Title
Studies of human depth perception in real and pictured environments

Author
Cooper, Emily Averill

Publication Date
2012

Peer reviewed|Thesis/dissertation

Studies of human depth perception in real and pictured environments

by

Emily Averill Cooper

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Neuroscience

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Professor Martin S. Banks, Chair
Professor Clifton M. Schor
Professor Dennis M. Levi
Professor David Whitney

Fall 2012

Abstract

Studies of human depth perception in real and pictured environments

by

Emily Averill Cooper

Doctor of Philosophy in Neuroscience

University of California, Berkeley

Professor Martin S. Banks, Chair

Determining the three-dimensional (3D) layout of the natural environment is an important but difficult visual function. It is difficult because the 3D layout of a scene and the shapes of 3D objects are not explicit in the 2D retinal images. The visual system uses both binocular (two eyes) and monocular (one eye) cues to determine 3D layout from these images. The same cues that are used to determine the layout of real scenes are also present in photographs and realistic pictures of 3D scenes. The work described in this dissertation examines how people perceive 3D layout in the natural environment and in pictures. Chapter 1 provides an introduction to the specific 3D cues that will be studied. Chapter 2 describes a series of experiments examining how the visual system's use of binocular cues might be adaptive to improve depth perception in the natural environment. Chapter 3 describes a series of experiments examining how the visual system interprets monocular linear perspective cues, and how this can lead to perceptual distortions in perceived 3D from pictures. Chapter 4, the Conclusion, discusses how the work described in this dissertation might be applied to creating better displays and pictures.

Acknowledgements

Thank you to my advisor Marty Banks; co-authors Johannes Burge (Chapter 2) and Elise Piazza (Chapter 3), whose work is reproduced with permission; dissertation committee—Cliff Schor, Dennis Levi, and David Whitney; Jim Maxwell for help with measuring torsional eye movements for Chapter 2; Darius Lerup for help in conducting the experiments in Chapter 3, Eric Kee for providing the 3D face model for Chapter 3; Lawrence Arend, James O'Shea, Maneesh Agrawala, David Whitney, and Stephen Palmer for helpful comments on the manuscript for Chapter 3; Hany Farid; Allyson Mackey; Wendy de Heer; Bill Sprague; Tony Leonard; Kati Markowitz. This work was supported by NIH Research Grant EY012851, NSF Grant BCS-0617701, the National Defense Science and Engineering Graduate Fellowship (NDSEG) Program, and National Science Foundation under Grant No. (DGE 1106400).

The majority of the work in this dissertation has been previously published in the following venues:

Articles:

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.

Cooper, E. A., Piazza, E. A., Banks, M. S. (2012). The Perceptual Basis of Common Photographic Practice. *Journal of Vision*, **12**(5):8, 1-14.

Abstracts:

Cooper, E. A., Burge, J., Banks, M. S. (2010). Do People of Different Heights Have Different Horopters? *Journal of Vision*, **10**(7):372.

Cooper, E. A., Piazza, E. A., Banks, M. S. (2011). Depth Compression and Expansion in Photographs. *Journal of Vision*, **11**(11):65.

Cooper, E. A., Banks, M. S. (2012). Perception of Depth in Pictures when Viewing from the Wrong Distance. *Journal of Vision*, **12**(9):896.

Table of Contents

Chapter 1. Introduction

1.1 Background

The human visual system evolved in a predictable, structured environment. Visual input from this environment provides a variety of reliable cues for determining 3D layout. These cues arise from the fact that certain aspects of 3D layout, such as the ordering of objects in depth and their 3D shapes, are correlated with certain patterns in the retinal images (Howard & Rogers, 2002). One such cue is binocular disparity. Binocular disparities are the differences between the retinal images of the two eyes, which arise because of the eyes' slightly different viewpoints (Figure 1.1a). For example, objects that are farther in depth than the point at which the eyes are fixating project to images shifted rightward in the left eye and leftward in the right eye (uncrossed disparity); objects that are nearer in depth project to images shifted leftward in the left eye and rightward in the right eye (crossed disparity). The visual system is highly sensitive to these differences in the eyes' images. Binocular disparities alone are sufficient to produce a vivid sense of 3D (Julesz, 1964; Blakemore, 1970).

There are also many monocular cues to 3D layout. Cues like relative size, texture gradients, and perspective convergence (together called perspective cues) all refer to the patterns that occur in a single eye's projected image of 3D space (Burton, 1945). For instance, if objects of uniform size are distributed in depth, those that are farther from the viewer project to smaller images and those that are closer project to larger images (Figure 1.1b). Similarly, parallel lines that recede in depth project to converging lines, because at farther distances the separation between the lines in the world projects to a smaller size on the retina. Perspective cues create a strong sense of 3D space even when only one eye is used to view the world (i.e., when there is no binocular disparity) (Smith & Gruber, 1958). The visual system uses these binocular and monocular cues together to recover the 3D layout of the natural environment.

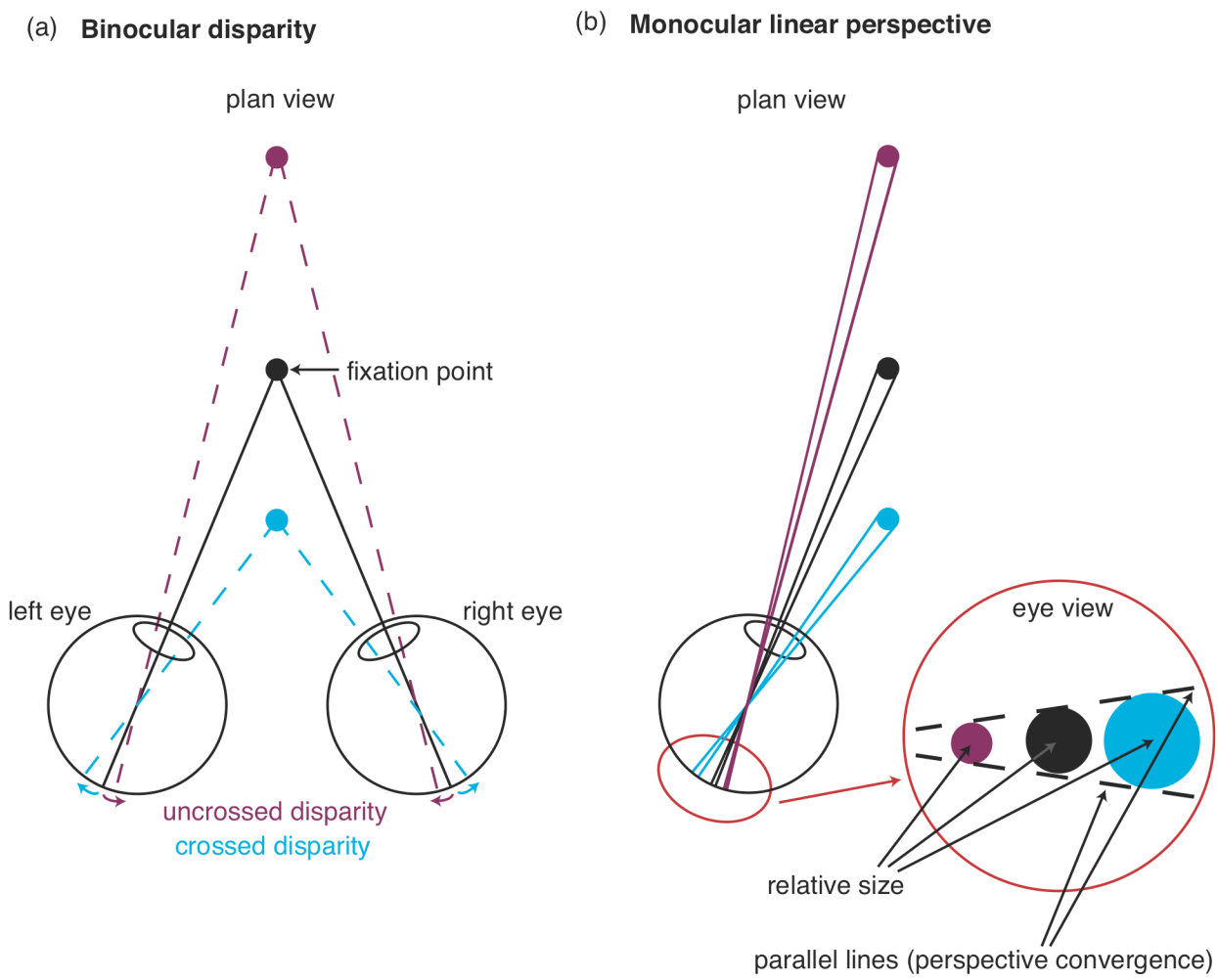

Figure 1.1. Depth cues. (a) Binocular disparity. The left and right eyes fixate a point in space (fixation point). The fixation point projects to the two foveas. A point further away in depth (maroon) projects to retinal locations that are shifted rightward in the left eye and leftward in the right eye. This is called uncrossed disparity. A point closer in depth (blue) projects to retinal locations that are shifted leftward in the left eye and rightward in the right eye. This is called crossed disparity. (b) Monocular linear perspective. Three objects of uniform size are distributed in depth. The closest object (blue) projects to a retinal image with the largest angular size. The other farther objects (black, maroon) project to retinal images that are relatively smaller. Inset: the relative sizes of the objects provide a cue to depth. If lines are drawn to connect the tops and bottoms of the objects, these are parallel lines in space. These lines converge in the eye's image (perspective convergence).

Depth cues are also a key reason why a picture that is drawn or printed on a flat surface can create a vivid sense of 3D layout. Two-dimensional (2D) pictorial representations, such as photographs, computer-graphics images, and perspective paintings, yield the perception of 3D space because the retinal image when looking at a picture contains many of the same cues as the retinal image when looking at the original 3D scene. For instance, a pair of converging lines drawn on a piece of paper can create the perception of two parallel lines receding in depth (Figure 1.2a; D'Amelio,

2004). The more cues contained in the picture, the more it can create a sense of 3D (Figure 1.2b). Adding binocular disparities to a picture, such that two separate pictures are projected to the two eyes (a stereo picture), can create an even more vivid sense of 3D. However, there is a critical limitation of any picture. A single picture on a flat surface only recreates the depth cues for a single viewpoint of a 3D scene. This means that there is only one place to view a picture from where the retinal image cast by the picture is the same as the image that would have been cast by the original 3D scene. That is, when looking at a picture from the correct position, the perspective cues in the picture correctly specify the 3D structure of the original scene. When pictures are viewed from locations other than the correct one, both the binocular and monocular depth cues can specify a 3D scene that is distorted relative to the original scene (Sedgwick, 1991). In addition to these distortions created by incorrect viewing position, perceptual distortions can occur because of depth cues outside of the picture. For instance, when viewing a photograph of a road receding into the distance, the monocular linear perspective—the convergence of the two sides of the road as it recedes in depth—specifies a 3D scene, but the binocular disparity cue specifies the flat, rectangular surface on which the photograph is printed (Figure 1.2c; Hagen, Glick & Morse, 1978). Some research has found that this flatness cue causes perceived depth in pictures to be flattened relative to real 3D scenes (Sedgwick, Nicholls, & Brehaut, 1995; Todorović, 2009). So while perception of 3D layout from pictures is very related to perception of 3D layout in the real environment, pictures pose their own unique challenges for the visual system to solve.

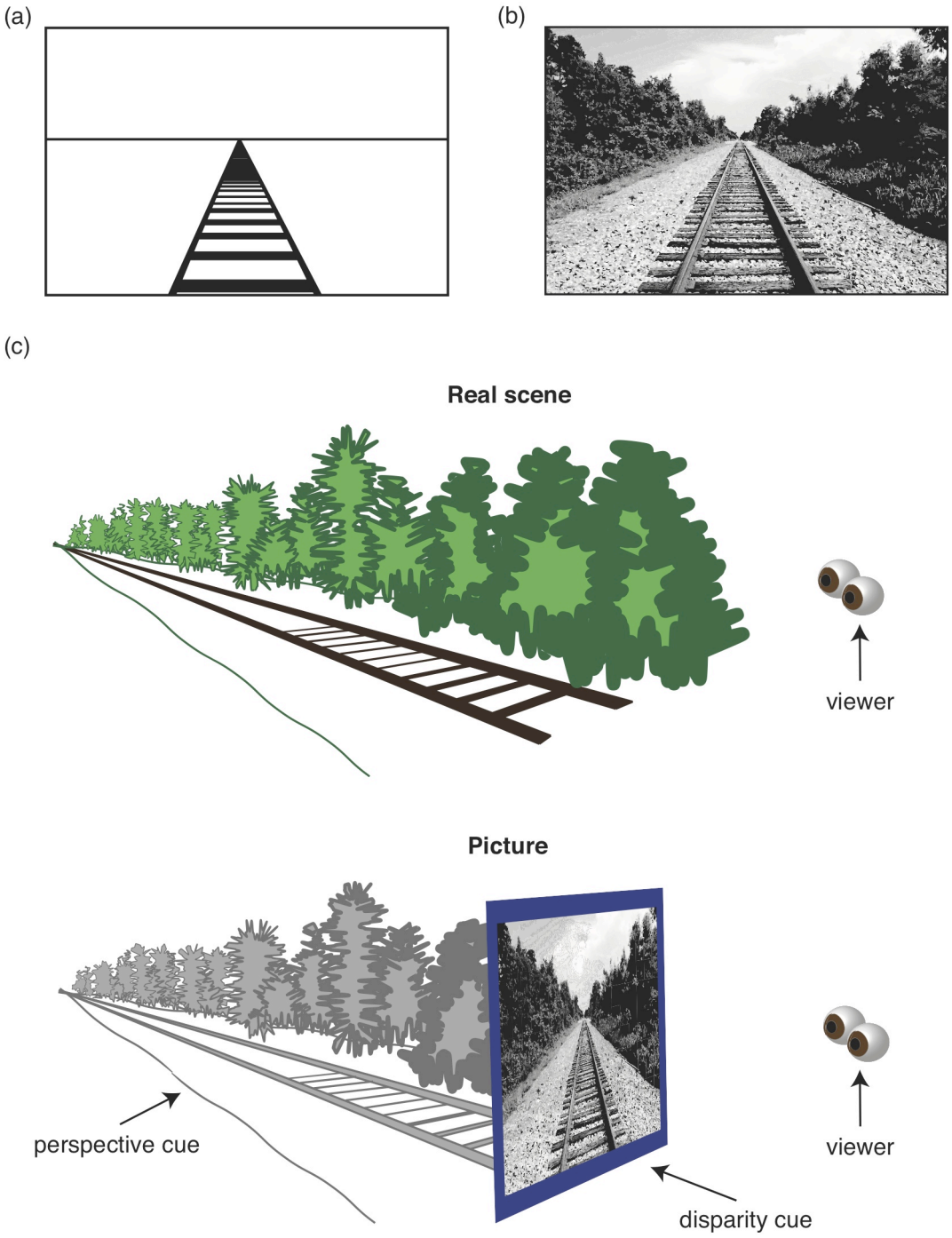

Figure 1.2. Depth cues in pictures. (a) A simple drawing with lines converging towards each other can create the appearance that parallel lines are receding in depth. (b) Adding details like shading and texture gradients can increase the appearance of depth in the picture. (c) Upper Panel: A viewer looks at a real 3D scene of a railroad. Lower Panel: When a viewer looks at this picture from the correct position, the perspective cue in the picture matches the 3D structure of the real scene. But because the picture shows only a single viewpoint of the scene, the disparity cue indicates a flat surface.

The aim of this dissertation is to examine the relationships between the visual system, the natural environment, and pictorial representations of the environment. The experimental work is separated into two chapters, each focusing in-detail on a different aspect of this problem.

1.2 Outline of dissertation

Chapter 2 describes the first set of experimental work. This work examines how the visual system might be adaptive for and adaptable by patterns in binocular disparities created by the natural environment. Depth estimates from binocular disparity are most precise when the visual input stimulates corresponding retinal points (Blakemore, 1970). Corresponding retinal points are points that, when stimulated, yield the same perceived-direction in the two eyes (i.e., they are perceived to have zero binocular disparity). Corresponding points tend to have uncrossed disparities in the upper visual field and crossed disparities in the lower visual field. They also tend to have uncrossed disparities in the left and right visual fields. Due to the disparities, the region of 3D space that stimulates corresponding points is pitched top-back—that is, depth perception from disparity is more precise behind fixation in the upper visual field and more precise in front of fixation in the lower visual field; in the left and right visual fields, depth perception from disparity is more precise at points that are farther than the point of fixation (Helmholtz, 1925). This region of 3D space is called the horopter, and the portion that falls along the vertical meridians of the eyes is called the vertical horopter.

Many have suggested that the top-back pitch of the vertical horopter is advantageous for discriminating depth in the natural environment, particularly relative to the ground (Helmholtz 1925; Breitmeyer, Battaglia, & Bridge, 1977; Nakayama, 1977). This is because when an observer looks at a point on the ground in front of them, the pitch of the ground plane qualitatively matches the pitch of the vertical horopter. To determine if the shape of the vertical horopter is in fact advantageous for depth perception, we asked whether the vertical horopter is adaptive (suited for perception of the ground) and adaptable (changeable by experience).

Experiment 1 tested whether the vertical horopter is adaptive for make depth discriminations relative to the ground plane of individual observers. We measured the disparities between corresponding points along the vertical meridians of the eyes in observers with a wide range of heights, from 4.3 to 7.0 ft (129.5 to 213.4 cm) tall. When standing on a flat surface, each eye height corresponded to a different top-back pitch angle of the ground plane in the visual field. Knowing each observer's eye height, we could then test whether the pitch of their vertical horopter matched the pitch of the ground. We confirmed that the horopter is pitched top-back in almost all observers. However, we did not find any evidence that it is well-suited to improve depth perception for the pitch of the ground plane for an individual. We also found that the vertical horopter is typically convex, curving back in depth away above and below the point of

fixation. This shape makes it generally ill-suited for depth perception relative to the ground, but possibly adaptive for other convex shapes.

Experiment 2 tested whether the vertical horopter is adaptable by experience. We hypothesized that the disparities between corresponding points may be influenced by systematic patterns in disparity created by the whole environment, rather than just the ground plane. So in this experiment, observers wore distorting spectacles for 7 days that systematically shifted the pattern of binocular disparities seen by their eyes. One pair of spectacles created uncrossed disparities in the upper visual field and crossed disparities in the lower visual field; the relationship was flipped in the other pair. We made daily measurements of corresponding points to determine if the pitch angle of the vertical horopter decreased or increased depending on the disparities created by the spectacles. Even though 3D percepts changed, we observed no change in the horopter, suggesting that it is not adaptable, at least over the course of a week. Thus, if the horopter is adaptive for the natural environment, it may be hard-coded and not affected by daily experience. In a subsequent analysis of the data from these experiments, we also showed that the horopter is not adaptive for long viewing distances. At long viewing distances, the disparities between corresponding points in the upper visual field cannot possibly be stimulated in the natural environment. We concluded that the vertical horopter seems to be adaptive for perceiving convex, slanted surfaces at short distances. We discuss how understanding the shape of this region of binocular vision could be applied to improving stereo pictures and displays.

Chapter 3 describes the second set of experimental work. This work focuses on perception of 3D spatial layout in pictorial representations of the environment. A picture presents the opportunity to recreate an accurate representation of the environment for an observer who was not present at the time or place when the picture was captured. Pictures can also be manipulated to create purposefully inaccurate and artistically distorted representations. By using lenses of different focal lengths to create a picture, photographers, cinematographers, and computer-graphics engineers can make a scene look compressed or expanded in depth, make a familiar object look natural or distorted, or make a person look smarter, more attractive, or more neurotic. One particular focal length, 50mm, is considered a normal lens that creates the most natural-looking picture (Kingslake, 1992; Belt, 2008; Modrak & Anthes, 2011; London, Stone, & Upton, 2010). The perceptual and behavioral mechanisms mediating these lens effects on perceived 3D shape and layout are not well understood. We asked why pictures taken with a certain lens focal length look natural, while those taken with other focal lengths look distorted.

Experiment 1 measured how perceived 3D shape from pictures differs from perceived 3D shape in the real world. We found that when pictures are viewed from the correct distance, perceived shape in the realistic picture is very similar to perceived shape in a real scene. However, with less realistic pictures, like line drawings, perceived depth is highly compressed (i.e., shapes appear flattened in depth). This is likely due to the greater salience of conflicting depth information in a line drawing. Specifically, with fewer cues to shape in the 3D environment (i.e., texture, shading, and lighting differences), the cues about the picture's flat surface interfere (Hagen, Glick, & Morse, 1978). Experiment 2 measured how changing the viewing distance from a picture—

viewing from farther or closer than the correct distance—affects perceived depth. We found that perceived depth is greatly distorted in all types of pictures, both more realistic pictures and line drawing pictures. When pictures are viewed from too far, perceived depth is expanded relative to the original 3D scene; when pictures are viewed from too close, perceived depth is compressed relative to the original scene. While these effects had previously been reported in the literature (e.g., Smith & Gruber, 1958; Adams, 1972; Lumsden, 1983), other studies suggested that the distortions are quite small, perhaps because a perceptual mechanism compensates for incorrect viewing distance (Yang & Kubovy, 1999). In particular, it has been argued that people may compensate for these distortions because they are rarely noticed in day-to-day picture viewing (Hagen, 1974; Bengston, Stergios, Ward & Jester, 1980; Yang & Kubovy, 1999). By using a variety of pictures with different cues, we were able to provide strong evidence that changes in viewing distance create substantial perceptual changes. Such distortions may be rarely noticed in day-to-day picture viewing because of the behaviors people adopt when they view pictures.

Experiment 3 measured what distances people prefer to view pictures from for many different types of pictures, captured with different focal lengths, and printed at different physical sizes. We found that preferred viewing distance is mainly affected by the physical size of the picture and not by the content. People tended to view small pictures from closer and larger pictures from farther, regardless of the type of picture or the focal length used to capture it. We also found that people preferred that smaller pictures take up a smaller angle in their field of view (~22°) and that larger pictures take up a larger angle in their field of view (~36°). For the larger picture sizes, this suggests that people naturally tend to view pictures from a distance that is very close to the correct viewing distance, if the picture was captured with a normal focal length lens. So by following the popular rule of thumb of using a 50-mm lens, photographers and picture-creators greatly increase the odds of a viewer looking at a photograph from the correct distance. This may explain why many pictures viewed in day-to-day life do not obviously appear distorted. However, our results also suggest that the precise recommended focal length to maximize accuracy in perceived 3D layout should vary depending on the size of the picture. This in turn can provide more useful, nuanced guidelines for creating pictures. By using these guidelines, photographers and other picture creators can get a better sense of how the 3D layout of their picture will be perceived, whether it is a news photo, a perspective drawing, a film, or any other type pictorial representation.

Determining the 3D layout of an environment is a complex problem that the human visual system solves remarkably well. These two chapters focus on different aspects of the problem of 3D perception: binocular and monocular depth cues, real environments and pictured environments.

