. B., Parton, A. D., Rose, D., & Langley, K. (2006). Surface orientation, modulation frequency and the detection and perception of depth defined by binocular disparity and motion parallax. *Vision Research*, *46*, 2636–2644.

Bradshaw, M. F., & Rogers, B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Research*, *36*, 3457–3468.

Brenner, E., Driesen, B., & Smeets, J. B. J. (2014). Precise timing when hitting falling balls. *Frontiers in Human Neuroscience*, *8*, 342.

Brenner, E., & Smeets, J. B. J. (2000). Comparing extra-retinal information about distance and direction. *Vision Research*, *40*, 1649–1651.

Brenner, E., Smeets, J. B. J, & Landy, M. S. (2001). How vertical disparities assist judgements of distance. *Vision Research*, *41*, 3455–3465.

Brenner, E., & van Damme, W. J. (1998). Judging distance from ocular convergence. *Vision Research*, *38*, 493–498.

Brenner, E., & van Damme, W. J. (1999). Perceived distance, shape and size. *Vision Research*, *39*, 975–986.

Brenner, E., van den Berg, A. V., van Damme, W. J. (1996). Perceived motion in depth. *Vision Research*, *36*, 699–706.

Brenner, E., & van den Berg, A. V. (1996). The special role of distant structures in perceived object velocity. *Vision Research*, *36*, 3805–3814.

Bridgeman B., & Cooke, I. (2015). Effect of eye height on estimated slopes of hills. *Perception*, *44*, 755–763.

Bülthoff, H. H. (1991). Shape from X: Psychophysics and computation. In M. S. Landy & J. A. Movshon, *Computational models of visual processing* (pp. 305–330). Cambridge, MA: MIT Press.

Collett, T. S., Schwarz, U., & Sobel, E. C. (1991). The interaction of oculomotor cues and stimulus size in stereoscopic depth constancy. *Perception*, *20*, 733–754.

Cooper, E. A., Burge, J., & Banks, M. S. (2011). The vertical horopter is not adaptable, but it may be adaptive. *Journal of Vision*, *11*(3), 20, 1–19.

Cuijpers, R. H., Kappers, A. M., & Koenderink, J. J. (2000). Investigation of visual space using an exocentric pointing task. *Perception & Psychophysics*, *62*, 1556–1571.

Cumming, B. G., Johnston, E. B., & Parker, A. J. (1991). Vertical disparities and perception of three-dimensional shape. *Nature*, *349*, 411–413.

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Roger (Eds.), *Perception of space and motion* (pp. 69–117). New York, NY: Academic Press.

Daum, S. O., & Hecht, H. (2009). Distance estimation in vista space. *Attention, Perception, & Psychophysics*, *71*, 1127–1137.

Duchon, A. P., & Warren, W. H. Jr. (2002). A visual equalization strategy for locomotor control: of honeybees, robots, and humans. *Psychological Science*, *13*, 272–278.

Duke, P. A., Oruç, I., Qi, H., & Backus, B. T. (2006). Depth aftereffects mediated by vertical

disparities: evidence for vertical disparity driven calibration of extraretinal signals during stereopsis. *Vision Research*, *46*, 228–241.

Enright, J. T. (1991). Exploring the third dimension with eye movements: Better than stereopsis. *Vision Research*, *31*, 1549–1562.

Enright, J. T. (1996). Sequential stereopsis: A simple demonstration. *Vision Research*, *36*, 307–312.

Erkelens, C. J., & Collewijn, H. (1985). Motion perception during dichoptic viewing of moving random-dot stereograms. *Vision Research*, *25*, 583–588.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, *87*, 411–434.

Frisby, J. P, Buckley, D., & Horsman, J. M. (1995). Integration of stereo, texture, and outline cues during pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception*, *24*, 181–198.