Chapter 2. Adaptiveness and adaptability of binocular corresponding points

2.1 Introduction

The ground is a prominent feature of the natural environment. It is usually perpendicular to the main axis of the head and body because humans tend to keep themselves aligned with gravity. The pervasiveness of the ground confers a simple relationship between distance and position in the visual field: Near points stimulate the lower field and far points stimulate the upper field. This relationship between environmental structure and position in the visual field yields a systematic pattern of binocular disparities on the retinas: crossed disparity below fixation and uncrossed disparity above fixation (Hibbard & Bouzit, 2005; Potetz & Lee, 2003; Yang & Purves, 2003). It would be useful to take advantage of this regularity when estimating the structure of the environment.

2.1.1 Geometric and empirical corresponding points

Depth estimates from disparity are most precise when the visual input strikes the retinas on empirical corresponding points (Blakemore, 1970). It is useful to describe those points with respect to geometric points. Geometric corresponding points are pairs of points with the same coordinates in the two retinas: By definition, they have zero disparity. The two anatomical vertical meridians of the eyes (great circles with zero azimuth) are an example of a set of geometric corresponding points. The locations in the world that stimulate geometric corresponding points define the geometric horopter. When fixation is in the head's mid-sagittal plane, the geometric vertical horopter is a vertical line through fixation (Figure 2.1a). Empirical corresponding points are generally defined by determining positions in the two retinas that, when stimulated, yield the same perceived-direction. Empirical and geometric points differ in that empirical points have uncrossed disparities in the upper visual field and crossed disparities in the lower field (i.e., in the upper field, points are offset leftward in the left eye relative to their corresponding points in the right eye; in the lower field, they are offset rightward). This pattern of offsets is often described as a horizontal shear between the empirical corresponding meridians, and this causes the empirical vertical horopter—the locus of points in the world that stimulate empirical corresponding points near the vertical meridians—to be pitched top-back (Figure 2.1b). The qualitative similarity between the disparities of empirical corresponding points and the disparities cast on the retinas by natural scenes has led to the hypothesis that corresponding points are adaptive for

precisely perceiving the 3D structure of the natural environment (Breitmeyer, Battaglia, & Bridge, 1977; Helmholtz 1925; Nakayama, 1977).

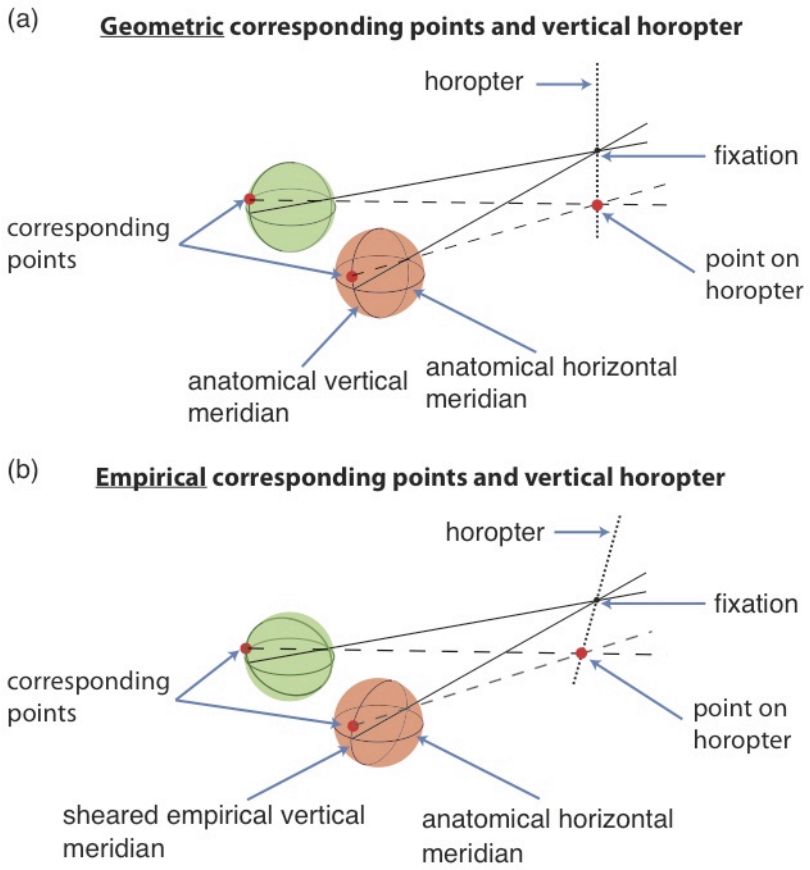

Figure 2.1. Geometric and empirical vertical horopters. Green and red spheres represent the left and right eyes, respectively. (a) The anatomical vertical meridians of the eyes are geometric corresponding points. When these points are projected into the world, they intersect at a vertical line in the head's mid-sagittal plane (here through fixation): this is the geometric vertical horopter. (b) The empirical vertical horopter has crossed disparity below fixation and uncrossed disparity above fixation, causing a top-back pitch.

The third column of Table 2.1 (marked θ_V) shows the measured angle between corresponding points near the vertical meridians in all published experiments that used the criterion of equal perceived-direction. The angle is positive in every case, consistent with corresponding points having crossed disparities in the lower visual field and uncrossed disparities in the upper field. However, the measured angle could be a consequence of cyclovergence, the disconjugate rotation of the eyes around the visual axes (Amigo, 1974). Cyclovergence causes equal rotations between the vertical and horizontal meridians, whereas the hypothesized shear of corresponding points should affect only the horizontal offsets between corresponding points near the vertical meridians. Therefore, to quantify the retinal shear angle, the cyclovergence angle must

be subtracted from the measured angle between corresponding points near the vertical meridians. Specifically,

$$\theta_r = \theta_v - \theta_h \qquad \text{Equation (2.1)}$$

where θ_r is the true retinal horizontal shear angle between corresponding points near the vertical meridians, θ_v is the measured angle, and θ_h is cyclovergence. (Obviously, if cyclovergence is zero, the true retinal shear and measured shear are equal.) The last column of Table 2.1 shows this adjustment for the studies in which both measurements were made. After adjustment, the shear angles are still all positive.

Citation	Subject	θ_v (deg)	θ_h (deg)	θ_r ($\theta_v - \theta_h$) (deg)
Helmholtz (1925)	HH[a]	2.66	0.3	2.36
	WV[b]	2.13	–	–
	WV[a]	2.15	–	–
	FS[b]	1.32	–	–
	FS[a]	1.44	–	–
Nakayama (1977)	AC	3.4	0.0	3.4
	CWT	4.8	0.0	4.8
Ledgeway and Rogers (1999)	TL	3.9	0.6	3.3
	BJR	5.8	1.3	4.5
	MLG	2.9	0.5	2.4
Siderov, Harwerth, and Bedell (1999)	AK[1]	0.56	–	–
	AK[2]	0.48	–	–
	HB[1]	0.49	–	–
	HB[2]	0.40	–	–
	LB[1]	0.62	–	–
	LB[2]	0.41	–	–
	MG[1]	0.32	–	–
	MG[2]	0.41	–	–
Grove, Kaneko, and Ono (2001)	PG	1.9	–	–
	HK	1.7	–	–
	NU	1.6	–	–
Schreiber et al. (2008)	PRM	2.8	0.0	2.8
	KMS	6.1	0.0	6.1
	HRF	3.6	0.0	3.6

Table 2.1. Previous studies of the shear between corresponding points. All of these studies used apparent-motion except: [a]binocular apparent vertical and horizontal; [b]monocular apparent vertical. For Siderov et al. (1999): [1]viewing distance = 200 cm; [2]viewing distance = 50 cm.

2.1.2 Empirical corresponding points and the ground plane

As illustrated in Figure 2.2, a positive shear between corresponding points near the vertical meridians would make the empirical vertical horopter parallel to the ground

plane when an observer's fixation is earth horizontal at infinity. Indeed, the horopter becomes coincident with the ground if the shear angle is

$$\theta_o = 2\tan^{-1}\left(\frac{I}{2h}\right) \qquad \text{Equation (2.2)}$$

where *I* is the observer's inter-ocular distance and *h* is the observer's eye height. We will call θ_o the optimal shear angle. With *I* = 6.5 cm and *h* = 160 cm, θ_o = 2.3° (Schreiber, Hillis, Fillipini, Schor, & Banks, 2008). The experimental measurements of the shear angle are reasonably consistent with this optimal value (Table 1). The similarity between observed and optimal shear angles suggests that the vertical horopter may be adaptive for making depth discriminations in the natural environment. Furthermore, for fixations on the ground in the head's sagittal plane, Listing's Law dictates that the horopter will remain coincident with the ground for an observer with an optimal shear value (Helmholtz, 1925; Schreiber et al., 2008). This is also illustrated in Figure 2.2.

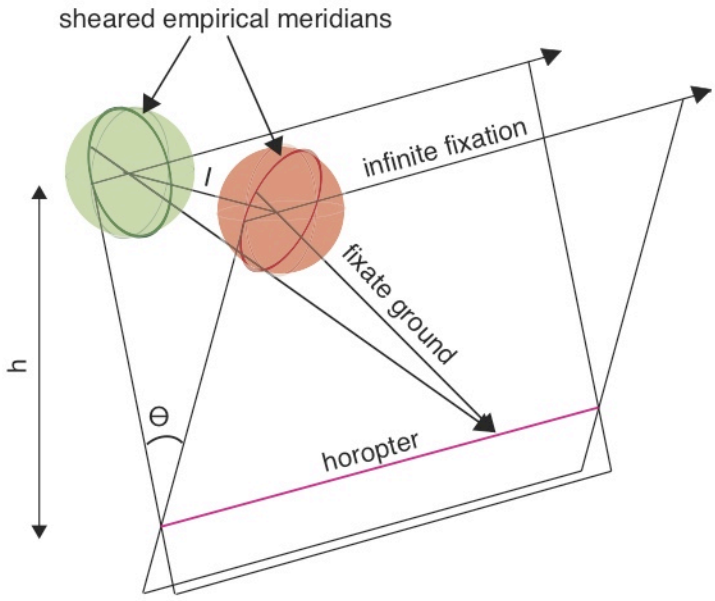

Figure 2.2. Vertical horopter and the ground. Green and red spheres represent the left and right eyes, respectively. The green and red circles represent the sheared empirical meridians associated with empirical corresponding points. When the eyes are fixated parallel to the ground at infinity, the vertical horopter is a horizontal line extending to meet fixation at infinity. For a given eye height and inter-ocular distance, the optimal shear angle places the horopter in the ground plane. Due to Listing's Law, the horopter remains in the ground when the eyes fixate the ground in the sagittal plane.

Cats and terrestrial owls are much shorter than humans, so by the above argument their optimal shear angles should be much larger than 2.3° (optimal shear angles for cats and owls are ~10.8° and ~10.6°, respectively). Indeed, physiological data indicate that their shear angles are near these optimal values (Cooper & Pettigrew, 1979), which are consistent with the hypothesis that the shear is adaptive for terrestrial species.

2.1.3 Hypotheses

Here, we examine two hypotheses about how the shear angle between corresponding points came to be: the adaptability hypothesis and the hard-coded hypothesis. The adaptability hypothesis is that an individual's shear angle is determined by his/her experience with the natural environment. According to this hypothesis, corresponding points adapt to optimize precision in depth estimation based on each individual's experience: If experience changes, the shear should change. The hard-coded hypothesis claims that the shear is hard-coded into the visual system because it confers an evolutionary advantage; that is, the shear is adaptive but not adaptable.

We performed a series of experiments to test these two hypotheses. While both hypotheses predict that the average shear angle in the population should be close to the average optimal value, the adaptability hypothesis makes the additional prediction that the shear should change with individual experience. We tested this prediction by determining whether observers with different inter-ocular distances and eye heights have different shear angles (Equation 2.2) and by determining whether an observer's shear angle changes when the experienced patterns of disparities are systematically altered by distorting lenses.

2.2 General methods for all experiments

In each experiment, we measured the locations of corresponding points using the apparent‑motion paradigm of Nakayama (1977). Our primary interest was to determine their locations near the eyes' vertical meridians, but we also measured their locations near the horizontal meridians so that we could subtract any contribution of cyclovergence.

2.2.1 Apparatus

Observers sat 114 cm from a large back-projection screen (61° wide by 51° high) and wore red–green anaglyph glasses. Display resolution was 1280 × 1024 pixels; each pixel subtended 3 arcmin. Observers were positioned and stabilized with a bite bar such

that the midpoint of their inter-ocular axis intersected a surface normal from the center of the screen. The room was dark except for the illuminated screen.

2.2.2 Stimulus and procedure

Observers were instructed to divergently fuse a pair of fixation targets that were presented at eye level and separated by the inter-ocular distance. Fusing these targets produced earth-horizontal fixation at infinity. To measure corresponding points accurately, it is essential to keep eye position constant across trials. The fixation targets were therefore constructed to allow observers to monitor their own fixation. The targets consisted of a radial pattern of 30-arcmin line segments (Schreiber et al., 2008). Some of the segments were presented to the left eye and some to the right (Figure 2.3a). By assessing the apparent vertical and horizontal alignments of the segments, observers could monitor horizontal and vertical vergence, respectively. We told observers to initiate trials only when the fixation targets were aligned and focused.

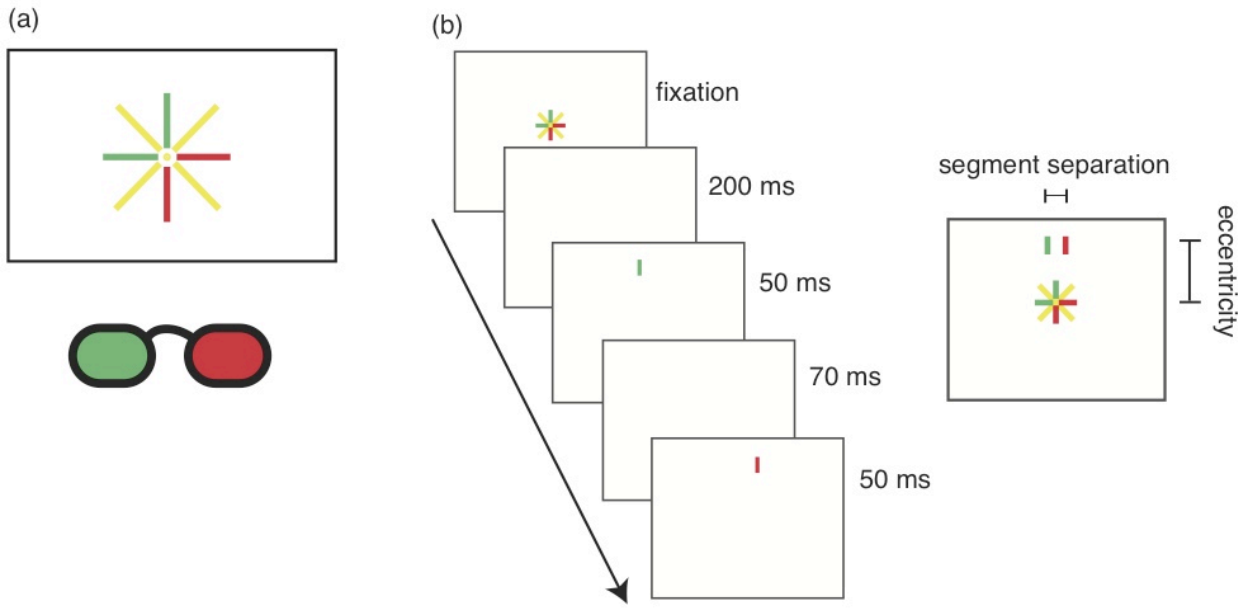

Figure 2.3. Stimulus and procedure. Observers wore red–green anaglyph glasses. Green, red, and yellow lines represent stimuli seen by the left, right, and both eyes, respectively. (a) The appearance of the fixation target when fused: perceived alignment of dichoptic vertical and horizontal segments indicated accurate horizontal and vertical vergences. Radial bioptic lines aided with maintenance of alignment. (b) Temporal sequence of screens for a trial and the resulting percept integrated over time.

For measurements near the vertical meridians, the experimental stimulus consisted of two dichoptic vertical line segments flashed in sequence. The presentation order between eyes was randomized. Each segment subtended 0.75° vertically and 3.4

arcmin horizontally. They were presented for 50 ms, with an inter-stimulus interval of 70 ms (Figure 2.3b). The line pairs had the same elevation but were displaced horizontally by equal and opposite amounts from the mid−sagittal plane. When the lines fell exactly on corresponding points, observers perceived no horizontal motion; otherwise, they appeared to move leftward or rightward. Line pairs appeared randomly at one of 14 vertical eccentricities from fixation (±2°, 3°, 4°, 5°, 6°, 7°, and 8°). By presenting stimuli at random eccentricities, we greatly reduced the usefulness of anticipatory eye movements. After each trial, observers made a forced-choice judgment of the direction of perceived motion. A 1-up/1-down adaptive staircase varied the horizontal separation between the lines at each eccentricity, with five step-size reductions and 14 reversals. Minimum step size was 1.7 arcmin. Data at each eccentricity were fit with a cumulative Gaussian using a maximum likelihood criterion (Wichmann & Hill, 2001). The mean of the best-fitting cumulative Gaussian (psychometric function) at each eccentricity was defined as the line segment separation that stimulated corresponding points. Figure 2.4 shows some of these fits for one observer.

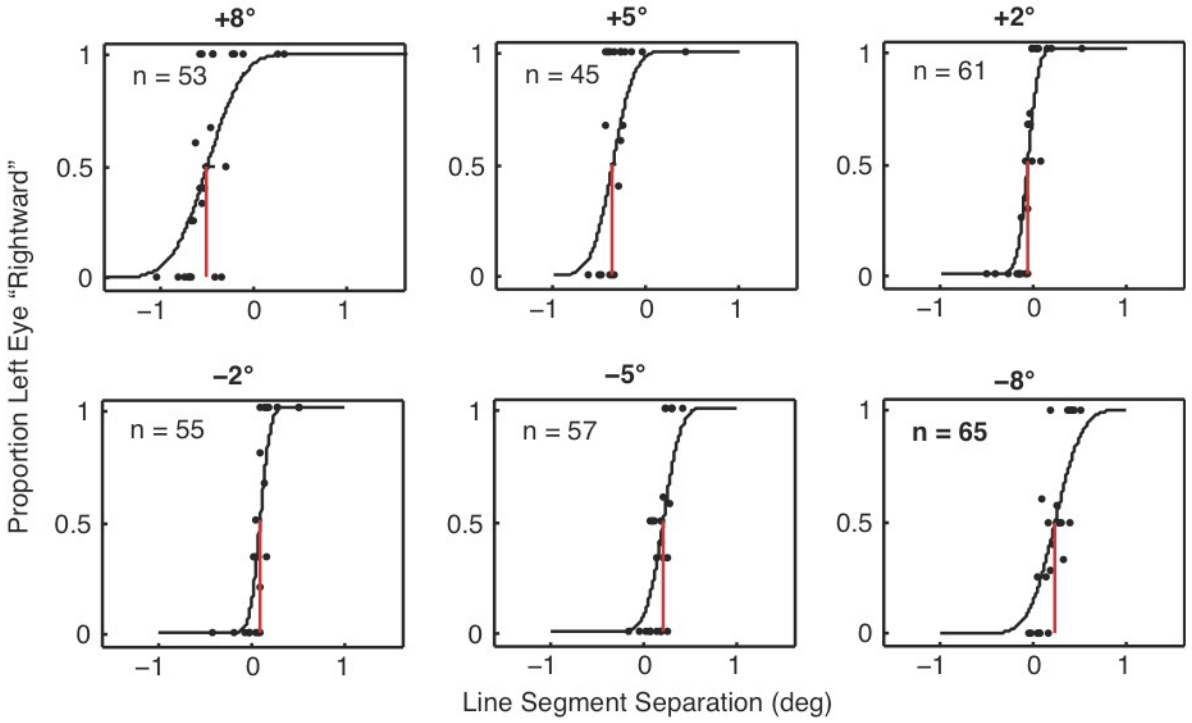

Figure 2.4. Psychometric functions for observer XMP. Each panel shows the responses for one vertical eccentricity (+8°, +5°, +2°, −2°, −5°, and −8°). The abscissas are the horizontal separation between the lines shown to the left and right eyes. Negative separations indicate uncrossed disparities, and positive separations indicate crossed disparities. The ordinates are the proportion of observer responses indicating that the line presented to the left eye was perceived to the right of the line presented to the right eye (i.e., indicating the lines had crossed disparity). The data were fit with cumulative Gaussians. The means of the Gaussians (indicated by the red vertical lines) were defined as the disparities between corresponding points at each eccentricity.

The horizontal separations of the points obtained from the Gaussian fits were plotted as a function of eccentricity (Figure 2.5a). We fit the resulting data with two lines via weighted linear regression. We defined the angle between the vertical meridians (θ_v) as the angle between the best-fit regression lines for the left and right eyes. Azimuth and elevation are plotted in Hess coordinates, a spherical coordinate system in which azimuth and elevation are both measured along major circles (i.e., longitudes). Lines in Cartesian coordinates project to major circles in spherical coordinates, and major circles are plotted as lines in Hess coordinates. Therefore, lines in the world map to lines in Hess coordinates. Using this coordinate system enabled us to readily assess whether the empirical horopter could lie in a plane.

We measured cyclovergence while observers performed the main experimental task. We did so by presenting dichoptic horizontal line segments near the horizontal meridians. The lines were displaced vertically and observers indicated whether apparent-motion was upward or downward (see Experiment 3 for validation of this method). The procedure was otherwise the same as the one we used to estimate corresponding points near the vertical meridians. For these measurements of vertical offsets, 14 additional staircases were randomly interspersed during a session with the other measurements.

Figure 2.5b plots the separations of line segments near the horizontal meridians that yielded no apparent-motion. We fit these data with regression lines and the angle between these lines was our estimate of cyclovergence. We then used these estimates (θ_h) to correct the measurements near the vertical meridians (θ_v) and thereby obtain an estimate of the retinal shear angle (θ_r; Equation 2.2). This is shown in Figure 2.5c.

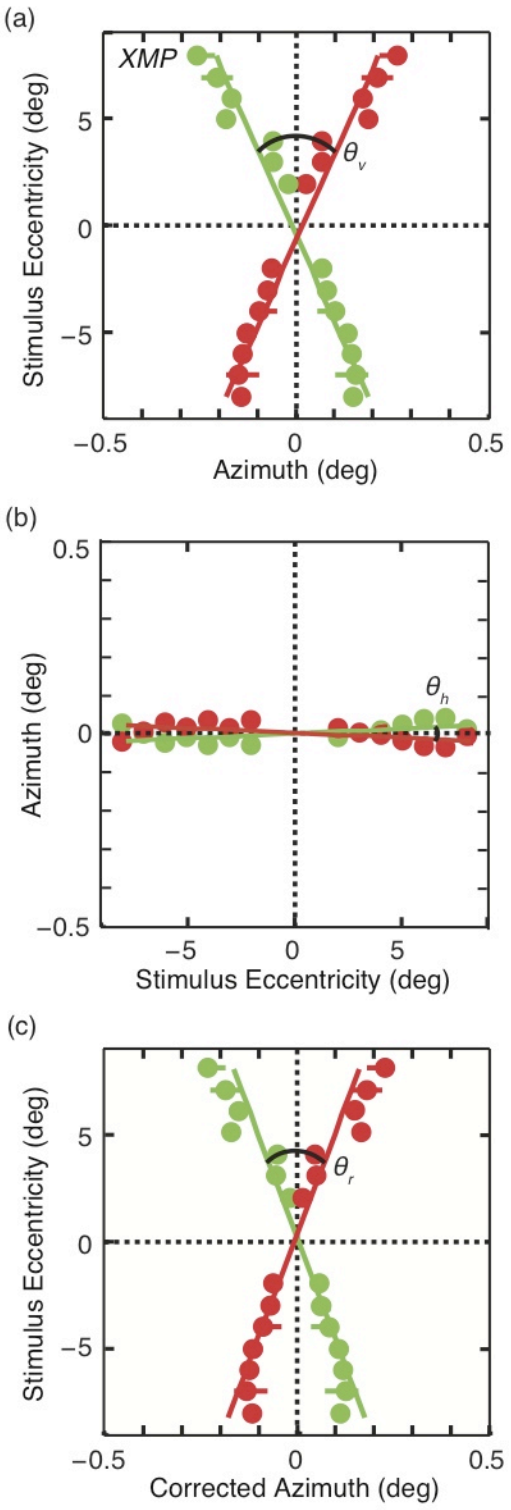

Figure 2.5. Data from one observer. The data in each panel are plotted in Hess coordinates. (a) Measurements near the vertical meridians before correction for cyclovergence. The abscissa is the horizontal line segment separation for which no motion was perceived. The ordinate is vertical eccentricity. The scale of the abscissa is expanded relative to the scale of the ordinate. At each

eccentricity, the green and red dots indicate the measured locations in the left and right eyes, respectively. Error bars represent 95% confidence intervals. The angle between the regression lines is the angle between the measured positions (θ_v). (b) Measured positions of corresponding points near the horizontal meridians. The abscissa is the vertical line separation for which no motion was seen. The ordinate is horizontal eccentricity. The angle (θ_h) is presumed to be due to cyclovergence. (c) Retinal positions of corresponding points near the vertical meridians once corrected for cyclovergence (θ_r; Equation 2.2).

2.3 Experiment 1: Adaptiveness of corresponding points

In Experiment 1, we measured the horizontal shear angle in observers with different inter-ocular distances and eye heights. The adaptability hypothesis predicts a positive correlation between observers' measured retinal shear (θ_r) and their optimal shear (θ_o from Equation 2.2). The hard-coded hypothesis predicts no correlation.

2.3.1 Methods

We recruited 39 observers with a range of optimal shear angles. It was impractical to recruit people based on their inter-ocular distance (most people do not know it), so we recruited people of various heights. Consequently, the population included members of the San Francisco Bay Area Chapter of the Little People of America and members of college basketball and crew teams. Their overall heights ranged from 4.3 to 7.0 ft (129.5 to 213.4 cm).

Eleven observers were excluded because they had reduced stereoacuity, significant exophoria or esophoria, or because they were unable to perform the task (staircases did not converge). One observer was an author; the others were unaware of the experimental hypotheses. All underwent training prior to data collection.

Figure 2.6a is a scatter plot of inter-ocular distances and eye heights. The two values were not significantly correlated ($r = 0.25$, $p = 0.2$, $df = 26$), so our population had a reasonably wide range of optimal shear angles: 1.5° − 2.8° (mean = 2.1°). Figure 2.6b is a histogram of these optimal values.

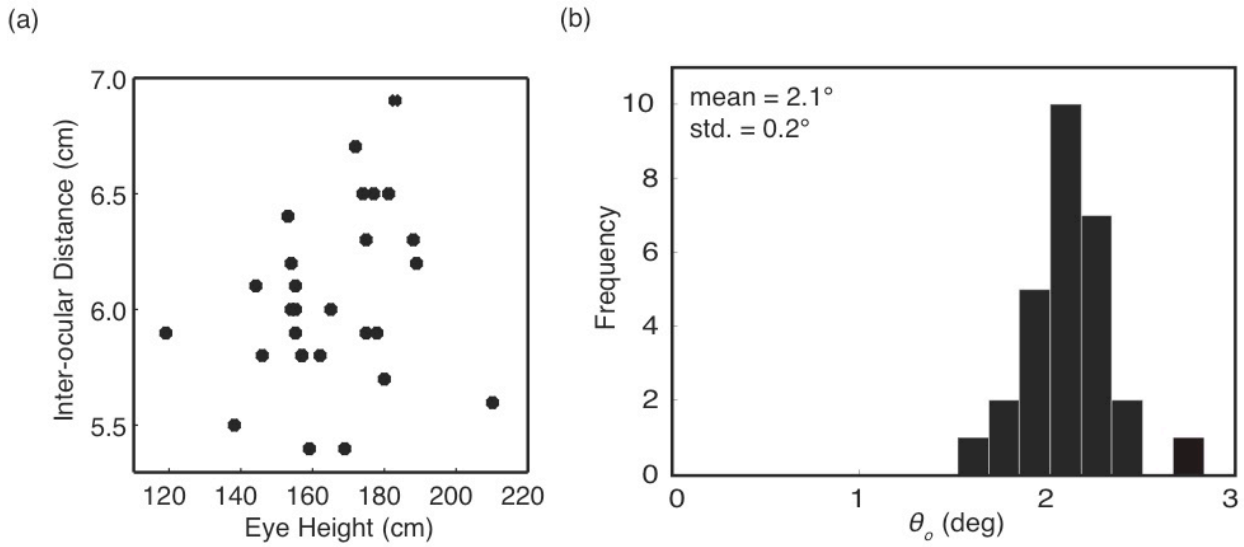

Figure 2.6. Eye height, inter-ocular distance, and optimal shear (θ_o) angles for the observer population. (a) Scatter plot of eye heights and inter-ocular distances. (b) Histogram showing the distribution of optimal shear angles for the sample population. The mean and standard deviation are shown in the upper left.

2.3.2 Results

Figure 2.7 plots the positions of corresponding points near the vertical meridians for all observers. Offsets due to fixation disparity were first eliminated from the data by shifting the data from the two eyes horizontally until they intersected at zero. Because no measurements were taken at 0° elevation, the amount of shift was determined by finding the x-intercept of a regression line fit to the data (we used quadratic regression lines because much of the data was poorly fit by lines). Rotations due to cyclovergence (θ_h) were also subtracted as shown in Figure 2.5 (Equation 2.1). In agreement with the previous literature, all but two observers had corresponding points with uncrossed disparity above fixation and crossed disparity below fixation. (Observer KKD had uncrossed disparity above but no clear pattern of disparity below fixation; LAT had no clear pattern at all.) This means that the vertical horopters of nearly all observers are pitched top-back.

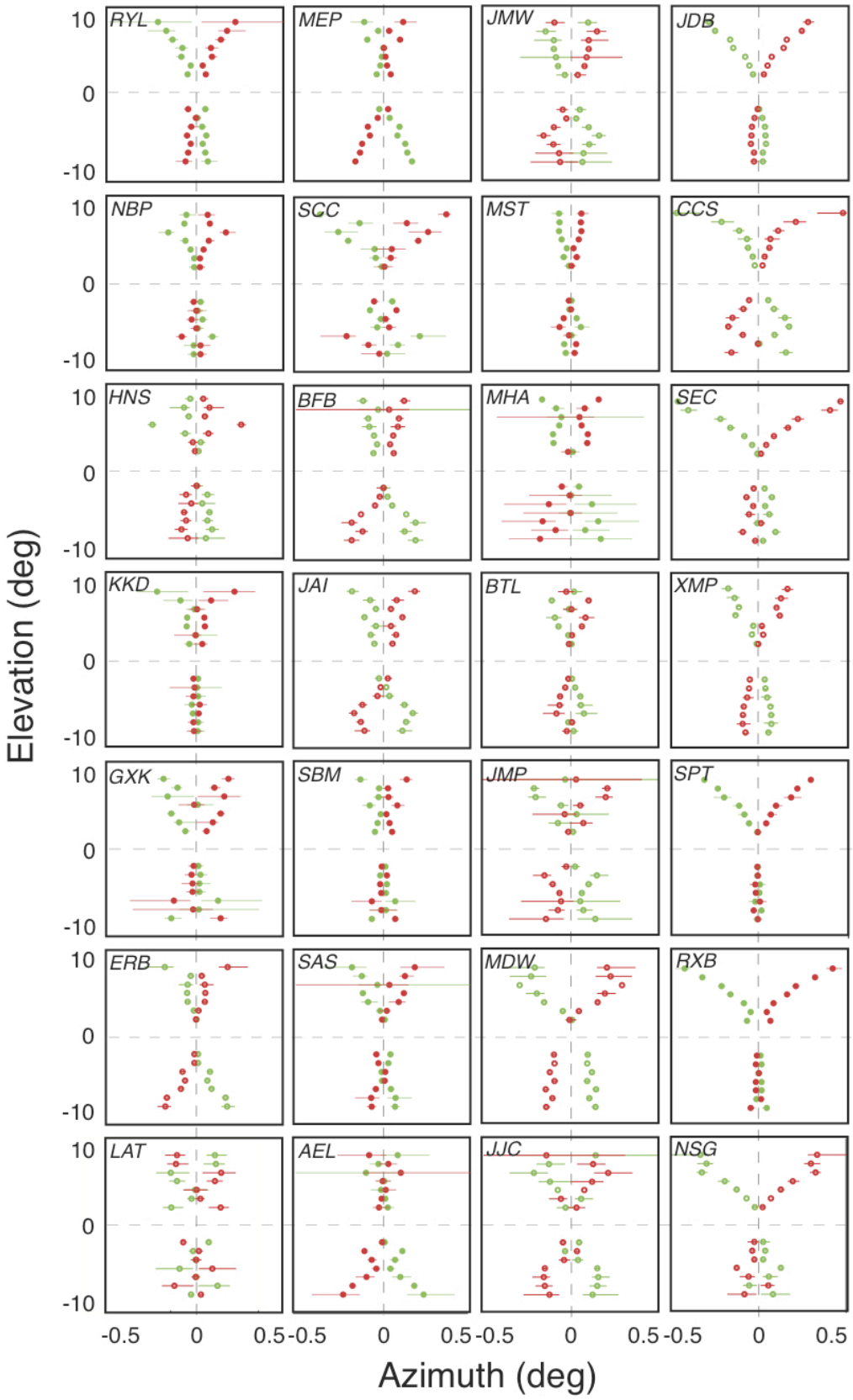
RYL
MEP
JMW
JDB
NBP
SCC
MST
CCS
HNS
BFB
MHA
SEC
KKD
JAI
BTL
XMP
GXK
SBM
JMP
SPT
ERB
SAS
MDW
RXB
LAT
AEL
JJC
NSG
10
0
-10
-0.5
0
0.5
Elevation (deg)
Azimuth (deg)

Figure 2.7. Locations of corresponding points for all observers. The abscissa is the azimuth and the ordinate is the vertical eccentricity. The scale of the abscissa is expanded relative to the scale of the ordinate. The data are plotted in Hess coordinates, a system in which lines in the world map to lines in azimuth–elevation plots. At each eccentricity, the green and red dots indicate the locations of corresponding points in the left and right eyes, respectively. Error bars are 95% confidence intervals for azimuth measurements. Rotations due to cyclovergence (θ_h) and offsets due to fixation disparity were first subtracted from the data. We corrected for fixation disparity by finding the abscissa value at which the regression lines intersected, and then shifting the data horizontally such that the intersection had an ordinate value of zero.

Figure 2.8a is a histogram of the measured retinal shear values (θ_r). The mean shear angle was 1.6° and the standard deviation was 0.8°. Figure 2.8b plots each observer's measured shear value against their optimal shear value. The two values were not significantly correlated ($r(26) = 0.07$, $p = 0.72$). The non-significant correlation between measured and optimal shear suggests that corresponding points are not adjusted to keep the horopter in the ground plane for individuals. This is counter to a prediction of the adaptability hypothesis. However, the average measured value was similar to the average optimal value (2.1°), which is consistent with the hypothesis that the shear is hard-coded to be adaptive for the population in general.

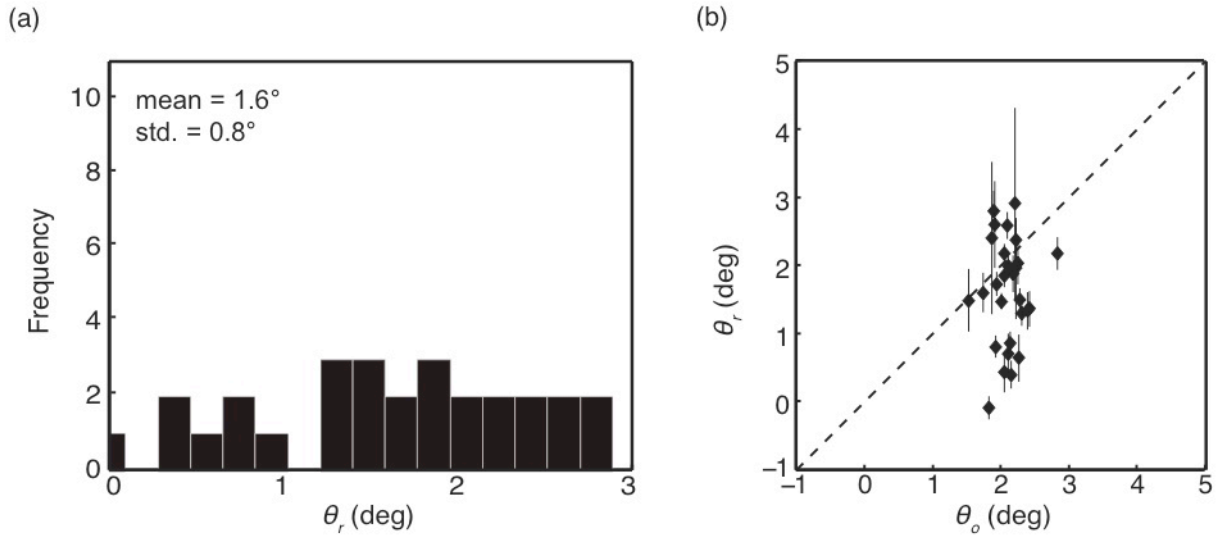

Figure 2.8. Retinal shear angle in individual observers. (a) Histogram showing distribution of measured retinal shear angles after correction for cyclovergence. The mean and standard deviation are shown in the upper left. (b) Scatter plot of optimal and measured retinal shear angles. The abscissa and ordinate are the optimal and measured angles (again after correction), respectively. Error bars are standard errors, determined by bootstrapping.

In the Hess coordinates we used, lines in the world map to lines in the corresponding point plots. Thus, if the vertical horopter lies in a plane, the data should

be well fit by lines. Figure 2.7 reveals that the data are generally not well fit by lines; this is particularly evident in the far right column. These data are best fit by curves with centers that are bent toward zero azimuth. Such convex patterns generate convex vertical horopters (i.e., relative to a slanted line, it is farther from the observer above and below fixation). To illustrate this, Figure 2.9 shows a side view of the horopter for two observers who are fixating on the ground plane at 0.5 and 1.5 m ahead. Eye position for declined gaze was determined using Listing's Law, so the horopter should be coincident with the ground if the shear is optimal (Equation 2.2). Observer JAI had a reasonably linear pattern of corresponding points, and the horopter is therefore approximately linear (left panel). Because the shear angle is close to the optimal value, the horopter is also approximately coincident with the ground plane. Observer SEC had a convex pattern of corresponding points, and the horopter is therefore convex and not coincident with the ground (right panel). The prevalence of convex correspondence patterns in our data is thus inconsistent with the original hypothesis that the pattern of corresponding points is an adaptation to the ground plane (Helmholtz, 1925; Schreiber et al., 2008).

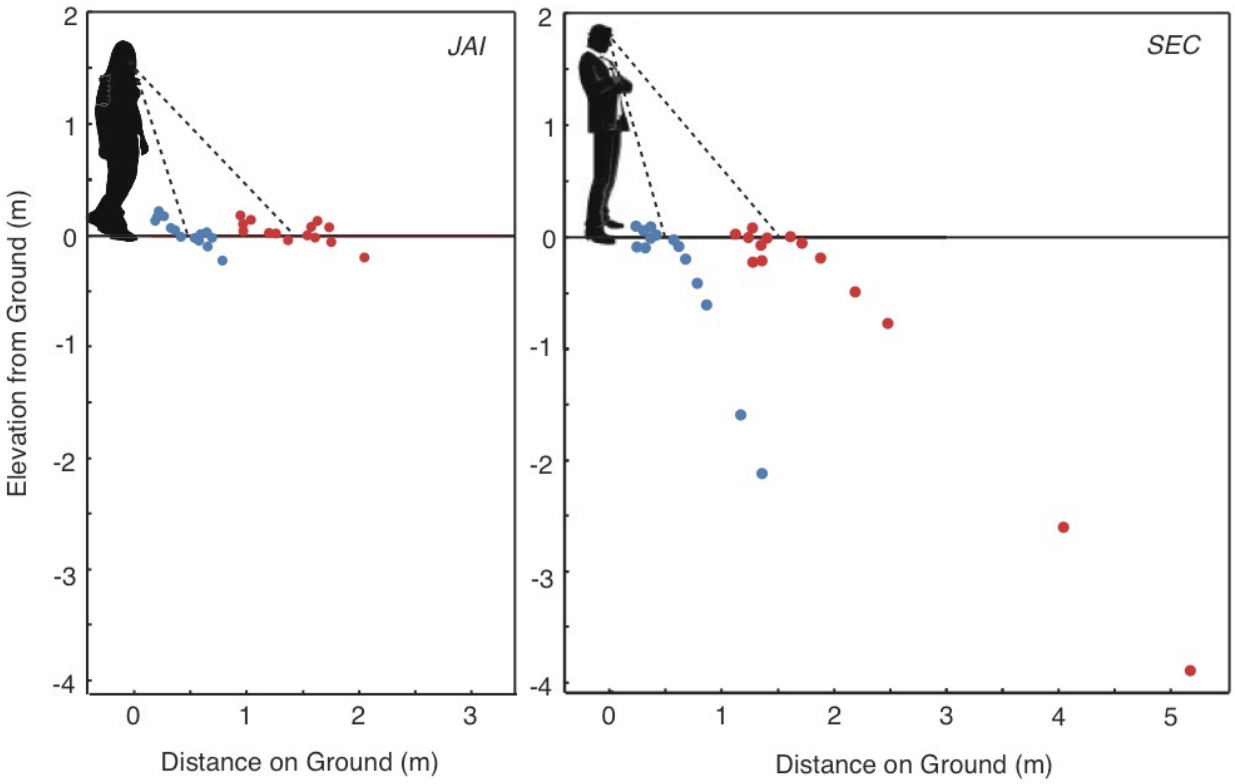

Figure 2.9. Vertical horopters for two observers. The abscissa is the distance along the ground from the observer's feet. The ordinate is elevation relative to the ground. Two gaze positions are shown with fixation on the ground at 0.5 and 1.5 m in the mid-sagittal plane. In calculating the horopters, eye position was consistent with Listing's Law. (Left) The vertical horopter for observer JAI with linear corresponding point data. This horopter coincides roughly with the ground. (Right) The vertical horopter for SEC with curved corresponding point data. The horopter is convex and therefore not coincident with the ground plane.

To check that the observation of curved correspondence patterns was not caused by a procedural or computational error, we also measured the curvature of corresponding points near the horizontal meridians. Along these meridians, the correspondence pattern should not be curved because non-zero vertical disparities in the apparent-motion task would presumably manifest non-zero cyclovergence. To test this prediction, we compared quadratic regressions of the measurements near the vertical and horizontal meridians for all observers. A two-tailed t-test also revealed that the coefficients on the quadratic terms for the fits near the horizontal meridians were not significantly different from zero (mean = 0.001, df = 26, p = 0.496), which means, as we expected, that there is no curvature in the pattern of correspondence near the horizontal meridians. In contrast, the coefficients for the fits near the vertical meridians were significantly less than zero (mean = −0.009, df = 26, p = 0.004), consistent with convex horopters.

2.4 Experiment 2: Adaptability of corresponding points

In Experiment 1, we observed that the horizontal shears for individual observers were not well correlated with the optimal shears (Equation 2.2). This result is inconsistent with the adaptability hypothesis. However, our calculation of the optimal shear assumes that the retinal shear is specifically adaptive for the pattern of disparities cast by the ground plane at standing height for fixations in the mid-sagittal plane. Perhaps this specific shear value does not reflect the majority of an observer's experience with disparities in the natural environment. Thus, to further test the adaptability hypothesis, we systematically sheared the disparities delivered to the eyes for seven consecutive days using distorting lenses. If the horopter is adaptable, the shear angle between corresponding points should change in the direction of the shear added by the lenses. If it is hard-coded, no change in shear should occur.

2.4.1 Methods

Five observers participated. All had corrected-to-normal vision and normal stereoacuity, and all underwent training prior to data collection. Two were authors; the others were unaware of the experimental hypotheses.

For all waking hours of seven consecutive days, observers wore lenses over the two eyes that systematically altered the pattern of incident disparities. The lenses were afocal unilateral magnifiers (Ogle & Ellerbrock, 1946); they magnify the image along one axis, and not the orthogonal axis, without introducing an astigmatic difference in focal power. The lenses were situated in a frame with their principal axes rotated by ±1.5°. They created equal and opposite shears of the images of a vertical line (±3°), and those

shears were opposite to the ones for a horizontal line (±3°). Our goal was to create only a horizontal shearing of vertical lines, but the additional vertical shearing of horizontal lines was a necessary byproduct of creating lenses that did not cause defocus. The overall effect is illustrated in Figure 2.10.

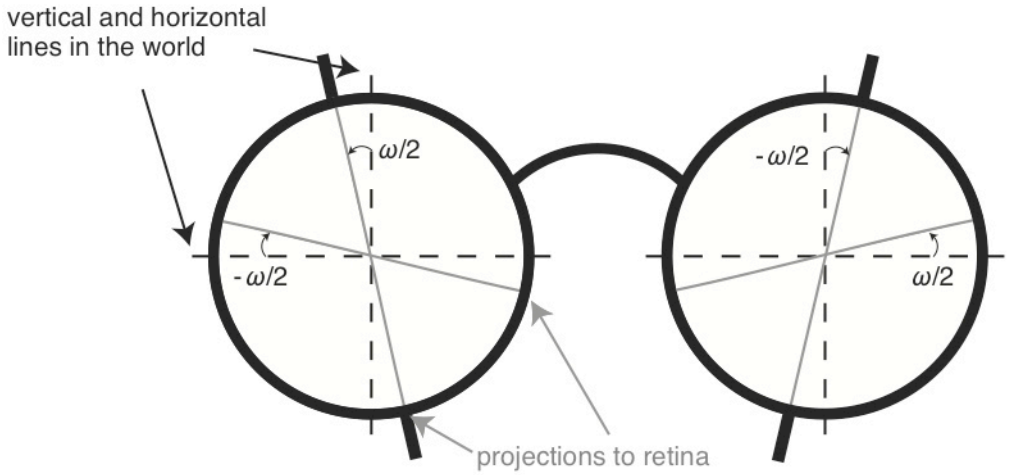

Figure 2.10. Distorting lenses worn by observers in Experiment 2. Each lens was an afocal unilateral magnifier. The lenses were rotated in opposite directions in the two eyes by ±ω/2°, creating a horizontal shear disparity of ω° in the projection of vertical lines. The lenses also created a vertical shear disparity of −ω° in the projection of horizontal lines.