Gardner, J. S., Austerweil, J. L., & Palmer, S. E. (2010). Vertical position as a cue to pictorial depth: height in the picture plane versus distance to the horizon. *Attention, Perception, & Psychophysics*, *72*, 445–453.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

Gibson, E. J., Gibson, J. J., Smith, O. W., & Flock, H. (1959). Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology*, *58*, 40–51.

Gillam, B. (1995). The perception of spatial layout from static optical information. In S. R. W. Epstein (Ed.), *Perception of space and motion* (pp. 23–67). London, United Kingdom: Academic Press.

Gillam, B., & Lawergren, B. (1983). The induced effect, vertical disparity, and stereoscopic theory. *Perception and Psychophysics*, *34*, 121–130.

Glennerster, A., McKee, S. P., & Birch, M. D. (2002). Evidence for surface-based processing of binocular disparity. *Current Biology*, *12*, 825–828.

Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., & Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology*, *16*, 428–432.

Gogel, W. C. (1990). A theory of phenomenal geometry and its applications. *Perception & Psychophysics*, *48*, 105–123.

Gogel, W. C., & Tietz, J. D. (1973). Absolute motion parallax and the specific distance tendency. *Perception & Psychophysics*, *13*, 284–292.

Gray, R., & Regan, D. (1996). Cyclopean motion perception produced by oscillations of size, disparity and location. *Vision Research*, *36*, 655–665.

Harris, J. M., & Rushton, S. K. (2003). Poor visibility of motion in depth is due to early motion averaging. *Vision Research*, *43*, 385–392.

Harris, J. M., & Wilcox, L. M. (2009). The role of monocularly visible regions in depth and surface perception. *Vision Research*, *49*, 2666–2685.

Held, R. T., Cooper, E. A., & Banks, M. S. (2012). Blur and disparity are complementary cues to depth. *Current Biology*, *22*, 1–6.

Higashiyama, A., & Ueyama, E. (1988). The perception of vertical and horizontal distances in outdoor settings. *Perception & Psychophysics*, *44*, 151–156.

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: optimal cue combination. *Journal of Vision*, *4*, 967–992.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. Oxford Psychology Series No. 29. Oxford, United Kingdom: Oxford University Press.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*, 3621–3629.

Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, *31*, 1351–1360.

Kelly, J. W., Loomis, J. M., & Beall, A. C. (2004). Judgments of exocentric direction in large-scale space. *Perception*, *33*, 443–454.

Kersten, D., Mamassian, P., & Knill, D. C. (1997). Moving cast shadows induce apparent motion in depth. *Perception*, *26*, 171–192.

Kilpatrick, F. P., & Ittelson, W. H. (1953). The size-distance invariance hypothesis. *Psychological Reviews*, *60*, 223–231.

Knill, D. (2007). Learning Bayesian priors for depth perception. *Journal of Vision*, *7*(8), 13.1–13.20.

Koenderink, J. J. (1986). Optic flow. *Vision Research*, *26*, 161–180.

Kudoh, N. (2005). Dissociation between visual perception of allocentric distance and visually directed walking of its extent. *Perception*, *34*, 1399–1416.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*, 389–412.

Li, W., Dallal, N., & Matin, L. (2001). Influences of visual pitch and visual yaw on visually perceived eye level (VPEL) and straight ahead (VPSA) for erect and rolled-to-horizontal observers. *Vision Research*, *41*, 2873–2894.

Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance 18*, 906–921.

Lugtigheid, A. J., Brenner, E., & Welchman, A. E. (2011). Speed judgments of three-dimensional motion incorporate extraretinal information. *Journal of Vision*, *11*(13), 1.1–1.11.

Lugtigheid, A., & Welchman, A. (2010). A surprising influence of retinal size on disparity-defined distance judgments. *Journal of Vision*, *10*(7), 63.

Mamassian, P., & Landy, M. S. (2001). Interaction of visual prior constrains. *Vision Research*, *41*, 2653–2668.

Marshall, J. A., Burbeck, C. A., Ariely, D., Rolland, J. P., & Martin, K. E. (1996). Occlusion edge blur: A cue to relative visual depth. *Journal of the Optical Society of America A. Optics, Image Science, and Vision 13*, 681–688.