Three observers wore lenses that created a horizontal shear of +3° (extorsion) and two wore lenses creating a shear of −3° (intorsion). When the lenses were initially put on, frontoparallel surfaces appeared to be pitched top-back (extorting lenses) or top-forward (intorting lenses), as expected from the geometry of the viewing situation (Ogle & Ellerbrock, 1946).

We measured the patterns of corresponding points before, during, and after wearing the lenses. The equipment, stimulus, and procedure were identical to Experiment 1. After an initial training session, subjects came in 24 hours prior to putting on the lenses and performed one measurement. There were eight measurements taken during lens wear because one measurement was taken immediately after putting the lenses on, and then measurements were taken at approximately 24 hour intervals for the next 7 days. The first post-lens measurement was taken immediately after the lenses were removed on the seventh day and the next measurement was taken 24 hours later. As before, we used measurements of vertical disparities near the horizontal meridians to measure cyclovergence and used those measurements to estimate the retinal shear near the vertical meridians.

2.4.2 Results

To determine how much the retinal shear angle (θ_r) changed in response to the distorting lenses, we had to take into account the effects of the optical shear caused by the lenses (because observers wore them during the experimental measurements) and of cyclovergence. To take the optical shear into account, we subtracted the horizontal shear due to the lenses (i.e., +3° or −3°) from the empirical measurements. To take the cyclovergence into account, we subtracted the measured cyclovergence values as in Experiment 1. We found that cyclovergence changed slightly during lens wear: The average increase was 0.5° for observers wearing lenses with ω = +3° and −0.1° for those wearing lenses with ω = −3°. Vertical shear disparity along the horizontal meridians induces cyclovergence (Crone & Everhard-Halm, 1975), so the change in cyclovergence we observed was surely due to the vertical disparities of the lenses near the horizontal meridians.

Figure 2.11 shows the corrected retinal shear angle before, during, and after wearing the lenses. There was no systematic change in the retinal shear angle between the vertical meridians for any of the observers. We conclude that corresponding points near the vertical meridians do not adapt in response to a 7-day change in the disparities delivered to the eyes. Consequently, the vertical horopter does not adapt to visual input, at least over the course of a week. This finding, coupled with the observation of no correlation between observer traits (eye height and eye separation) and retinal shear (Figure 2.8), implies that the vertical horopter is not adaptable.

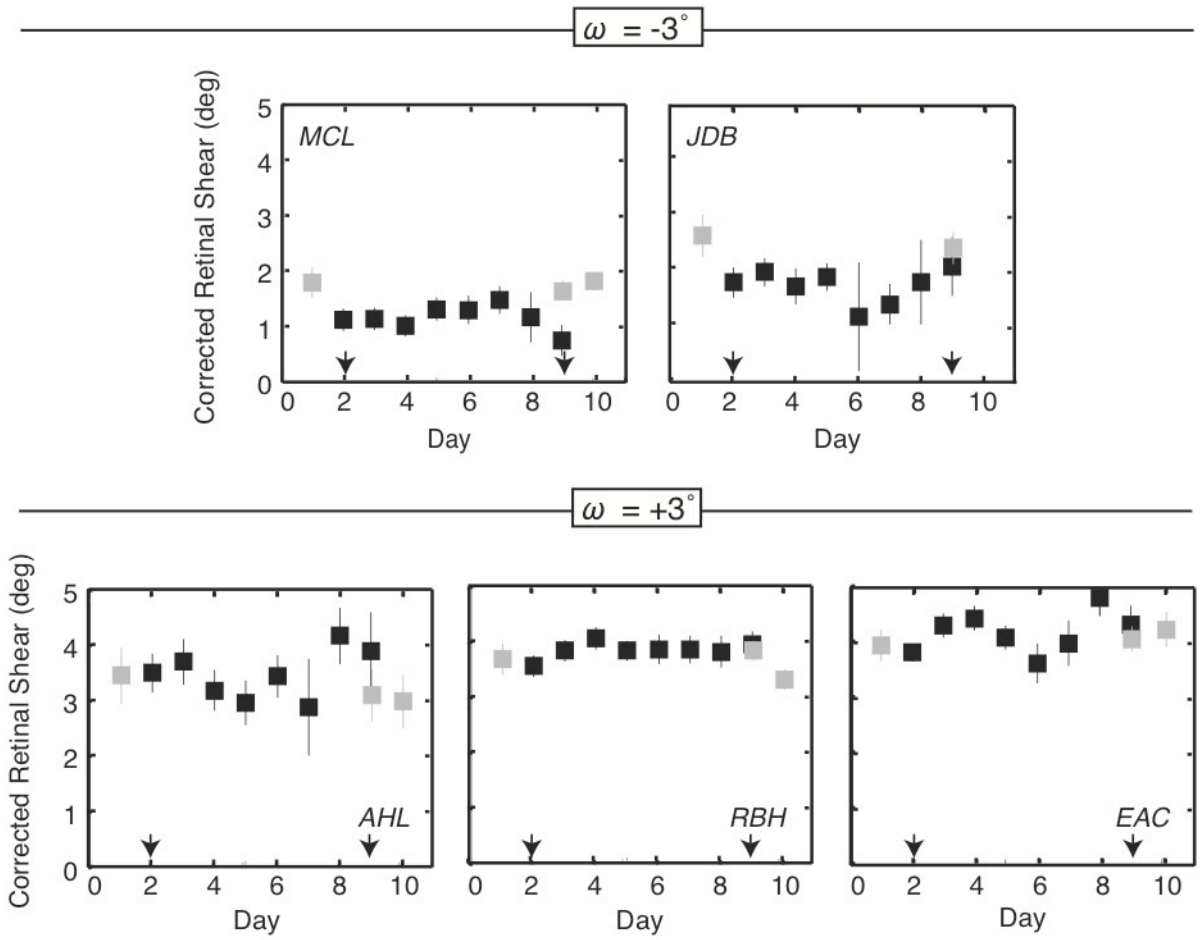

Figure 2.11. Retinal shear angle before, during, and after wearing the distorting lenses. Corrected retinal shear angle is plotted for each day. The angle has been corrected for cyclovergence and for the shearing effects of the lenses. Thus, the plotted values reflect changes in the retinal positions of corresponding points. One measurement was made before putting the lenses on (gray squares), eight were made while the lenses were on (black squares), and two were made after the lenses were taken off (gray squares). The arrows indicate the days the lenses were first put on and first taken off. Error bars are 95% confidence intervals obtained by bootstrap.

Despite no change in retinal corresponding points, all five observers experienced perceptual adaptation. They reported that the world appeared distorted when they first wore the lenses (e.g., depending on the sign of ω, frontoparallel surfaces appeared slanted top-back or top-forward, and perceived height was increased or decreased, respectively). However, after 5 days of lens wear, everyone reported that the world appeared undistorted. Four of them also reported perceptual distortions in the opposite direction when the lenses were removed. Closing an eye eliminated the perceptual aftereffect, which suggests that the perceptual effects were due to changes in the interpretation of binocular information and not due to changes in monocular shape representation. However, additional tests would be necessary to confirm this.

2.5 Experiment 3: Methods for measuring cyclovergence

Experiments 3 and 4 are control experiments conducted to test the assumptions made about cyclovergence in Experiments 1 and 2. In Experiment 3, we assessed the validity of our method for measuring cyclovergence. This subjective method has been frequently used (Banks, Hooge, & Backus, 2001; Ledgeway & Rogers, 1999; Nakayama, 1977) but rarely compared to an objective measurement (but see Crone & Everhard-Halm, 1975; Howard, Ohmi, & Sun, 1993). To assess validity, we measured cyclovergence in the same observers at the same time using the subjective method and an objective measurement of eye position.

2.5.1 Methods

Four observers participated; all were unaware of the experimental hypotheses. They performed the subjective task (described earlier) while the torsional position of both eyes was measured using an eye tracker. Stimuli were projected on 70° × 70° screen 100 cm from the midpoint of the inter-ocular axis. Pixels subtended 3.5 arcmin. The eye tracker was an infrared video-oculography system (SensoMotoric Instruments, Teltow, Germany). It captures 60-Hz video of the pupils and irises and determines cyclovergence by measuring the relative rotation of the irises.

At the beginning of each session, observers fixated a dichoptic target that created stable earth-horizontal fixation at infinity. The center of the target was identical to the fixation target used in the previous experiments. A larger stabilizing stimulus was added to aid alignment of the horizontal meridians of the eyes. This stimulus was a pattern of six 40°-long bioptic (identical in both eyes) radial line segments. After fusing this stimulus for 1 min, eye position was recorded and used as the reference position for the eye tracker. Next, the observer's eyes were induced to change cyclovergence. To induce such changes, we presented a pattern of 40°-long horizontal lines that were rotated in opposite directions for the two eyes by ±3°, ±2°, ±1°, and 0° (Crone & Everhard-Halm, 1975). This new stimulus was also viewed for 1 min. We then began the eye tracking and subjective task. To maintain cyclovergence, the inducing stimulus was presented for 3 s between each apparent-motion trial.

We used the data from the eye tracker as the objective measure of cyclovergence and the data from the apparent-motion task as the subjective measure. The time series data from the eye tracker were thresholded by dropping measurements with reliabilities less than 75%. These reliabilities are calculated automatically by the eye tracker for each time point and reflect the agreement between the current image of the iris and the initial calibration image (Pansell, Schworm, & Ygge, 2003). Time stamps from the eye tracker and the apparent-motion stimulus were used to select the eye-tracking data obtained within ±100 ms of the middle of each apparent-motion trial. Only

eye-tracking data obtained within these time windows were included for analysis. The apparent-motion data were analyzed as before to obtain the subjective measure of cyclovergence.

2.5.2 Results

Figure 2.12 plots the objective and subjective estimates of cyclovergence as a function of the rotation of the cyclovergence stimulus. Each panel shows an individual observer's data. The agreement between the two measures was excellent, which validates the subjective method for measuring cyclovergence.

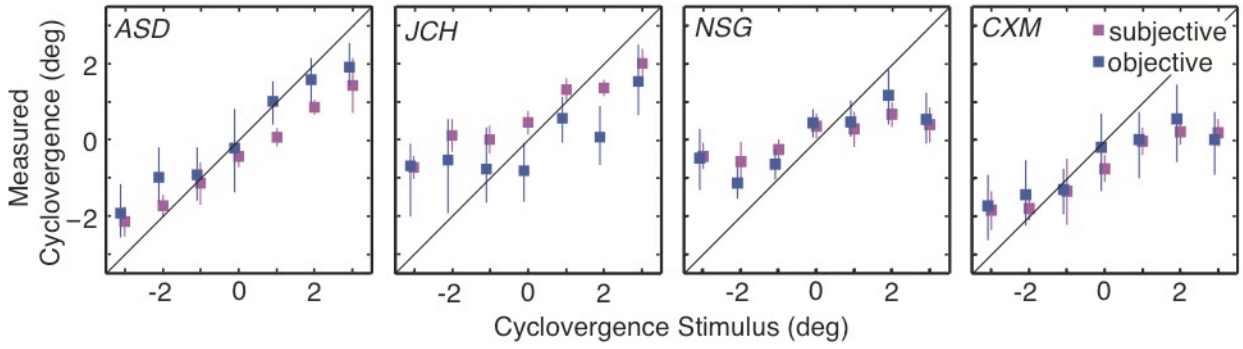

Figure 2.12. Subjective and objective estimates of cyclovergence. Each panel shows the data from an individual observer. The abscissa is the angular rotation of the inducing lines and the ordinate is the cyclovergence measured with the apparent-motion task (red) and the eye tracker (blue). Error bars are 95% confidence intervals.

2.6 Experiment 4: Cyclovergence in natural viewing

We were concerned about how to properly correct for cyclovergence when estimating the shear angle between corresponding points. Previous studies subtracted cyclovergence from the measured shear in order to estimate the retinal shear angle (Equation 2.1 and Table 2.1). However, what if cyclovergence under natural viewing conditions—such as gazing at the horizon while standing upright—is not zero? In that case, subtracting cyclovergence would not yield an estimate of the surface that stimulates corresponding points in natural viewing. To do this, we need to know the cyclovergence of the eyes in natural viewing and then subtract that value from the measured shear. To this end, we next measured cyclovergence for upright observers when a binocular stimulus simulating a floor and hallway was present. We then compared that value to the cyclovergence when only a dichoptic fixation target was present (as in Experiments 1 and 2).

2.6.1 Methods

Five of the original 28 observers participated. We measured cyclovergence using the apparent-motion task. As before, fixation was earth horizontal at infinity. In one condition, the only visible stimuli were the dichoptic fixation target and the flashed lines used in the apparent-motion task (Figure 2.3); the room was otherwise completely dark. In a second condition, we added a random-dot stereogram that simulated the walls and floor of a hallway. The slant of the simulated floor was adjusted for each individual observer to be the same as the slant of a ground plane viewed while standing upright. The conditions were presented in random order.

2.6.2 Results

The average cyclovergence angles in the first (the one with only the fixation target) and second (the one with the floor and walls added) conditions were 0.50° (standard error = 0.14°) and 0.03° (standard error = 0.10°), respectively. These estimates were significantly different (pairwise t-test; df = 8, p = 0.01). They were significantly greater than 0° in the first condition (blank) but not in the second condition (hallway). Figure 2.13 shows cyclovergence angles for individual subjects in both conditions. Assuming that the second condition is more representative of natural viewing than the first condition, these results illustrate the need to measure and correct cyclovergence to zero in Experiments 1 and 2. Thus, our correction procedure was justified.

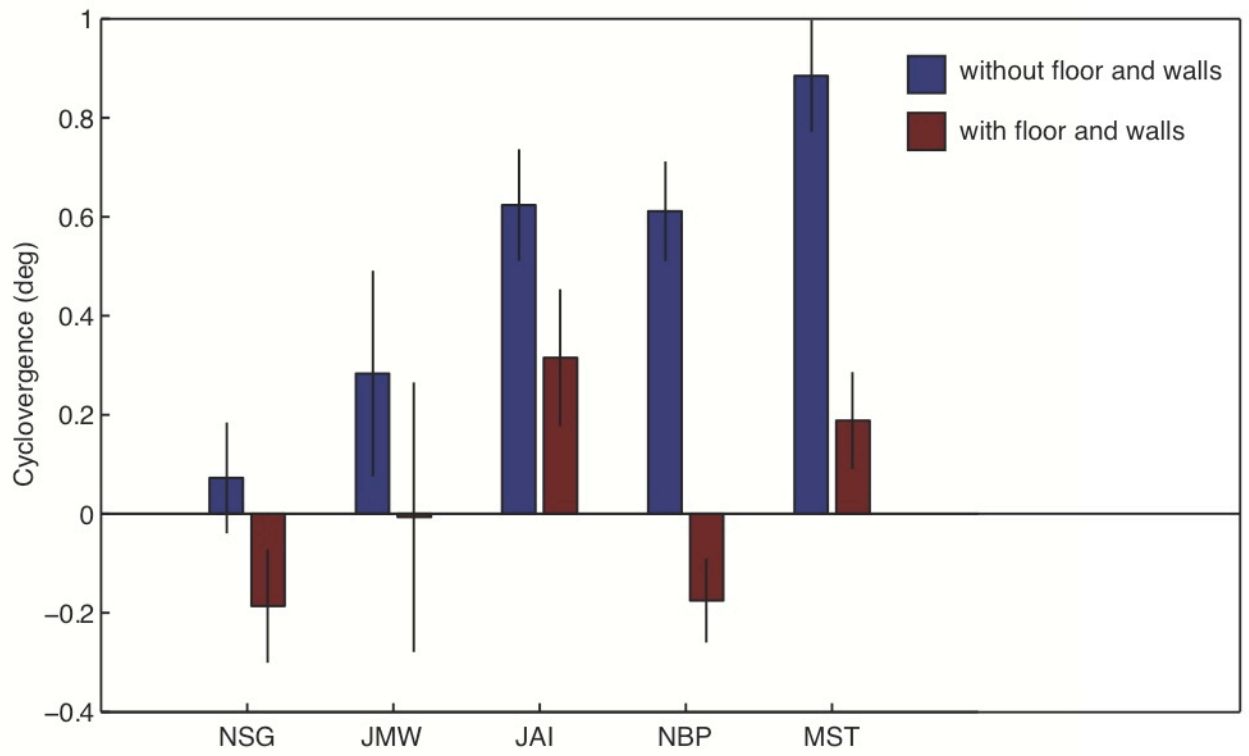

Figure 2.13. Cyclovergence angles for each subject when viewing a screen with only a fixation target (blue), or with a fixation target, and simulated floor and walls (red). Bars indicate cyclovergence angle, errors bars are standard deviation.

2.7 Discussion

2.7.1 Summary

We found no evidence that the horopter is adaptable: There was no correlation between an observer's eye height/separation and their retinal shear angle and no effect on corresponding point locations after wearing distorting lenses for a week. While null effects can be difficult to interpret, we can conclude that there is no strong effect of experience on the vertical horopter, at least in adults.

We did find that the shear was positive in all but one observer, and that the average shear angle of our observers was similar to the predicted optimal shear angle based on eye height/separation. This is consistent with the hypothesis that the location of the horopter is hard-coded because it is adaptive. However, we also found that many observers had patterns of retinal correspondence that yielded convex horopters. Curved horopters are inconsistent with the hypothesis that the horopter is especially adaptive for making depth discriminations relative to the ground plane. Perhaps the shear is an adaptation to some other common situation in natural viewing.

2.7.2 Convexity

With the exception of Amigo (1974), all previous papers on the vertical horopter have described the deviation of empirical corresponding points from geometric points as a horizontal shear, i.e., horizontal offsets proportional to elevation. We will refer to this correspondence pattern as linear because the offsets can be fit with lines. As noted earlier, we observed systematic deviations from linearity in most observers (Figure 2.7).

It turns out that this deviation has been observed before. Figure 2.14 plots the data from the three previous studies of the vertical horopter that used the criterion of perceived-direction and that reported data for individual observers. Non-linear correspondence patterns are quite evident in some observers (PG, PRM, KMS) and perhaps present in others (CWT, AC, NU). Importantly, whenever a deviation from linearity occurs (in our data and theirs), it is always convex (i.e., centers bent toward zero azimuth). Convex patterns of correspondence are evidently common.

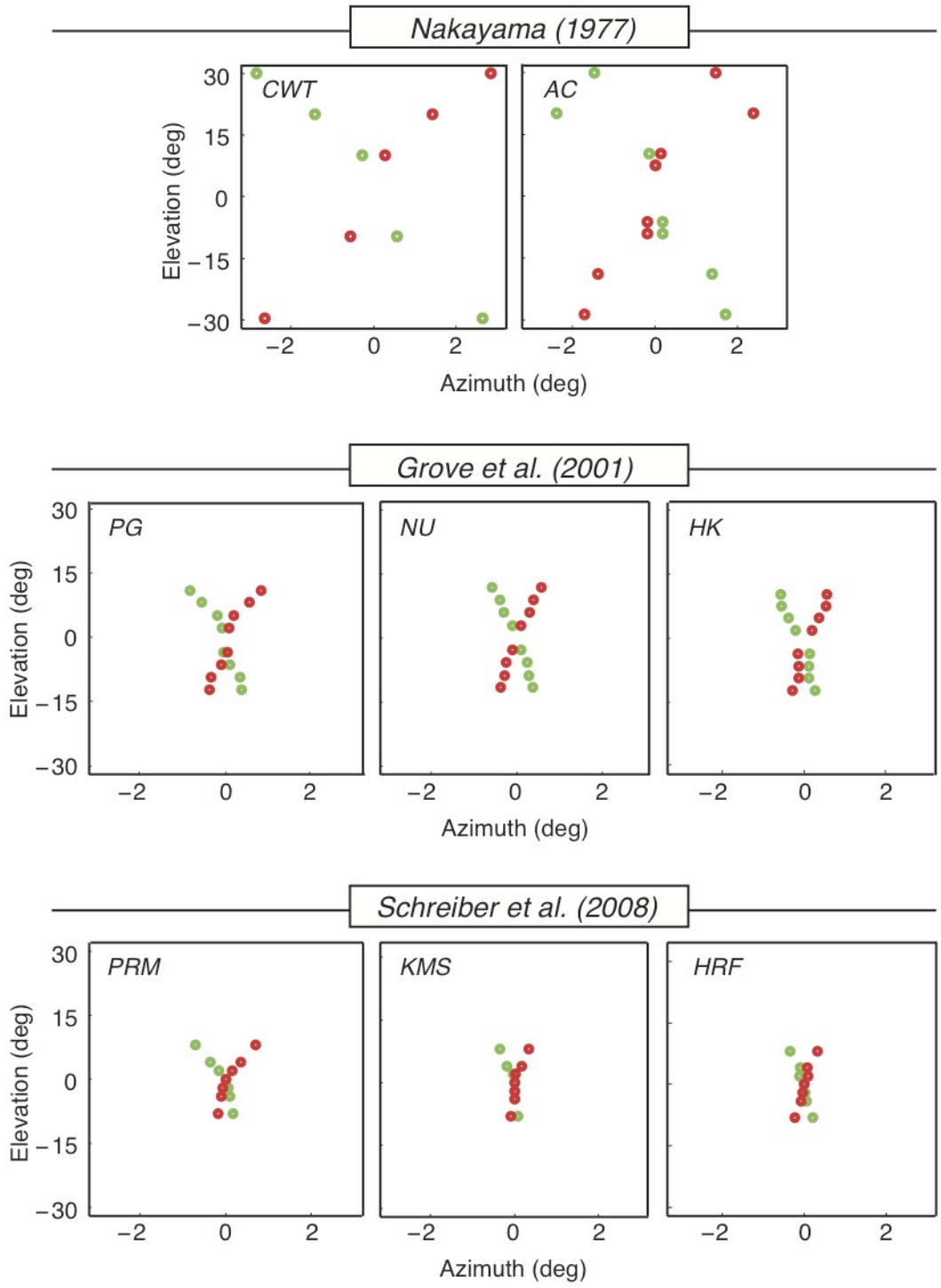

Figure 2.14. Patterns of corresponding points from previous studies. The data in each panel are plotted in Hess coordinates. The abscissa is the azimuth of the line segments for which no motion is perceived. The ordinate is vertical eccentricity. The scale of the abscissa is expanded relative to that of the ordinate. The green and red dots indicate the measured locations of corresponding points in the left and right eyes, respectively.

In the Hess coordinates we used, convex patterns of corresponding points yield convex vertical horopters. Such horopters cannot be coincident with the ground plane. The observed convexity is thus inconsistent with the theory that the vertical horopter manifests an adaptation for depth perception relative to the ground. Perhaps the horopter is adaptive for a different property of the natural environment.