McIntosh, R. D., & Lashley, G. (2008). Matching boxes: Familiar size influences action programming. *Neuropsychologia 46*, 2441–2444.

McKee, S. P. (1983). The spatial requirements for fine stereoacuity. *Vision Research*, *23*, 191–198.

Meng, J. C., & Sedgwick, H. A. (2001). Distance perception mediated through nested contact relations among surfaces. *Perception & Psychophysics*, *63*, 1–15.

Mitchison, G. J., & Westheimer, G. (1984). The perception of depth in simple figures. *Vision Research*, *24*, 1063–1073.

Muller, C. M., Brenner, E., & Smeets, J. B. J. (2008). Testing a counter-intuitive prediction of optimal cue combination. *Vision Research*, *49*, 134–139.

Muller, C.M.P., Brenner, E., & Smeets, J. B. J. (2009). Maybe they are all circles: Clues and cues. *Journal of Vision*, *9*(9), 10.1–10.5.

Nachmias, J. (2008). Judging spatial properties of simple figures. *Vision Research*, *48*, 1290–1296.

Ono, H., & Ujike, H. (2005). Motion parallax driven by head movements: Conditions for visual stability, perceived depth, and perceived concomitant motion. *Perception*, *34*, 477–490.

Ooi, T. L., Wu, B., & He, Z. J. (2001). Distance determined by the angular declination below the horizon. *Nature*, *414*, 197–200.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, *43*, 2451–2468.

O'Shea, R. P., Govan, D. G., & Sekuler, R. (1997). Blur and contrast as pictorial depth cues. *Perception*, *26*, 599–612.

Peh, C. H., Panerai, F., Droulez, J., Cornilleau-Pérès, V., & Cheong, L. F. (2002). Absolute distance perception during in-depth head movement: Calibrating optic flow with extra-retinal information. *Vision Research*, *42*, 1991–2003.

Philbeck, J. W., Loomis, J. M., & Beall, A. C. (1997). Visually perceived location is an invariant in the control of action. *Perception & Psychophysics*, *59*, 601–612.

Ramamurthy, M., Bedell, H. E., & Patel, S. S. (2005). Stereo thresholds for moving line stimuli for a range of velocities. *Vision Research*, *45*, 789–799.

Read, J. C. (2010). Vertical binocular disparity is encoded implicitly within a model neuronal

population tuned to horizontal disparity and orientation. *PLoS Computational Biology*, *6*(4), e1000754.

Regan, D. (1997). Visual factors in hitting and catching. *Journal of Sports Sciences*, *15*, 533–558.

Regan, D. M., Kaufman, L., & Lincoln, J. (1986). Motion in depth and visual acceleration. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 1. Sensory processes and perception* (pp. 19.1–19.46). New York, NY: Wiley.

Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception*, *24*, 155–179.

Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, *8*, 125–134.

Rosenholtz, R., & Malik, J. (1997). Surface orientation from texture: Isotropy or homogeneity (or both)? *Vision Research*, *37*, 2283–2293.

Schor, C. M., & Badcock, D. R. (1985). A comparison of stereo and vernier acuity within spatial channels as a function of distance from fixation. *Vision Research*, *25*, 1113–1119.

Sedgwick, H. A. (1986). Space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance; sensory processes and perception* (pp. 1–57). New York, NY: Wiley.

Seydell, A., Knill, D., & Trommershäuser, J. (2010). Adapting internal statistical models for interpreting visual cues to depth. *Journal of Vision*, *10*(4), 1.1–1.27.

Siderov, J., & Harwerth, R. S. (1995). Stereopsis, spatial frequency and retinal eccentricity. *Vision Research*, *35*, 2329–2337.

Siderov, J., Harwerth, R. S., & Bedell, H. E. (1999). Stereopsis, cyclovergence and the backwards tilt of the vertical horopter. *Vision Research*, *39*, 1347–1357.