There is good evidence that the visual system has an expectation, or Bayesian prior, for convex shapes (Langer & Bülthoff, 2001; Liu & Todd, 2004; O'Shea, Agrawala, & Banks, 2010; Sun & Perona, 1998). Such an expectation makes sense because most objects are mostly convex. Perhaps the convexity of the vertical horopter is an adaptation for the most likely shape of surfaces in the natural environment. To evaluate this idea, we need to also examine the shape of the horizontal horopter. The geometric horizontal horopter is the Vieth–Müller Circle: the circle containing the fixation point and the nodal points of the eyes. Figure 2.15a shows a plan view of the geometric horizontal horopter. As shown in Figure 2.15b, the empirical horizontal horopter is less concave than the geometric horopter, and the Hering–Hillebrand deviation (*H*) quantifies the difference:

$$H = \cot(\alpha_L) - \cot(\alpha_R) \qquad \text{Equation (2.3)}$$

where α_L and α_R are the angular locations of corresponding points along the horizontal meridians in the left and right eyes, respectively. Note that the empirical and geometric horizontal horopters are the same when $H = 0$. Table 2.2 shows the *H* values reported from several previous studies that used the perceived‑direction criterion. *H* is always greater than zero except for observer HRF in Schreiber et al. (2008) and she has intermittent strabismus.

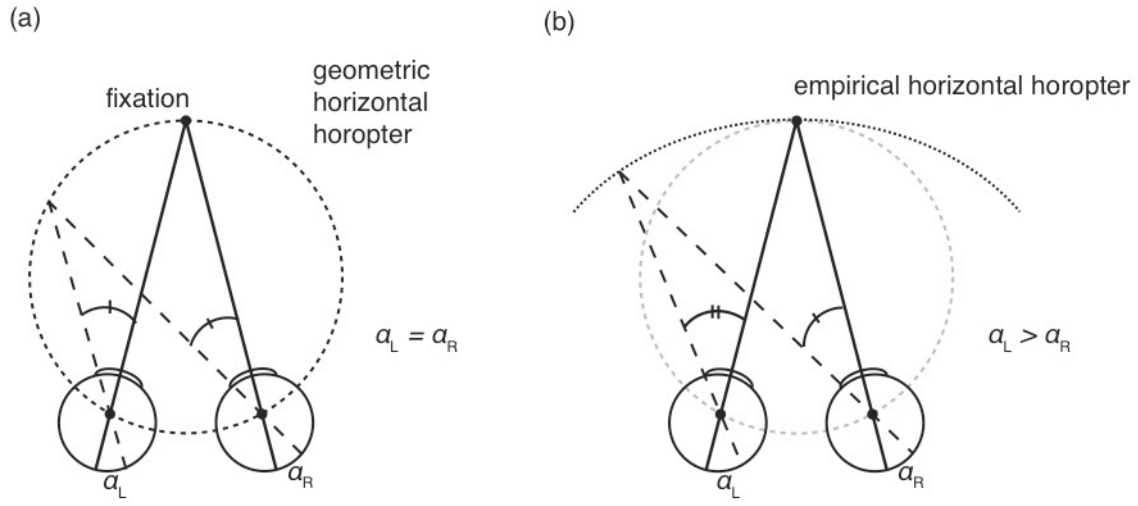

Figure 2.15. Geometric and empirical horizontal horopters. Circles represent the left and right eyes that are fixating a point in the mid-sagittal plane. (a) Geometric corresponding points along the horizontal meridians have equal horizontal offsets in the eyes: $\alpha_L = \alpha_R$. When these points are projected into the world, they intersect at a circle containing the fixation point and the nodal points of the eyes. This is the geometric horizontal horopter, or Vieth–Müller Circle. (b) Empirical corresponding points along the horizontal meridians have unequal offsets: $\alpha_L > \alpha_R$. This is quantified by the Hering–Hillenbrand deviation (H). The deviation is such that the empirical horizontal horopter is less concave than the geometric horopter.

Paper	Observer	Distance (cm)	H
von Liebermann (1910)*		97	0.04
Lau (1921)*		150	0.10
Helmholtz (1925)*		71	0.07
Ogle (1950)	KNO	76	0.08
	AA	76	0.05
	FDC	76	0.05
	WH	60	0.04
Amigo (1967)	PDL	67	0.13
	GA	67	0.07
Hillis and Banks (2001)	MSB	172	0.24
	JMH	172	0.22
	ND	172	0.43
Schreiber et al. (2008)	PRM	40	0.25
	KMS	40	0.36
	HRF	40	−0.11

Table 2.2. H values (Hering–Hillebrand deviation) from previous studies. Note: *Values obtained from Ogle (1950). When values were given for various fixation distances, the farthest distance was used.

We next compared the shapes of the vertical and horizontal horopters and determined how those shapes changed with viewing distance. For this analysis, we used our measurements of corresponding points near the vertical meridians and data from the literature for corresponding point data near the horizontal meridians (Table 2.2). First, we calculated average horopters from the vertical and horizontal data. To do this, we found the average offset between corresponding points at each retinal eccentricity (between −8 and 8°) and determined the locations in space that would stimulate those points. Figure 2.16a shows top and side views of the average horizontal and vertical horopters, respectively, for three fixation distances (0.5, 1, and 2 m). Fixation is earth horizontal in the mid-sagittal plane. Note that the horizontal horopter is approximately planar at the near distance and becomes increasingly convex at greater distances. The vertical horopter is pitched top-back and is convex at all distances, but the pitch and convexity increase with distance. Next, we quantified the changes in horopter curvature as a function of fixation distance. We plotted each observer's horopter as a function of eccentricity (azimuth for the horizontal horopter and elevation for the vertical). Then, we fit a second-order polynomial to those data and calculated the

second derivative ($\partial^2 Z/\partial E^2$). Figure 2.16b plots the second derivative as a function of fixation distance for the vertical and horizontal horopters. Positive and negative values indicate convex and concave shapes, respectively. Greater magnitudes indicate greater curvature. As expected, the vertical horopter is convex at all distances and becomes increasingly so with increasing distance. The horizontal horopter is concave at near distance (see inset where the second derivative is less than zero), becomes planar at approximately 0.46 m (the abathic distance), and becomes increasingly convex at greater distances. For comparison, we also plot the results of the same analysis for a basketball (men's size 7, radius = 11.8 cm). We used a cross-section of the basketball to determine the osculating circle for a parabola and calculated the second derivative. The basketball is more convex than the horizontal horopter at all plotted distances (at sufficiently great distance, the horopter becomes more convex than the ball). The basketball is more convex than the vertical horopter for distances less than 2 m and less convex than the horopter at greater distances.

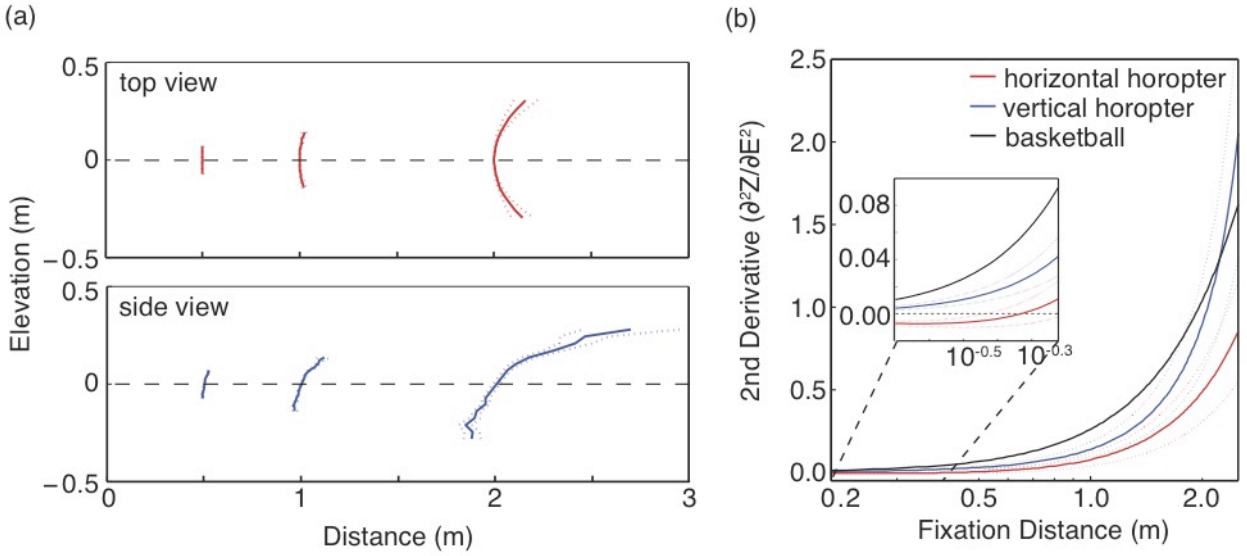

Figure 2.16. The shapes of the vertical and horizontal horopters as a function of viewing distance. (a) The upper panel is a top view of the empirical horizontal horopter at three fixation distances (0.5, 1, and 2 m). Fixation is earth horizontal in the mid-sagittal plane (gaze direction is indicated by the dashed horizontal lines). In each case, the observer is fixated (that is, converged) where the surface has an ordinate value (distance parallel to inter-ocular axis) of 0. The *H* value (Hering–Hillebrand deviations) used to generate these horopters is 0.13, the average of the values in Table 2.2. Standard error is indicated by the dotted lines. The lower panel is a side view of the empirical vertical horopter at the same three fixation distances. Again, the observer is fixated where the surface has an ordinate value (distance perpendicular to inter-ocular axis) of 0. The positions of corresponding points used to generate these horopters are the average values across our observers. Standard error is indicated by the dotted lines. (b) Curvature of the vertical and horizontal horopters as a function of fixation distance. The average vertical and horizontal horopters were fit with second-order polynomials. The second derivative of distance as a function of eccentricity ($\partial^2 Z/\partial E^2$), where *Z* is distance and *E* is elevation for the vertical horopter and azimuth for the horizontal horopter, of those fitted functions is plotted as a function of fixation distance. The inset is a magnified view of the values for short fixation distances; the ordinate values have been magnified more than the abscissa

values. In the main plot and the inset, the values for the horizontal and vertical horopters are indicated by the red and blue curves, respectively. We also used a basketball as an osculating circle to determine a parabola and then computed the second derivative as a function of viewing distance. Those values are indicated by the black curves.

This analysis shows that corresponding points are best suited for surfaces that are generally convex and pitched top-back. It would be quite interesting to see if such surfaces are commonplace in natural scenes, particularly surfaces at relative close range where stereopsis is precise.

2.7.3 Subjective measurement of cyclovergence

We estimated cyclovergence in our experiments by measuring the perceived offsets between dichoptic horizontal line segments (i.e., a nonius task). Two previous studies compared subjective (nonius) and objective (eye tracking) estimates of cyclovergence. Howard et al. (1993) compared subjective estimates from vertical nonius lines presented above and below the foveae to objective estimates obtained with scleral search coils. The two estimates did not yield the same values, so they concluded that subjective measurements do not provide an accurate measure of cyclovergence. Crone and Everhard-Halm (1975) compared estimates from horizontal nonius lines slightly above and below the foveae to the locations of ocular blood vessels. They observed close agreement between the two estimates and concluded that subjective methods do allow one to estimate cyclovergence. Unfortunately, Crone and Everhard-Halm made few measurements, so their data were not very convincing.

We propose that the disparity between these two reports stems from the difference in the orientation and location of the nonius lines. The perceived alignment of vertical lines above and below fixation will be affected by both cyclovergence and the shear of retinal corresponding points above and below the foveae. This is not true for horizontal lines near the horizontal meridians because corresponding points in those retinal regions have the same anatomical elevations in the two eyes; as a consequence, perceived misalignment of horizontal lines is caused by cyclovergence alone. The results of our control experiment confirm that horizontal nonius lines presented near the horizontal meridians provide an accurate estimate of cyclovergence across a wide range of eye positions. Thus, horizontal lines to the left and right of fixation should be used when estimating cyclovergence subjectively.

2.7.4 Natural situations in which corresponding points cannot be stimulated

The top-back pitch and curvature of the vertical horopter may be adaptive for short viewing distances, but they are not beneficial for long ones. Here, we show that the pitch and curvature of the horopter preclude the stimulation of corresponding points in the upper visual field at long viewing distances.

In natural viewing, there can never be greater uncrossed disparities than the disparities created by light rays that are parallel to one another (i.e., coming from infinite distance). Because the corresponding points above fixation have uncrossed disparity, there is a fixation distance beyond which those points could never be stimulated by the natural environment. We calculated these critical fixation distances for each retinal eccentricity. In the left panel of Figure 2.17, a binocular observer fixates a point in the head's mid-sagittal plane at distance Z_0 while a point at distance Z_1 stimulates the retina at locations α_L and α_R relative to the foveae. The horizontal disparity due to Z_1 is the difference in those locations. The horizontal disparity in radians is given by

$$\delta = \frac{(Z_1 - Z_0)I}{Z_1 Z_0}, \qquad \text{Equation (2.4)}$$

where I is the inter-ocular separation (Held, Cooper, O'Brien, & Banks, 2010). Rearranging, we obtain

$$Z_1 = \frac{Z_0 I}{I - Z_0 \delta}. \qquad \text{Equation (2.5)}$$

This is the object distance that is associated with a given disparity and fixation distance. Those distances are plotted in the right panel of Figure 2.17. Blue and red curves correspond to combinations of fixation distances (Z_0) and object distances (Z_1) for positive (uncrossed) disparities and negative (crossed) disparities, respectively; the disparities have been converted to degrees. For each positive disparity, there is a greatest fixation distance Z_0 at which it is possible for that disparity to arise from the natural environment. That greatest distance is I/δ (Equation 2.5 for $Z_1 = \infty$). For disparities of +0.1° and +1.0°, the greatest fixation distances are 34.4 and 3.44 m, respectively (indicated by arrows in the figure). Greater distances could not possibly give rise to the observed disparity.

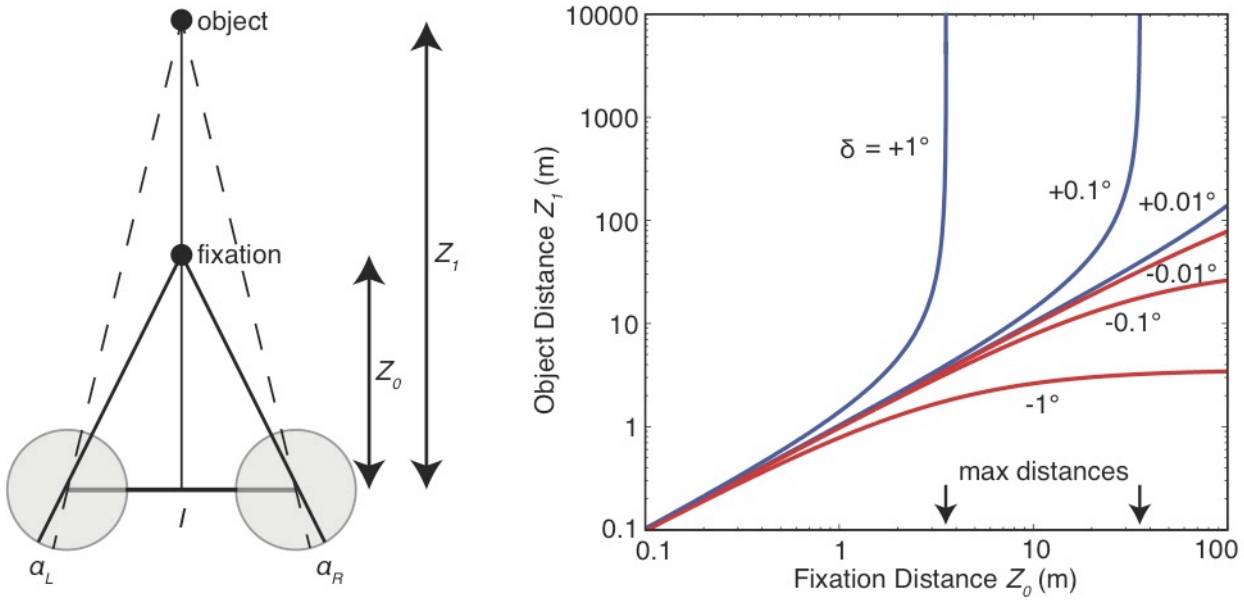

Figure 2.17. The set of distances that can yield disparities of different values. (Left) The viewing geometry. A binocular observer with inter-ocular distance *I* fixates a point at distance Z_0. Another object at distance Z_1 stimulates the left and right retinas at locations α_L and α_R, respectively, creating horizontal disparity δ. (Right) Given disparity δ, different combinations of fixation and object distance are possible. Object distance Z_1 is plotted as a function of fixation distance Z_0, both in meters. The blue curves are the fixation–object combinations that can occur with positive (uncrossed) disparities and the red curves are the combinations that can occur with negative (crossed) disparities. Each curve is labeled with the specific disparity value, expressed in degrees. We assumed an inter-ocular distance *I* of 0.06 m.

Corresponding points nearly always have uncrossed disparity above fixation. The analysis in Figure 2.17 shows that for each retinal eccentricity, there is a fixation distance beyond which real stimuli cannot stimulate corresponding points. Figure 2.18 shows those distances as a function of eccentricity in the upper visual field. We conclude there are many natural viewing situations in which corresponding points in the upper visual field cannot possibly be stimulated. (This situation does not generally occur in the lower visual field because corresponding points there almost always have crossed disparity.) Because disparity-based depth discrimination is most precise when corresponding points are stimulated, the precision of depth perception is compromised in such viewing situations.

There are many reasons that stereopsis is not well suited for long viewing distance (Howard & Rogers, 2002). The fact that real scenes cannot stimulate corresponding points in a large portion of the visual field when the viewer fixates more than a few meters away is yet another.

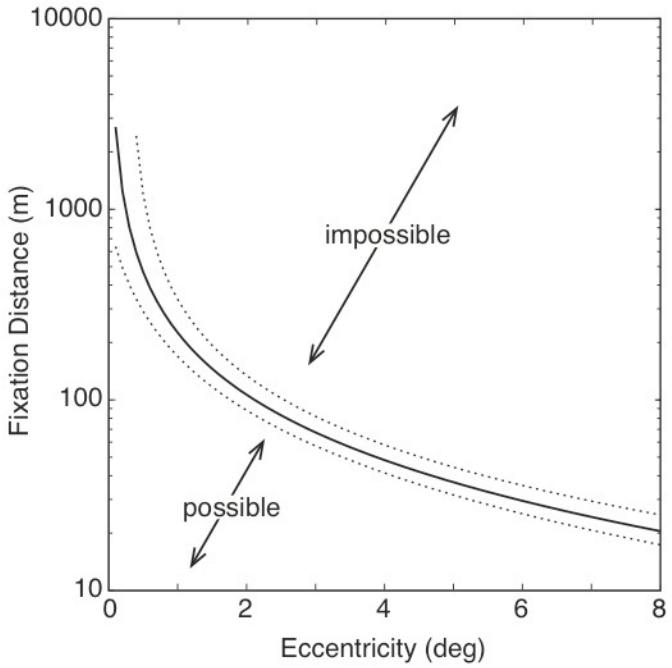

Figure 2.18. The set of distances and retinal eccentricities for which corresponding points in the upper visual field cannot be stimulated by real stimuli. The critical fixation distance is plotted as a function of eccentricity in the upper visual field. We used the average correspondence pattern across our observers to determine the disparity of corresponding points at each retinal eccentricity. The greatest fixation distance at which such disparity could occur is I/δ, where I is inter-ocular distance (assumed to be 0.06 m) and δ is disparity in radians. The dashed lines represent standard errors at each eccentricity calculated from the curves that fit each observer's data. The average correspondence pattern was used to generate this figure. Therefore, the impossible viewing distances apply to the average observer, not necessarily to an individual observer.

2.7.5 Conclusion

This chapter described a series of experiments and analyses examining how one aspect of human vision—the vertical horopter—might be adapted to the natural environment. We specifically tested the hypothesis that the vertical horopter is adapted to improve the precision of depth estimates relative to the ground plane. We confirmed that the vertical horopter is pitched top-back, like the ground in the visual field, but we also found that it is typically convex. We observed no change in the horopter after wearing distorting lenses, suggesting that it is not adaptable by experience. We also showed that the horopter is not adaptive for long viewing distances because at such distances uncrossed disparities between corresponding points cannot be stimulated. Taken together, these findings do not support the original hypothesis that the vertical horopter is specifically adaptive for making depth estimates relative to the ground plane. Instead, our findings suggest that the vertical horopter may be adaptive for perceiving

slightly convex, slanted surfaces at short distances. Are such surfaces commonly viewed in the natural environment? Future work should try to answer this question. Such work would need to measure not only the 3D structure of the environment, but also the vergence and version of binocular gaze. One could then reconstruct the disparities that are most common in natural viewing. Previous studies have looked at similar questions, but they used either virtual environments or simulated rather than actual fixations (Hibbard & Bouzit, 2005; Liu, Bovik & Cormack, 2008). Ideally, a study would include data about both the 3D structure and the 3D fixations within the same environment. If the pattern of disparities observed in the vertical horopter is adaptive, we would expect to find that this pattern of disparities is commonly encountered in the natural environment.

Chapter 3. Perception of 3D shape from linear perspective in pictures

3.1 Introduction

Every day the Associated Press publishes ~3,000 new photographs, and Facebook users post nearly 250 million (Associated Press, 2009; Shaffer, 2011). Clearly, people rely on pictures to communicate with one another, so it is important to understand how people perceive pictures in typical viewing situations.

Photographers, cinematographers, and computer-graphics engineers create pictorial effects in various ways. For example, photographs of scenes captured with short-focal-length lenses appear expanded in depth, while those captured with long lenses appear compressed. These effects can be seen in still photographs and video. In the latter, the technique creating the effect is called a “dolly-zoom” shot. Figure 3.1a shows two example photographs from the website Flickr. On the left, the goat looks stretched in depth; on the right, the pitcher and batter appear to be much closer to one another than they actually are. Figure 3.1b shows two example frames from a dolly-zoom shot in the movie Goodfellas. Objects through the window appear much farther away in the frame on the left (from early in the scene) than in the frame on the right (from later). Figure 3.1c shows how depth compression and expansion can also affect the appearance of a face. Long lenses can make a person look smarter, more attractive, and less approachable; short lenses have the opposite effect (Perona, 2007).

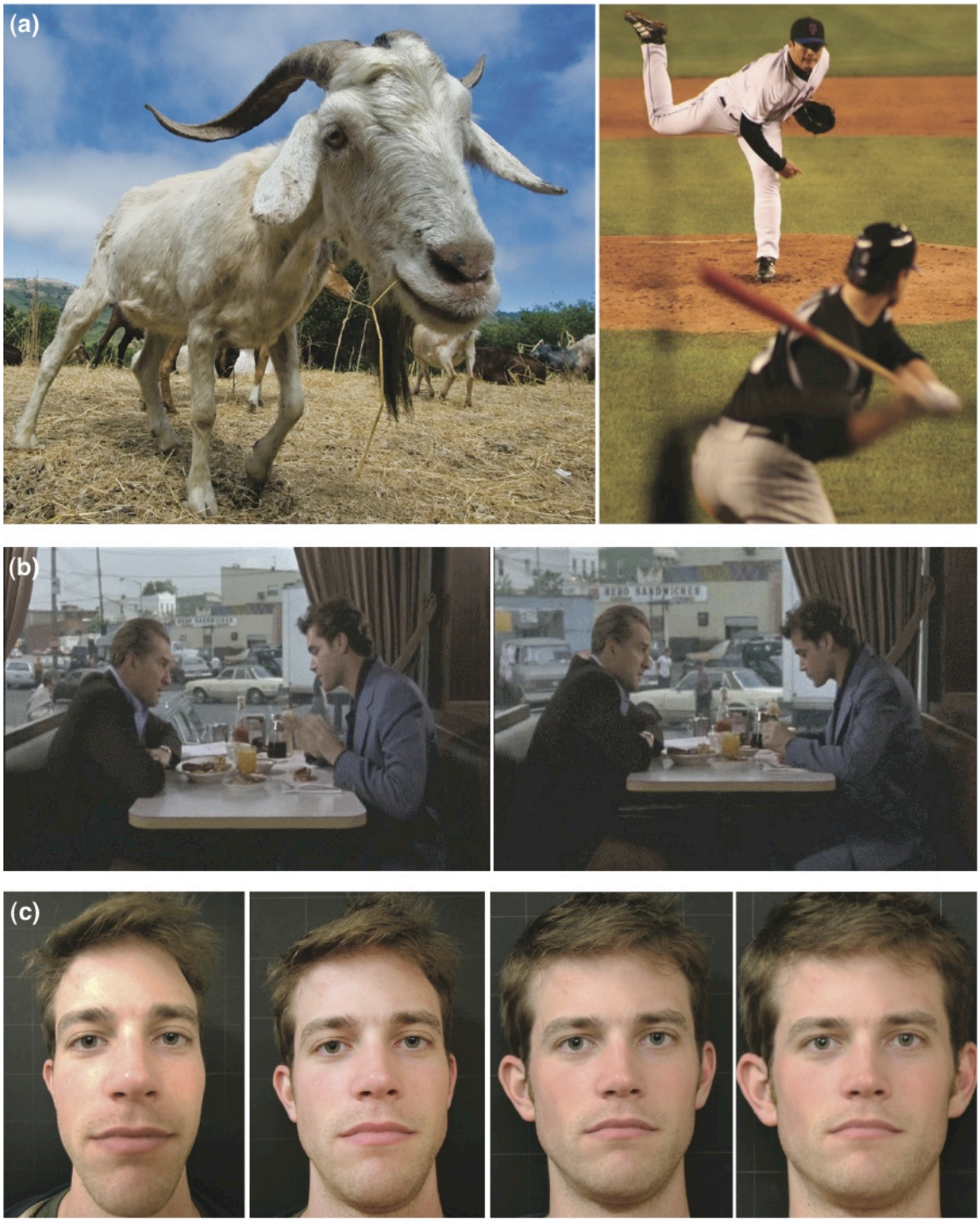

Figure 3.1. Depth compression and expansion with different focal lengths. (a) Left panel: wide-angle effect (short focal length). This picture was taken with a 16-mm lens (all focal lengths are reported as 35-mm equivalent). The goat looks stretched in depth (©Eliya Selhub; http://www.flickr.com/photos/eliya/2734997796/). Right panel: telephoto effect (long focal length). This picture was taken with a 486-mm focal length. The distance between the pitcher's mound and home plate on an official Major League Baseball field is 18.4 meters. This distance appears compressed (©Mitali

Mookerjee; http://www.flickr.com/photos/cool/261259100/). (b) Two stills taken from the movie Goodfellas (Warner Brothers). In this scene, the cinematographer slowly moves from being close up and using a short-focal-length lens to being far away and using a long-focal-length lens, while keeping the actors the same size in the frame. The visual effect is a smooth change from depth expansion to compression (a dolly-zoom shot). The effect is particularly noticeable for the car and buildings seen through the window. (c) Photographs of the same person were taken with focal lengths from left to right of 16, 22, 45, and 216 mm. Lens distortion was removed in Adobe PhotoShop, so the pictures are nearly correct perspective projections. Camera distance was proportional to focal length, so the subject's interocular distance in the picture was constant. The subject's face appears rounder with a short focal length and flatter with a long focal length.

The apparent expansions and compressions in depth are often called perspective distortion, as if these effects are due to a distortion in the physical projection from the scene to the film plane. The effects occur, however, when the projections are geometrically correct. For example, Figures 3.2 and 3.3 contain computer-generated (CG) frames of dolly-zoom shots of a kitchen and a face. Both were created using accurate 3D models and correct perspective projection. Thus, the perceptual effects are not caused by physical distortion in the projections. To explain them, one must consider perceptual mechanisms and people's viewing habits, and that is the purpose of our paper.

Figure 3.2. Frames from a dolly-zoom shot of a kitchen. Using Maya (Autodesk), we imaged a 3D model of a kitchen (Birn, 2008) while simultaneously zooming in and dollying out. The angle subtended by the

bottle was kept constant. The focal length of the lens used in each frame is indicated in the bottom left. The resulting images look very different even though they are all correct perspective projections of a 3D model.

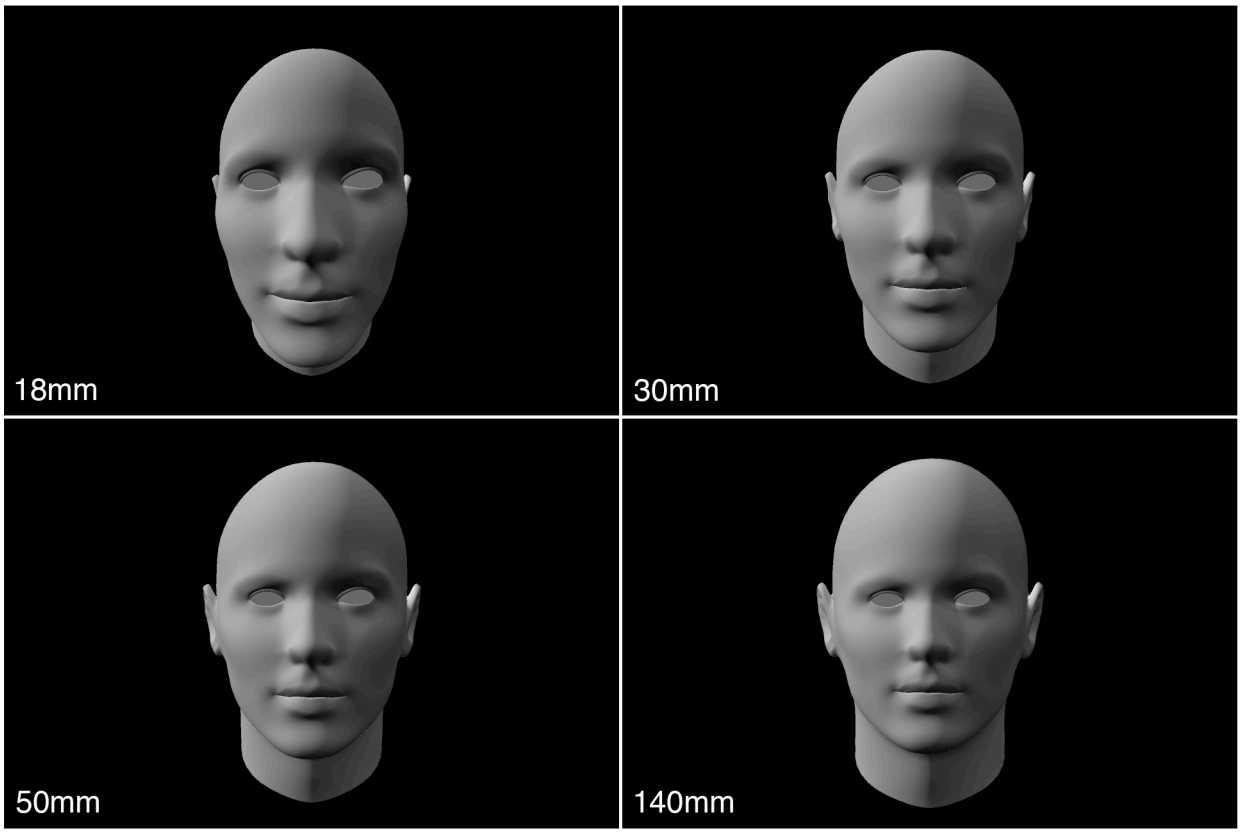

Figure 3.3. Frames from a dolly-zoom shot of a face. The same technique from Figure 3.2 was used. The horizontal angle subtended by the temples of the face was kept constant.

A rule of thumb among professional photographers is to use a focal length of 50 mm for standard 35-mm film (more generally, a focal length equal to the diagonal length of the film or sensor) to create natural-looking images (Kingslake, 1992; Belt, 2008; Modrak & Anthes, 2011; London, Stone, & Upton, 2010). Photography texts offer explanations for this rule's efficacy, but they are either vague or merely restatements of the phenomenon. For example, Reframing Photography claims that using 50-mm lenses “approximates the angle of view and magnification of human vision” (Modrak & Anthes, 2011). The Elements of Photography states that “the normal focal length for a given format most closely approximates human sight, and projects an image with the least distortion and compression of space from foreground to background” (Belt, 2008). In an article for the Photographic Society of America Journal, Current (1990) suggested that 50 mm is the “normal” lens because most people view pictures from a distance equal to the diagonal length of the picture, and at this distance “the perspective is correct and we are most comfortable” when the picture was captured with a 50-mm

lens. We sought a more rigorous explanation of why the 50-mm rule works and why deviations from it yield perceptual distortions.

3.1.1 Pictures and perspective projection

Pictures (i.e., photographs, computer-generated images, and perspective paintings) are created by projecting the light from a 3D scene through a point—the center of projection or COP—onto a flat surface (Figure 3.4). This is perspective projection. The field of view of a captured projection is:

$$\theta = 2\tan^{-1}\left(\frac{l_s}{2f}\right) \qquad \text{Equation (3.1)}$$

where l_s is the diagonal length of the film or sensor, f is focal length, and θ is diagonal field of view. If the image on the sensor is magnified by m, the resulting picture has a diagonal length of ml_s. The distance to the COP is then:

$$d_{COP} = fm. \qquad \text{Equation (3.2)}$$

If the viewer's eye is positioned at the picture's COP (i.e., the viewing distance d_v is equal to d_{COP} and the eye lies along the central surface normal), the image cast by the picture onto the retina matches the image that would be cast by the original scene.

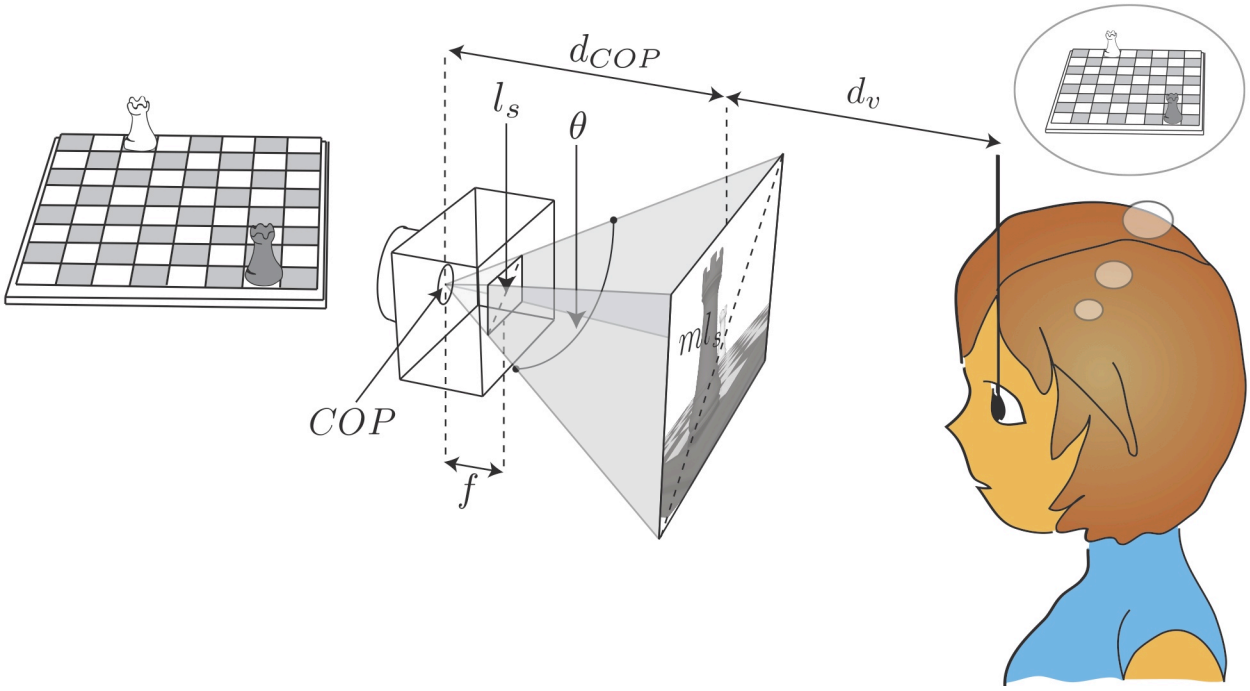

Figure 3.4. Camera, picture, and viewing parameters. A camera with focal length *f* captures a picture on the sensor. The camera's diagonal field of view is θ. The sensor's diagonal length is l_s, and the print is magnified by *m* to have a diagonal length of ml_s. The COP is located at the optical center of the camera. The distance to the COP, d_{COP}, is fm. The print is then viewed from a location along the central surface normal at a distance of d_v. If d_v is equal to d_{COP}, the diagonal field of view subtended by the picture is also θ, and the image cast on the retina matches the image that would be created by the original scene. (Note that the image shown on the print has been rotated to be right-side up.)

Of course, one cannot reconstruct the original scene rigorously from a single retinal image, whether it was generated by a real scene or a picture. But most of the time the brain reconstructs reasonably accurately by using assumptions about perspective (e.g., the chess pieces are the same size, the chessboard is composed of square tiles, the opposite sides of the chessboard are parallel; La Gournerie, 1859; Pirenne, 1970; Sedgwick, 1991; Todorović, 2005). Because viewing a picture from the COP generates the same retinal image as the original scene, it is not surprising that a picture yields a faithful impression of the scene layout or the physical characteristics of a person (Smith & Gruber, 1958; Vishwanath, Girshick, & Banks, 2005; Koenderink, van Doorn, & Kappers, 1994).

3.1.2 Viewing distance and picture viewing

However, people do not necessarily position themselves at the COP when viewing pictures; they may be too far or too near. If viewers fail to compensate for an incorrect distance, the interpretation of the pictured scene may be distorted. For example, Figures 3.5a and b show two pictures of the same scene for two COP

distances; the pictures differ. Figures 3.5c and d show how the apparent 3D scene may differ when one of the pictures (Figure 3.5a) is viewed from two different distances. When viewed from twice the COP distance, the layout specified by linear perspective is stretched in depth: The near chess piece projects to a larger image than the distant piece and, given the assumption that chess pieces are the same size, they appear farther from each other than they actually are. Similarly, for a viewer positioned too close to a picture, the specified layout may be compressed in depth.

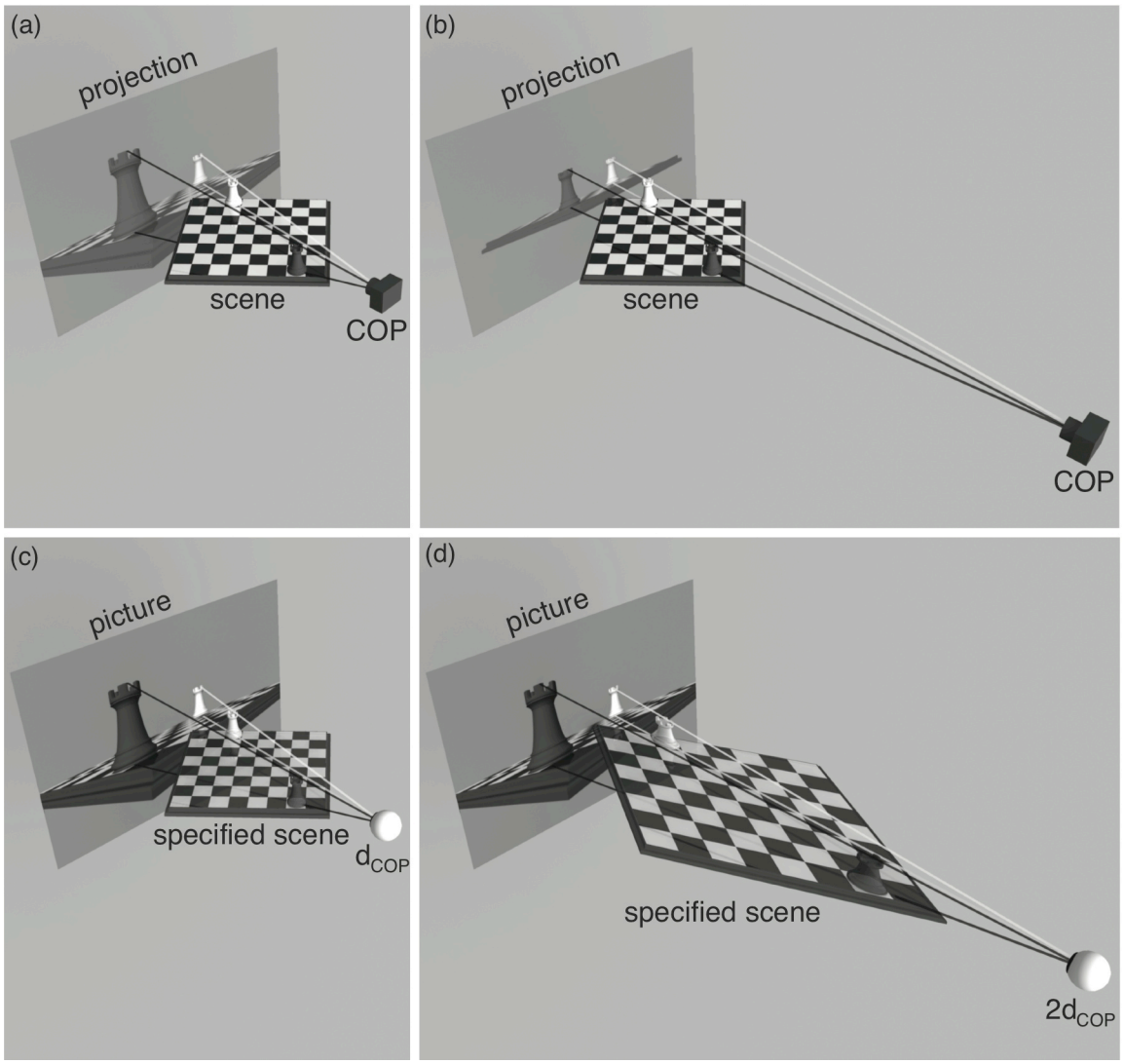

Figure 3.5. (a, b) Perspective projection. The original scene—a chessboard—is projected from two different COPs onto a projection plane. (c) If the picture from panel a is viewed from d_{COP}, the specified scene is the same as the original chessboard. (d) If the same picture from panel a is viewed from twice the COP distance ($2d_{COP}$), the specified scene is stretched in depth relative to the original chessboard.

The perceptual distortion illustrated in Figure 3.5d is geometrically determined from the picture, the viewing position, and linear perspective. Is 3D perception really distorted in this way when people view pictures from the wrong distance? Because pictures do not tend to look distorted in day-to-day viewing, some have suggested that people have a specialized perceptual compensation mechanism to undo these distortions (Pirenne, 1970; Bengston, Stergios, Ward, & Jester, 1980; Yang & Kubovy, 1999). Many studies have asked whether people viewing from incorrect distances do in fact perceive the spatial layout in a pictured scene to be distorted. One early study used photographs of rich scenes to measure perceptual distortions at different viewing distances, and found that perceived distortions were indeed very close to the predicted geometric distortions (Smith & Gruber, 1958). Subsequent studies have measured perceived distortions with simpler pictures, like line drawings or very simple scenes, and reported that perceived distortions were smaller than predicted (e.g., Adams, 1970; Bengston, Stergios, Ward, & Jester, 1980; Yang & Kubovy, 1999; Todorović, 2009). Some have interpreted these findings as evidence of perceptual compensation for pictorial distortions: i.e., observers perceive the 3D scene geometry reasonably accurately even when the depicted geometry from linear perspective is distorted due to viewing from distances closer or farther than the COP (Yang & Kubovy, 1999). Some studies also found that perceived depth from pictures, even when viewed from the COP, is quite inaccurate because it is flattened overall relative to the original scene (Adams, 1972; Hagen, Glick, & Morse, 1978; Lumsden, 1983). This flattening may be due to conflicting depth cues from the picture surface (e.g., binocular disparity, motion parallax) interacting with the perspective cues in the picture content. Again, these studies tended to use pictures with sparse content. For photographs it remains unclear (1) if perceived depth is flattened overall relative to perceived depth when viewing real 3D scenes, and (2) if perceived depth is distorted by incorrect viewing distances.

3.1.3 Hypothesis

We propose that the effectiveness of the 50-mm rule, and the perceptual distortions from other focal lengths, can be explained by the effect of viewing distance on perceived 3D layout and by people's natural viewing habits. We present three experiments whose results confirm the main tenets of the hypothesis. Experiment 1 examines how accurately people perceive 3D shape in pictures with varying levels of realism when viewing from the COP, as compared to a real 3D scene. Experiment 2 examines how perceived 3D shape changes as a function of COP distance and viewing distance, again in pictures with varying levels of realism. Experiment 3 tests how people naturally set their viewing distance when looking at pictures. Based on the findings from these experiments, we describe new guidelines for constructing pictures when the picture creator's intention is to yield accurate percepts of 3D structure.

3.2 Experiment 1: Perception of 3D shape in pictures versus real scenes

The goal of this experiment was to determine if perceived 3D shape in pictures viewed from the COP is significantly flatter than perceived 3D shape in a real scene. To this end, participants were asked to make 3D shape judgments about real objects, and about pictures of objects with varying levels of realism and flatness cues.

3.2.1 Methods

Five young adults participated. All but one were unaware of the experimental hypotheses. Participants gave informed consent under a protocol approved by the Institutional Review Board of the University of California, Berkeley.

The stimuli consisted of two rectangular grids joined together at a hinge. There were three main categories of stimuli: line drawing pictures, realistic CG images, and real wooden hinges (Figure 3.6). Line drawings were rendered using Psychtoolbox and OpenGL, and consisted of red lines on a black background (Figure 3.6a). Realistic CG images were rendered using Autodesk Maya and consisted of photographs of wood that were texture-mapped onto the two sides of the hinge, wallpaper in the background, and a wood-textured floor (Figure 3.6b). Line drawings and CG images were rendered with a COP distance of 28 cm and displayed on a CRT (40.6 x 30.4 cm, 2048 x 1536 pixels). The CRT was corrected for distortion using Psychtoolbox. Real wooden hinges were created using small wooden rectangles with gridlines drawn on to them (Figure 3.6c). Real hinges were placed inside a display box and viewed through an aperture.