Sobel, E. C., & Collett, T. S. (1991). Does vertical disparity scale the perception of stereoscopic depth? *Proceedings of the Royal Society B: Biological Sciences*, 244, 87–90.

Sousa R., Brenner, E., & Smeets, J. B. J. (2009). Slant cue are combined early in visual processing: Evidence from visual search. *Vision Research*, *49*, 257–261.

Sousa R., Brenner, E., & Smeets, J. B. J. (2010). A new binocular cue for absolute distance: Disparity relative to the most distant structure. *Vision Research*, *50*, 1786–1792.

Sousa R., Brenner E., & Smeets, J. B. J. (2011). Judging an unfamiliar object's distance from its retinal image size. *Journal of Vision*, *11*(9), 10.1–10.6.

Sousa R., Smeets, J. B. J., & Brenner, E. (2012). Does size matter? *Perception*, *41*, 1532–1534.

Sousa, R., Smeets, J. B. J., & Brenner, E. (2013). The influence of previously seen objects' sizes in distance judgments. *Journal of Vision*, *13*(2), 2

Sumnall, J. H., & Harris, J. M. (2000). Binocular three-dimensional motion detection: Contributions of lateral motion and stereomotion. *Journal of the Optical Society of America A. Optics, Image Science, and Vision*, *17*, 687–696.

Sumnall, J. H., & Harris, J. M. (2002). Minimum displacement thresholds for binocular three-dimensional motion. *Vision Research*, *42*, 715–724.

Todd, J. T. (1985). Perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 689–710.

Tresilian, J. R. (1993). Four questions of time to contact: A critical examination of research on interceptive timing. *Perception*, *22*(6), 653–680.

van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, *81*, 1355–1364.

van Beers, R. J., van Mierlo, C. M., Smeets, J. B. J., & Brenner, E. (2011). Reweighting visual cues by touch. *Journal of Vision*, *11*(10), 20.1–20.16.

van den Berg, A. V. (1992). Robustness of perception of heading from optic flow. *Vision Research*, *32*, 1285–1296.

van den Berg, A. V., & Brenner, E. (1994). Humans combine the optic flow with static depth cues for

robust perception of heading. *Vision Research*, *34*, 2153–2167.

van Ee, R., Adams, W. J., & Mamassian, P. (2003). Bayesian modeling of cue interaction: Bistability in stereoscopic slant perception. *Journal of the Optical Society of America A. Optics, Image Science, and Vision*, *20*, 1398–1406.

van Ee, R., van Dam, L. C., & Erkelens, C. J. (2002). Bi-stability in perceived slant when binocular disparity and monocular perspective specify different slants. *Journal of Vision*, *2*, 597–607.

van Mierlo, C. M., Louw, S., Smeets, J. B. J., & Brenner, E. (2009). Slant cues are processed with different latencies for the online control of movement. *Journal of Vision*, *9*(3), 25.1–25.8.

Warren, P. A., & Mamassian, P. (2010). Recovery of surface pose from texture orientation statistics under perspective projection. *Biological Cybernetics*, *103*, 199–212.

Warren, P. A., & Rushton, S. K. (2008). Evidence for flow-parsing in radial flow displays. *Vision Research*, *48*, 655–663.

Warren, P. A., & Rushton, S. K. (2009). Optic flow processing for the assessment of object movement during ego movement. *Current Biology*, *19*, 1555–1560.

Warren, W. H., & Hannon, D. J. (1988). Direction of self-motion is perceived from optical flow. *Nature*, *336*, 162–163.

Watt S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, *5*(10), 834–862.

Westheimer, G., & McKee, S. P. (1975). Visual acuity in the presence of retinal-image motion. *Journal of the Optical Society of America*, *65*, 847–850.

Westheimer, G., & McKee, S. P. (1978). Stereoscopic acuity for moving retinal images. *Journal of the Optical Society of America*, *68*, 450–455.