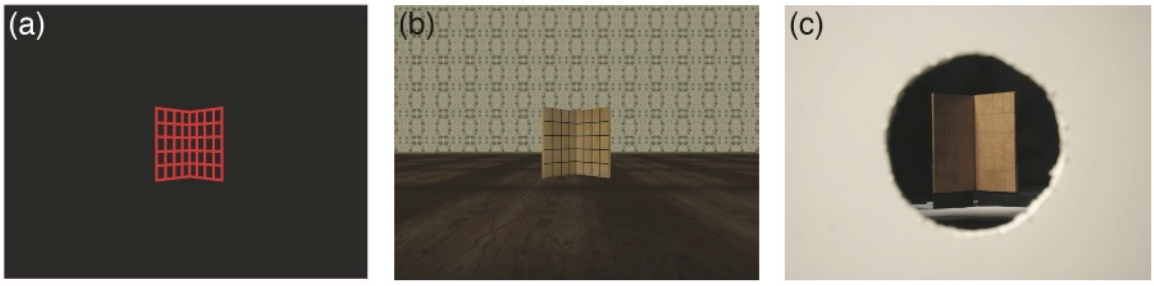

Figure 3.6. Categories of stimuli for Experiment 1. (a) and (b) are line drawing and realistic CG pictorial stimuli; (c) is a photograph of a real wooden hinge that was viewed directly by observers (not a picture).

For both pictorial stimuli (line drawings and realistic CG), the shape of the hinges was randomized to prevent subjects from learning specific pictorial cues. There were nine different hinge shapes that were randomly selected from for each trial. For the realistic CG images, the nine hinges each had slightly different lighting directions (all from the upper left), different wood texturing, and one of three different background environments. The different lighting and texturing prevented subjects from using a purely luminance-based strategy to decide the angle of the hinge. For the real hinge

stimuli, there were three different hinge shapes that were randomly selected from for each trial. The lighting was always from the same location, the upper left.

For the pictorial stimuli, subjects were positioned on a bite bar 28 cm (the COP distance) from the CRT. They viewed either monocularly through an aperture or binocularly, with either the viewing eye or the midpoint of the interocular axis centered in front of the screen. They were told that the two sides of the hinge were rectangular. After each 1.5-second stimulus presentation, subjects indicated whether the hinge angle was greater or less than 90°. A 1-up/1-down adaptive staircase varied the hinge angle symmetrically about the midsagittal axis with four step-size reductions and 10 reversals, and a minimum step size of 2°. Data were fit with a cumulative Gaussian (psychometric function) using a maximum-likelihood criterion (Wichmann & Hill, 2001). The mean of the best-fitting function was defined as the angle perceived as 90°. For the real hinge stimuli, the procedure was the same, except the experimenter manually adjusted the hinge angle out of the subject's sight between trials, and stimulus presentation was controlled with a shutter. The minimum step size for these stimuli was 2.5°.

3.2.2 Results

We wanted to determine if the interaction between 3D cues in a picture and other cues to the flat picture surface causes perceived depth to be compressed in pictures relative to real scenes. We predicted that increasing flatness cues (adding binocular disparity) would lead to increased compression of perceived picture space. We also predicted that decreasing pictorial cues to depth (removing texture gradients and shading) would lead to increased compression of perceived picture space. In the present task, if observers perceived depth as compressed, they would set the hinge angle to be more acute than 90° (e.g., if a 70° angle is perceived as 90°, then perceived depth in the picture is compressed). If observers perceived depth as expanded, they would set the hinge angle to be more obtuse than 90° (e.g., if a 110° angle is perceived as 90°, then perceived depth in the picture is expanded).

Figure 3.7 shows the angle perceived as 90° in each of the experimental conditions. Conditions are ordered roughly in decreasing realism and increasing flatness cues from left to right. The purple and maroon bars on the left show the results for the real scene condition, under binocular and monocular viewing. Observers perceived the 3D shape very accurately in both conditions, setting the angle on average to 100° and 92°, respectively. While both values were greater than 90°, suggesting slight depth expansion relative to the ground truth, individual condition t-tests revealed that neither mean was significantly different from 90° (p = 0.09 and p = 0.60). Perceived depth in the realistic CG images was slightly less than 90° both under binocular and monocular viewing (means of 80° and 78° deg, respectively). This indicates slight depth compression, but again t-tests reveal no significant difference from 90° (p = 0.38 and p = 0.11). This is surprising because in the binocular viewing condition, there was a conflicting depth cue of binocular disparity indicating the flat picture surface. These results suggest that when a realistic picture is viewed from the COP distance, perceived depth is near veridical, even when binocular disparity indicates a flat picture surface.

To compare these results to stimuli more similar to previous experiments, we included a line drawing stimulus, which was viewed binocularly. The average angle perceived as 90° with the line drawing was 52°. This was significantly less than 90° (p = 0.05). This angle was substantially more acute than the perceived angle in the other conditions, indicating a large amount of compression in perceived depth in the line drawing image. There were several differences between the realistic and the line drawing stimuli that may account for the different amounts of depth compression. Things like the textured environment, the shading and texture on the hinges, and the presence of a ground plane might account for this difference. These differences in the realistic stimuli may increase the reliability of perspective information and allow for a more accurate estimate of the hinge angle. Future work is necessary to understand why perceived depth in these pictures was so different.

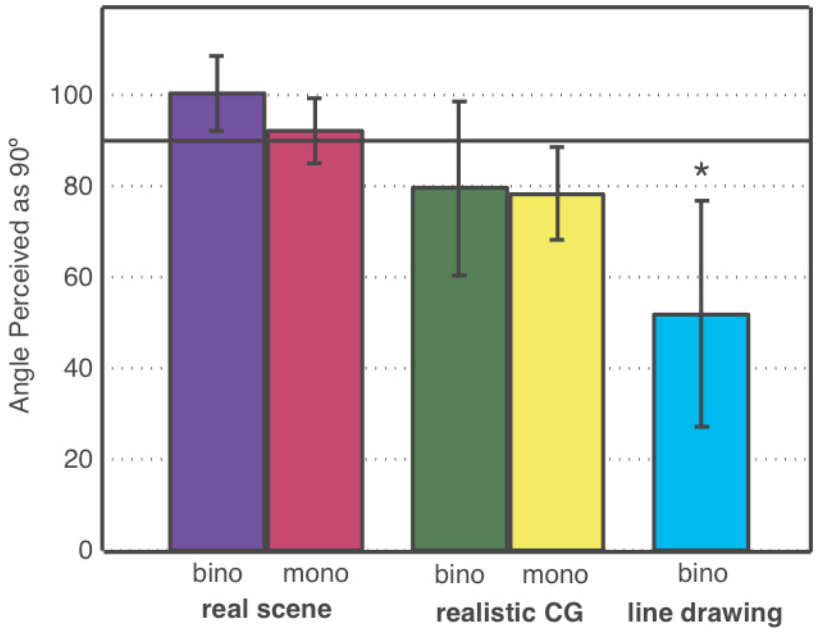

Figure 3.7. Effect of realism and flatness cues on perceived depth in real scenes and pictures viewed from the COP. Each bar shows the mean angle perceived as 90° in each viewing condition. Error bars indicate 95% confidence intervals. Asterisk indicates a mean that is significantly different from 90°.

In all of the above-described conditions, the sides of the hinge opened outward towards the observer, creating a concave surface. Some research has suggested that the visual system has a convexity prior, and therefore observers tend to perceive ambiguous stimuli as convex rather than concave (Kleffner & Ramachandran, 1992). A prior for convexity might tend to make concave surfaces appear less concave. It is possible that some of the depth compression observed in our results is due to the fact that our stimuli were all concave and therefore affected by such a prior. To test this idea, we included a control condition in which the line drawing stimulus was reversed to make a convex angle, opening away from the observer. Observers viewed this stimulus binocularly. The average angle perceived as 90° (not shown the Figure 3.7) in this

convex condition was 45° (compared to 52° in the concave condition) which suggests little if any difference in depth compression for convex and concave stimuli. A paired t-test performed on the two line drawing conditions (concave and convex) revealed no significant difference in the angle perceived as 90°. From these results, we can conclude that the depth compression in line drawing pictures was likely not caused by a convexity prior.

3.3 Experiment 2: Perception of 3D shape in pictures when viewed from the wrong distance

The goal of this experiment was to examine how perceived 3D shape changes as a function of COP distance and viewing distance in pictures with varying levels of realism.

3.3.1 Methods

The same five subjects from Experiment 1 participated. Stimuli and methods were similar to Experiment 1, with the following exceptions. Only line drawing and realistic CG pictures were used: no real objects. The images were rendered with five different COP distances—11, 16, 28, 55, and 79 cm. Observers again viewed from 28 cm, so two conditions had closer COP distances, two had farther COP distances, and one was the correct distance. Staircases for each of the five COP distances were randomly interspersed and repeated six times. Observers viewed all stimuli binocularly.

The line drawing stimuli were identical to Experiment 1. For the realistic CG pictures, there were four additional stimuli in which the realism and content of the CG environment were varied (Figure 3.8). We included multiple versions of the realistic CG stimuli to try to understand the cues that are necessary for perceptual compensation for viewing distance distortions, if such compensation exists. The main realistic CG version contained familiar objects (cubes) and realistic shading with predictable lighting conditions (from the upper left) (Figure 3.8a). The cubes were included in the scene in order to provide potential geometric cues for perceptual compensation (La Gournerie, 1859; Yang & Kubovy, 1999). The shading on the hinges also provided a potential compensation cue because the oblique lighting direction created a situation in which acute angles had larger differences in shading between the two sides of the hinge, and obtuse angles had smaller differences in shading, regardless of the COP distance.

The other versions of the realistic CG stimuli had one or both of these potential compensation cues removed. The ‘no cubes’ condition (Figure 3.8b) contained realistic shading and predictable lighting, but not familiar objects. The ‘no cubes, random lighting’ condition (Figure 3.8c) had the lighting direction randomized to reduce the reliability of the shading cue. The ‘no shading’ condition (Figure 3.8d) had uniformly bright ambient lighting that removed any differential shading between the two sides of

the hinge, but the scene still contained cubes. The ‘no shading, no cubes’ condition (Figure 3.8e) had the same ambient lighting, but the cubes were removed.

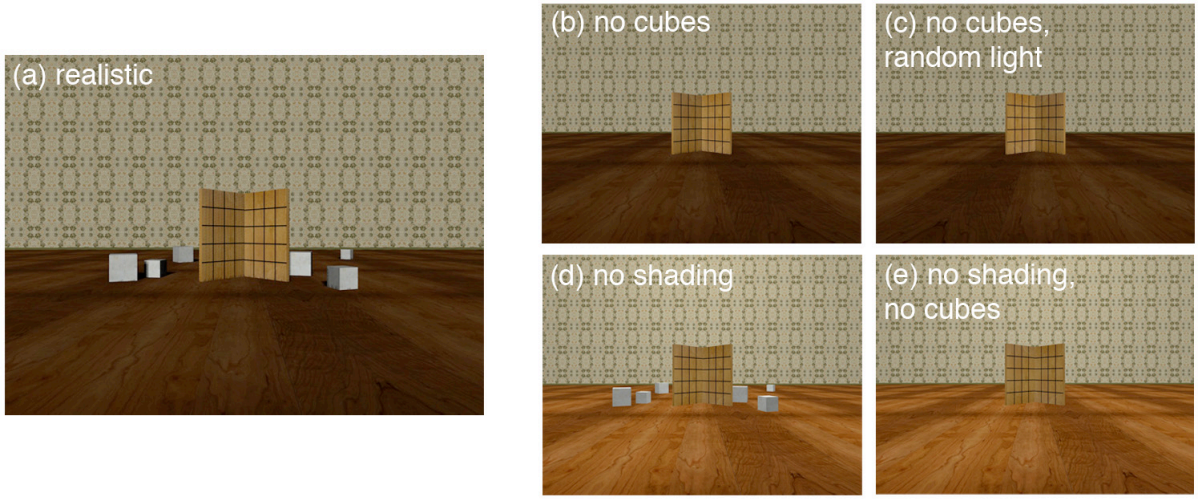

Figure 3.8. Realistic CG stimuli for Experiment 2. (a) The main stimulus contained geometric and lighting cues that could potentially be used for perceptual compensation. Each additional condition had one or both of these cues removed: (b) lighting but no cubes, (c) lighting randomized and no cubes, (d) only ambient lighting, with cubes, (e) only ambient lighting, no cubes. A line drawing stimulus was included in this experiment also, but is not shown in this Figure.

3.3.2 Results

If subjects were able to compensate for their viewing distance relative to the COP distance, they would perceive the depicted hinge angle correctly and would set the hinge to 90° in scene coordinates. If subjects failed to compensate for the difference between their viewing distance and the COP distance and instead interpreted the scene directly from the geometry of the retinal image (assuming that the hinge is composed of rectangular planes), they would set the depicted hinge angle to different values in scene coordinates for each COP distance. The predicted settings for this second hypothesis can be calculated from geometric analyses of perspective projections such as those presented in Sedgwick (1991) and Rosinksi and Farber (1980). These analyses show that the angle in scene coordinates and the depicted angle are related by the ratio of the viewing distance to the COP distance. With no compensation, the predicted hinge angle perceived to be 90° is:

$$\omega = 2tan^{-1}\left(\frac{d_v}{d_{COP}}\right) \qquad \text{Equation (3.3)}$$

where d_{COP} is the COP distance of the picture and d_v is viewing distance.

The results of Experiment 2 are shown in Figure 3.9. Figure 3.9a shows the results for the line drawing condition (blue) versus the main realistic CG condition (red). The perceptual compensation prediction is plotted as a dashed line in 3.9a and 3.9b. The no-compensation prediction (from Equation 3.3) is plotted as the solid curve in 3.9a and 3.9b. The results were very different for the two types of stimuli. For realistic CG images, the data were quite consistent with the no-compensation prediction. When the COP distance was less than the viewing distance, subjects perceived a larger (more obtuse) angle as 90°, which means that they experienced depth expansion. When the COP distance was greater than the viewing distance, they perceived a smaller (more acute) angle as 90°, meaning that they experienced depth compression. When the COP distance and viewing distance were the same, a 90° hinge was perceived as close to 90°, so they experienced neither expansion nor compression. There were slight, but systematic differences between the realistic CG data and the no-compensation prediction. Generally, subjects set the hinge angle to slightly less than the predicted value, which means that they perceived the angles as somewhat flatter than dictated by the geometry of the retinal image. (The one exception to this is at the greatest COP distance where they set the angle slightly larger than predicted.) For the line drawing images, the trend in the data is also consistent with the no-compensation prediction, but all perceived angles were consistently less than the predicted values, meaning subjects perceived the angles as much flatter than dictated by the geometry of the retinal image. We believe that the cause of this flatness bias (both the small bias in the realistic CG stimuli and the large bias in the line drawing stimuli) is the flatness specified by a number of cues including binocular disparity and focus cues, as discussed in the previous section on Experiment 1 (Watt, Akeley, Ernst, & Banks, 2005).

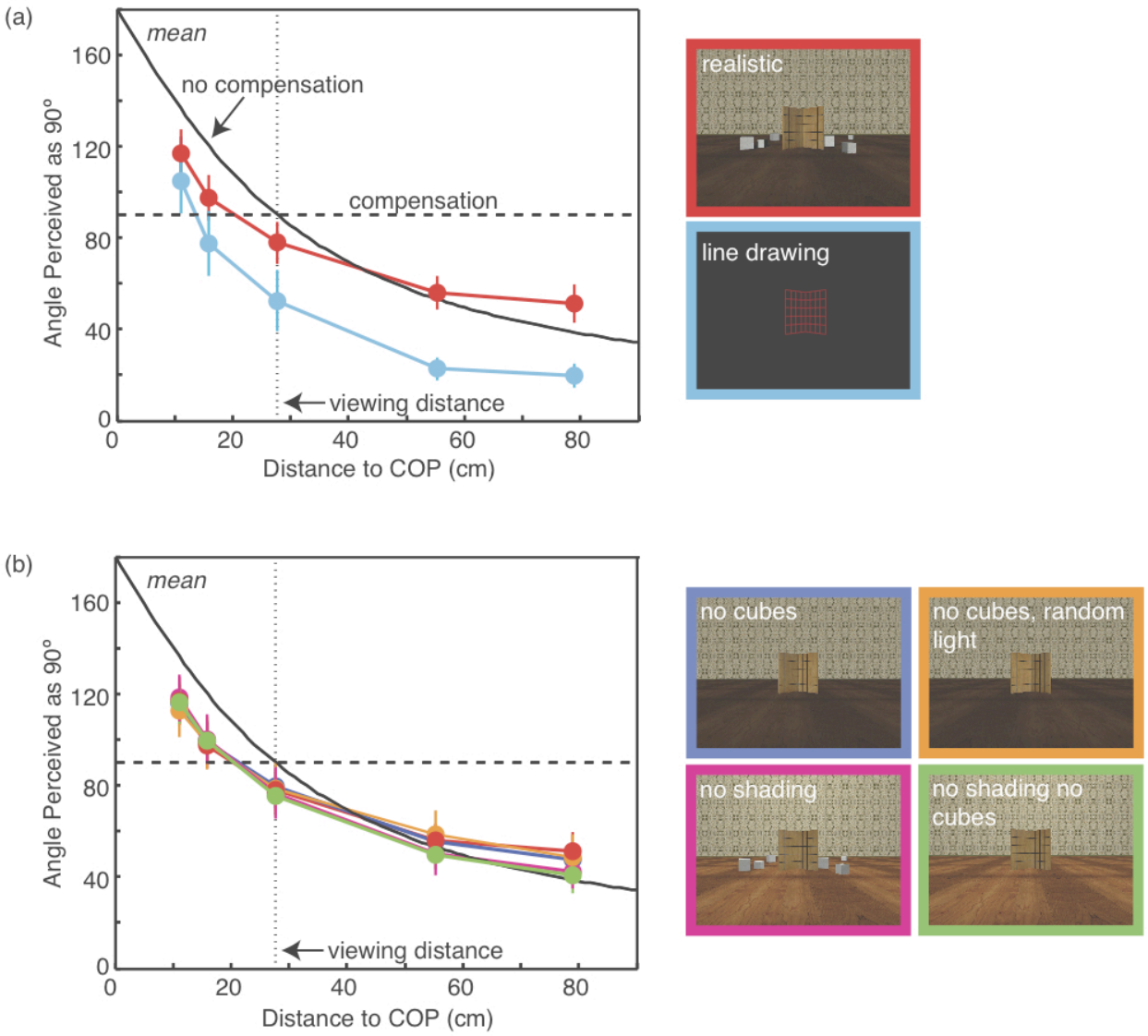

Figure 3.9. Effect of distance to COP on the angle perceived as 90°. Circles represent the mean angle perceived as 90° across subjects; error bars are standard errors. The dotted vertical line indicates the viewing distance. Example images from each type of stimulus are shown to the right of the data plots. The color of the border on the example indicates the color of the circles in the data plot. (a) Comparison of results for realistic CG and line drawing stimuli. (b) Comparison of results for all versions of the realistic CG stimuli.

Figure 3.9b shows the results for each of the versions of the realistic CG stimuli (four versions with cues removed, plus the original all-cue version re-plotted in red). There were no systematic differences between the results for any of these conditions, suggesting that neither the shading nor the familiar objects played a large role in determined the angle perceived as 90°. This makes sense because these manipulations were expected to show an effect of reducing perceptual compensation, but we found no evidence for perceptual compensation to begin with.

With regard to the small deviations in the data from the no-compensation prediction, some previous studies have reported similar, but larger deviations (i.e., the perceptual distortions were significantly less than predicted by perspective geometry) and suggested that there might be partial compensation for incorrect viewing distance (Adams, 1972; Lumsden, 1983; Yang & Kubovy, 1999). These studies used line drawings or very simple scenes with little perspective information. From our line drawing stimuli, we found that geometric predictions are not a good model for these types of stimuli, most likely due to the many depth cue conflicts. Rather than positing a specialized perceptual mechanism for viewing distance compensation, a more parsimonious explanation might be that deviations from the geometric predictions of picture viewing are mostly caused by cue conflicts (Sedgwick, 1991; Sedgwick & Nicholls, 1994; Todorović, 2009).

From Experiments 1 and 2, we can conclude that viewers perceive pictorial depth in realistic pictures and photographs in a way that is very similar to the geometric perspective predictions. Generally, the perceived depth in a picture will be mostly accurate when it is viewed from the COP distance. Perceived depth will be compressed when a picture is viewed from too close and expanded when a picture is viewed from too far.

3.4 Experiment 3: Preferred viewing distance for pictures

In this experiment, we measured people's preferred viewing distance for pictures of different focal lengths, magnifications, and print sizes. The results enabled us to determine whether people use consistent strategies for setting viewing distance and, if so, what those strategies are. We could then analyze the relationship between viewing distance and COP distance in natural viewing.

3.4.1 Methods

Participants

Eight young adults participated in the main experiment, and 11 additional young adults participated in a follow-up experiment. All were unaware of the experimental hypotheses. Participants gave informed consent under a protocol approved by the Institutional Review Board of the University of California, Berkeley.

Stimuli

Scenes for the pictures were selected from five categories described in the scene-recognition literature: indoor, street, outdoor open (e.g., coastline, mountains), outdoor closed (trees), and portrait (Torralba & Oliva, 2003; Torralba, 2009). For each of

the first four categories, we used three unique scenes: one photographed scene and two CG scenes. For the fifth category, we used two photographed scenes. Figure 3.10 provides example pictures of scenes from each category.

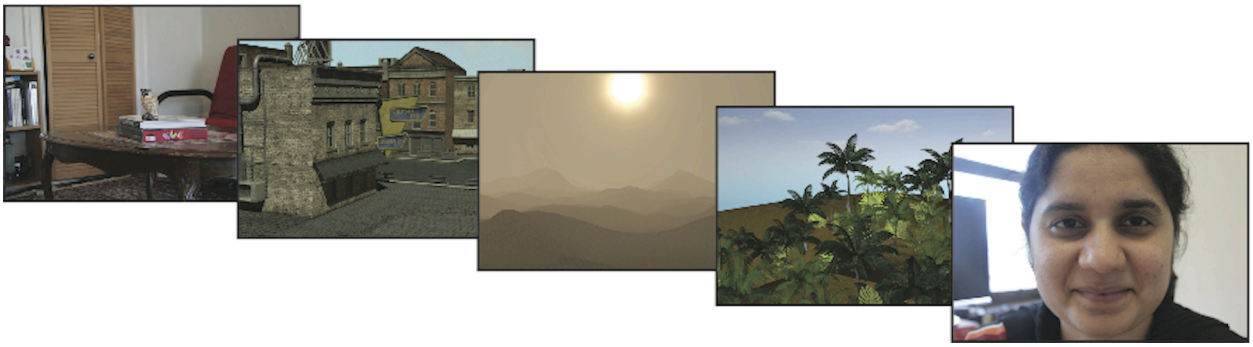

Figure 3.10. Example pictures from each scene category. Left to right: indoor, street, outdoor open, outdoor closed, and portrait.

The photographs were taken with a Canon EOS 20D SLR camera and saved as low-compression JPEG files. CG images were saved as TIFF files. Both types of image files were saved at 300 dots/inch (dpi) and printed on photographic paper with a resolution of 300 dpi and an aspect ratio of 3:2. All CG images were rendered with infinite depth of field (i.e., no blur) and were illuminated with a combination of a directional and ambient light sources. For the photographs, we used the smallest aperture allowed by the lighting environment to minimize differences in depth of field and exposure between photographs taken with different focal lengths.

There were two primary stimulus manipulations: focal length and magnification. To manipulate focal length, we selected a focal object in each scene and created a series of five images taken with five different focal lengths—22, 32, 56, 112, and 160 mm (35-mm equivalent)—while keeping the camera at one location. All of those pictures were magnified eight-fold and printed at 18 × 12 cm (7 × 5 in). To manipulate magnification, we took photographs with a 56-mm lens and printed them at 18 × 12 cm (same as the previous), and four additional sizes (6 × 4, 9 × 6, 29 × 19, and 39 × 26 cm).

By changing focal length, the focal object became different sizes in the prints (Figure 3.11a). To determine whether the varying size of that object affected preferred viewing distance, we also created five images in which the focal length was fixed at 56 mm, but the camera was dollied in and out so that the size of the focal object would match those from the five focal lengths (Figure 3.11b). These were all printed at 18 × 12 cm.

Figure 3.11. Changing focal length and camera distance to maintain constant size of the focal object (in this case, a pillow). (a) The effect of changing focal length while keeping camera position constant. The focal lengths from left to right are 160, 56, 32, and 22 mm. (b) The effect of changing camera distance while holding focal length constant. From left to right, the camera is moved farther and farther from the focal object. Focal length was always 56 mm. By moving the camera farther from the focal object, the sizes of the focal object are matched to those in the upper row without changing COP distance. Differences between the images in panels (a) and (b) are particularly noticeable in the apparent shape of the bed and slant of the wall.

We were curious to see whether these results would generalize to larger picture sizes, so we conducted a follow-up experiment with larger pictures. Eleven new subjects participated. The stimuli were the same with a few exceptions. Only four scenes were used: one indoor, one street, one outdoor open, and one outdoor closed. All pictures were CG. We created pictures with three focal lengths (22, 56, and 160 mm) and printed each at four sizes (18 × 12, 53 × 35, 73 × 49, and 100 × 67 cm). We dollied the camera away from the focal object as we increased the focal length in order to match the size of the object across focal lengths. Subjects were shown each focal length twice and each print size twice with a random selection of two of the four scenes.

Procedure

At the start of each trial, a picture was mounted on a wall at the subject's eye level. Subjects stood initially 5 m from the picture. They were instructed to walk back and forth along a line that was perpendicular to the picture until they were at “the best distance to view the picture from.” If they asked for clarification, we told them to select “whatever distance you prefer to view from.” Once they indicated that they were at the preferred distance for that picture, the experimenter recorded the distance with a photograph. The trials were recorded so preferred distances could be measured off-line using the ruler tool in Adobe Photoshop.

Subjects were presented with a picture from each level of each manipulation eight times, with a random selection of 8 of the 14 scenes. Therefore, subjects did not see the same scene/manipulation combination twice. The main experiment took place over four sessions. The order of presentation was randomized. We also wanted to measure test-retest reliability, so we presented eight pictures four times each (once per session). Each subject thus completed a total of 136 trials.

The procedure of the follow-up experiment was identical to the main experiment with a few exceptions. The subjects began each trial standing 6.5 m from the picture and again moved toward and away until they were at their preferred viewing distance. We measured viewing distance with a laser range finder. These measurements occurred in two sessions. Presentation order was randomized. To assess test-retest reliability, we randomly presented three pictures four times (twice per session). Each subject therefore completed a total of 36 trials in this phase of the experiment.

We also investigated whether the manner of picture viewing—standing in front of a wall-mounted picture as opposed to holding a picture while seated—affects preferred viewing distances. Three subjects from the main experiment participated in these measurements. They sat in a chair and held each picture in their hands. They varied distance by adjusting their arms until they achieved the preferred value. We measured that distance using the laser range finder. A subset of the stimuli from the main experiment was used with one focal length (56 mm) and two print sizes (9 × 6 and 18 × 12 cm). For each print size, 10 of the 14 scenes were randomly selected. Each subject completed a total of 20 trials.

3.4.2 Results

We first asked whether the data from the follow-up experiment differed from the main experiment. A one-way ANOVA performed on the data from overlapping conditions revealed no significant effect ($p = 0.53$), so from here on we combine the data from these two experiments.

The results for the main stimulus manipulations—focal length and magnification—are illustrated in Figure 3.12. Figure 3.12a shows mean preferred viewing distance as a function of focal length. The results are plotted separately for each magnification. Some magnifications only have one focal length because the two variables were not completely crossed in the main experiment. There was clearly no effect of focal length on preferred viewing distance for a given magnification. Figure 3.12b shows the same data, but with mean preferred viewing distance plotted as a function of magnification. There was a strong effect of magnification/picture-size on preferred viewing distance, independent of focal length. The dashed line shows a linear regression of these data ($p < 0.0001$). Equations for the line as a function of picture diagonal (l_p) and magnification (m) are shown next to the line. Notably, the y-intercept of the line (25 cm) is the same as the nearest comfortable viewing distance for young adults (Ray, 2000).

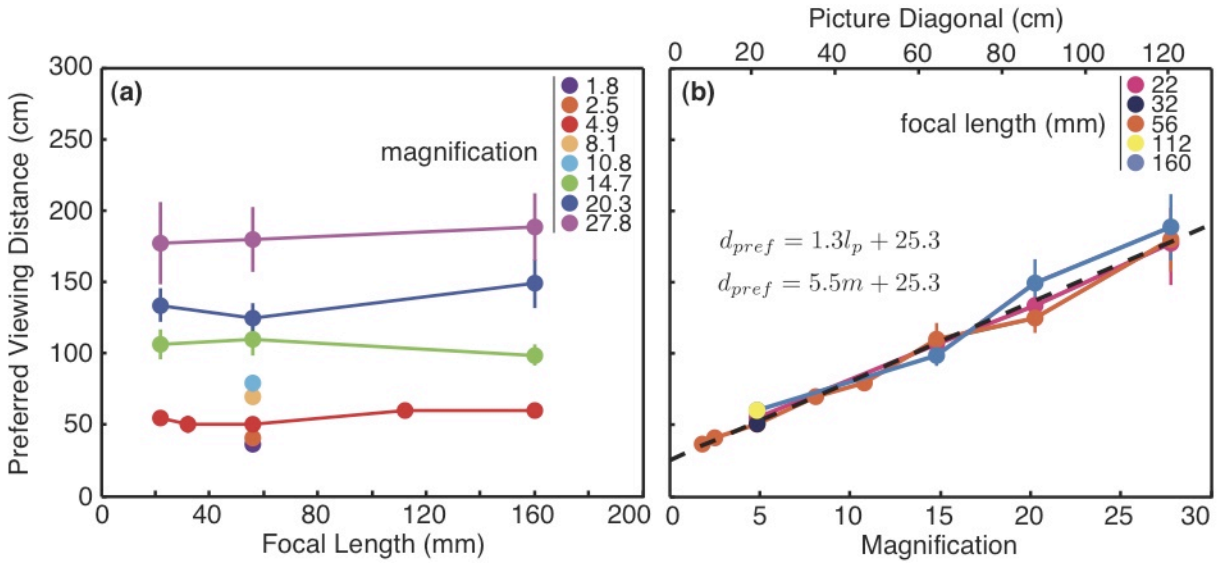

Figure 3.12. Effects of focal length and magnification on preferred viewing distance. (a) Preferred viewing distance is plotted as a function of focal length for each magnification. Circles represent the data: the mean preferred viewing distance across subjects. Error bars represent standard errors of the mean. Each color represents a different picture magnification (and therefore a different picture size), as indicated by the legend. (b) Data from panel a replotted as a function of magnification for each focal length. The diagonal length of the picture for different magnifications is indicated at the top. A linear regression of the data is represented by the dashed black line and the two equations. All five focal length levels are plotted for magnification = 4.9, but the circles are largely overlapping because there was so little effect of focal length.

We re-plotted a subset of the data in Figure 3.12 in a way that allows us to examine the picture properties that determine preferred viewing distance. Figure 3.13a shows two subsets of stimuli for one example scene: five focal lengths for one magnification and eight magnifications for one focal length. Figure 3.13b shows the average preferred viewing distance for these subsets of all stimuli. If subjects preferred that pictures subtend a particular visual angle, or field of view (FOV), preferred distance would be proportional to print size, and the data would fall along one of the blue lines in Figure 3.13b, depending on the desired angle. Alternatively, if subjects always moved to the distance of the picture's COP (d_{COP}), the preferred viewing distance would be proportional to focal length and magnification (Equation 3.2), and the data would lie on the red lines in Figure 3.13b. The left panel shows that preferred viewing distance was barely affected by COP distance. From the nearest to farthest COP, preferred distance increased by only 20%, significantly less than the 614% change that would have occurred if subjects matched viewing distance to COP distance. The right panel shows that preferred viewing distance was strongly dependent on magnification (or equivalent picture size). But subjects were not establishing a constant field of view; rather, they preferred a small field (~22°) with small prints and a larger field (~36°) with large prints. This smaller preferred field of view for small prints likely reflects a trade-off between viewing comfort and angle subtended by the print. We conclude that picture viewers do

not naturally set their viewing distance to a picture's COP distance. Instead they adjust distance according to the field of view (albeit smaller fields for small prints and larger fields for large prints). These data are consistent with television-viewing studies, which show that preferred viewing distance is determined by the size of the screen rather than image content or television resolution (Ardito, 1994; Lund, 1993).

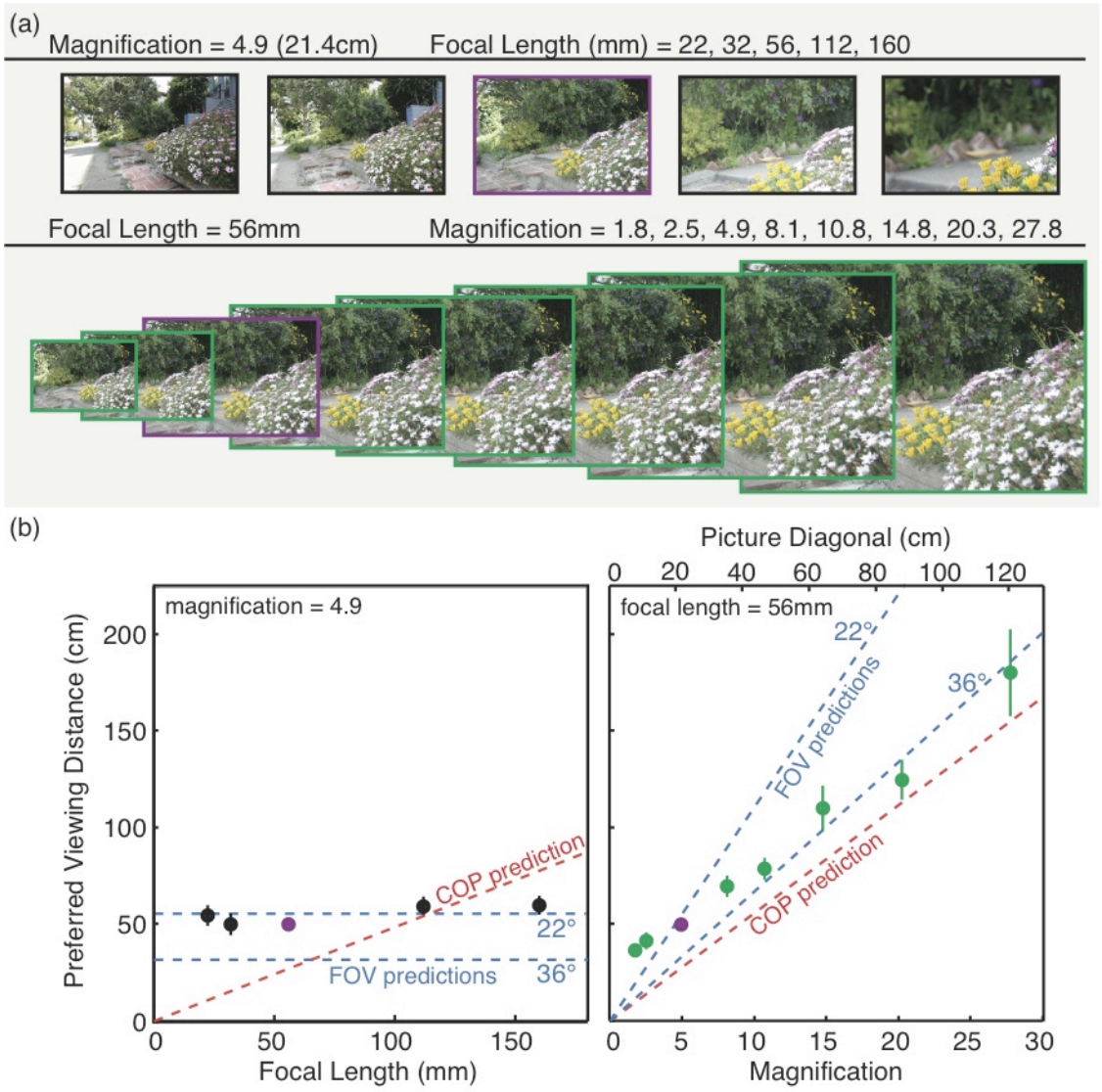

Figure 3.13. Subset of data from Experiment 3. (a) Example stimuli for two subsets of conditions. One subset contains five focal lengths with a magnification of 4.9 (diagonal length of the printed picture was 21.4 cm). The other subset contains eight magnifications with a focal length of 56 mm. The relative sizes of the stimuli actually changed by a factor of 15.4, but we cannot show such a large change in the Figure. Therefore, the change in relative size shown previously is qualitative. The purple boxes around two of the pictures indicate the one that was in both subsets. (b) Two plots of average preferred viewing distance across subjects for each manipulation. Black and green circles represent the focal length and magnification manipulations, respectively, and correspond to the boxes around the pictures in panel a.

The purple circles in both plots represent data from one magnification and focal length (4.9 and 56 mm, respectively). Error bars represent standard errors of the mean.

We also examined whether there were differences in the effects of focal length and magnification between photographs and CG images. The average difference between the preferred viewing distance for photographs and CG images across the whole data set was only 3.2 cm (standard deviation = 4.6 cm). Because the experimental conditions were not fully crossed, we could not perform an ANOVA on these results. A one-way ANCOVA on the effect of focal length and magnification on preferred distance for photographs versus CG revealed no significant difference for either manipulation (p = 0.55 and 0.99).

Although magnification had by far the largest effect on preferred viewing distance, there was a small but significant effect of focal length (for example, the slope of a linear regression of the data plotted in the left panel of Figure 3.14b was = 0.1, p = 0.008). This effect could have been due to the picture's COP distance or to the size of the focal object (i.e., the object centered in the frame). Recall that we included a control condition in which the size of the focal object was manipulated by dollying the camera rather than changing the focal length (Figure 3.11b). A one-way ANCOVA on preferred distance as a function of normalized focal object size for the two groups (focal length and camera distance) revealed no significant difference between the effects of focal length and camera distance (p = 0.46). We conclude that this small effect was due to the size of the focal object and not due to an effect of COP distance on preferred viewing distance.

To assess test-retest reliability, we also calculated the standard deviation of preferred viewing distance for each subject for each of the repeated pictures. The mean standard deviations across all images and subjects were 14 cm for the main experiment and 22 cm for the follow-up experiment. These values are small relative to the means, so the preferred distances were reasonably repeatable.

Finally, we examined the effect of standing (where subjects adjusted their viewing distance by walking to and fro) and sitting (where subjects held the pictures in their hands) on preferred viewing distance. A two-way ANOVA performed on overlapping conditions from the two sets of data revealed no effect (p = 0.59), so we conclude that people behave similarly when viewing wall-mounted pictures while standing and when viewing handheld pictures while sitting (provided that picture size is not so large for arm length to limit the ability to set distance to the desired value).

3.5 Discussion

3.5.1 Summary

We can now explain why focal length affects apparent depth in pictured scenes and facial appearance in portraits. Recall that long- and short-focal-length pictures look

respectively compressed and expanded in depth (Figures 3.1, 3.2, and 3.3). We propose that people's preferred field of view when looking at most pictures leads them to view long-focal-length pictures from too near and short-focal-length pictures from too far. Perceptual compression and expansion occur because people do not take their incorrect viewing distances into account. Thus, scenes captured with long lenses look compressed in depth, which makes faces apparently flatter. Likewise, scenes captured with short lenses appear expanded in depth, which makes faces look rounder.

3.5.2 Focal length recommendations

However, this does not tell us why pictures created with a 50-mm lens look most natural, i.e., neither expanded nor compressed. To investigate this, we calculated for each picture size the focal length for which the subjects' average preferred viewing distance would be equal to the COP distance. We call this the recommended focal length:

$$f_{rec} = 43.3\frac{d_{pref}}{l_p} \qquad \text{Equation (3.4)}$$

where d_{pref} is the average preferred viewing distance, l_p is the diagonal length of the picture, and 43.3 is the diagonal length of standard 35-mm film in millimeters. The recommended values from our data, calculated by averaging the preferred viewing distance across all focal lengths for each picture size from Experiment 3, are plotted in Figure 3.14. The regression line from Figure 3.12b is also replotted in terms of recommended focal length. The equation for the line is:

$$f_{rec} = 55 + \frac{1096}{l_p}. \qquad \text{Equation (3.5)}$$

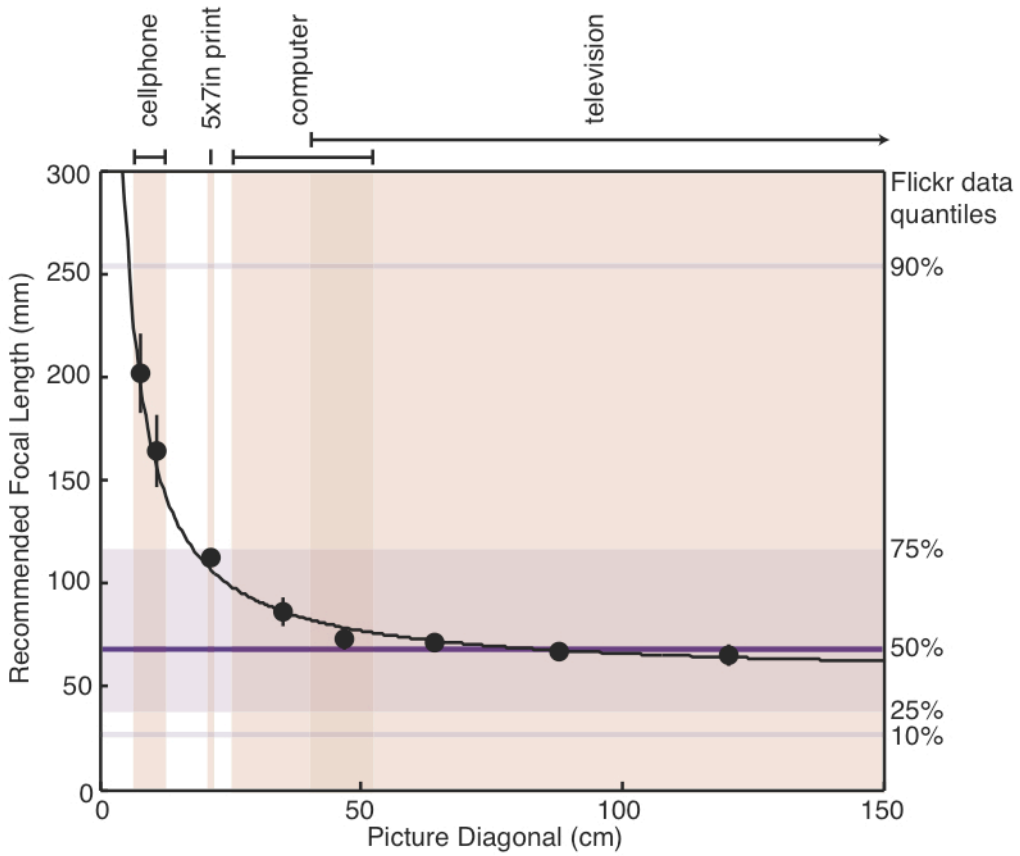

Figure 3.14. Recommended focal length as a function of picture size. We calculated recommended focal length for each picture size by determining the average preferred viewing distance across all focal lengths from Experiment 3 (Figure 3.12b) and then calculating the focal length that would produce a COP distance equal to the preferred distance (Equation 3.4). Circles represent those values, and error bars represent standard errors. The black curve shows the linear regression from Figure 3.12b replotted in terms of recommended focal length. Vertical bands indicate some typical image sizes for various formats. Horizontal bands indicate quantiles from several cumulative probability values for 3,930 Flickr photographs taken with SLR cameras.

Thus, for prints 35 cm or larger, the recommended focal length is ~50 mm. Most prints, particularly professional ones, are at least that size. We claim therefore that following the 50-mm rule of thumb maximizes the odds of a viewer looking at the photo from the COP distance and thereby makes it most likely that the percept will be undistorted. This rule has presumably evolved over time based on collective experience. Similar recommendations apply for cinematographers, computer-graphics engineers, and painters of realistic images. Some typical image sizes for various formats (Take, 2003) are superimposed as vertical bands in the figure. For most venues, the recommended focal length is ~50 mm (35-mm equivalent). With the small screens of mobile devices, longer focal lengths should be used. If image creators know the size of a typical print or projection of their work, they can use Equation 3.5 to make

a better choice of focal length or to change the distance of the COP in postprocessing (Carroll, Agarwala, & Agrawala, 2010).

Most photography texts advocate the 50-mm rule (Kingslake, 1992; Belt, 2008; Modrak & Anthes, 2011; London et al., 2010), but we wondered whether the rule is actually used in practice. To find out, we collected 3,930 photographs from the website Flickr that were taken with single-lens reflex (SLR) cameras. (These cameras tend to be used by professionals and serious hobbyists.) We obtained the 35-mm-equivalent focal length for those photos from their EXIF data. The median is 68 mm (50% quantile horizontal line in Figure 3.14). Interestingly, 68 mm is closer than the advocated 50 mm to our recommended focal length for a wide range of sizes. Thus, current practice deviates slightly from the 50-mm rule, but is more consistent with our experimental data.

Our recommended focal length is much longer for small picture sizes, such as those on mobile devices. The viewing of images on mobile devices is becoming much more common (Choney, 2009; Carlsson & Walden, 2007). People tend to view smart phones from ~30 cm (Knoche & Sasse, 2008). When standard content is viewed at that distance, the smart-phone user is generally much farther from the display than the COP distance, making the images of objects subtend small angles and producing expansion in apparent depth. Interestingly, smart-phone viewers prefer standard content to be magnified and cropped (Knoche et al., 2007; Song et al., 2010), which increases the COP distance, much like increasing focal length; this practice should make the viewed content appear less expanded than it otherwise would.

3.5.3 Conclusion

We claim that the 50-mm rule emerged because of people's tendency to view pictures from a distance that establishes a desirable field of view and their inability to compensate when that tendency yields an incorrect viewing distance. Our data can be used to create better guidelines, based on empirical results, for creating effective pictures for all viewing situations.

Chapter 4. Conclusion

The human visual system has evolved to determine the 3D layout of the natural environment. However, we often use this same system to determine the 3D layout of pictured environments. Studying 3D perception can provide fundamental insights into how the human visual system works, as well as practical insights into how to create and display information in pictures.

4.1 Corresponding points and stereo displays

Chapter 2 described the hypothesis that the locations of binocular corresponding points may be adaptive for optimizing the precision of depth perception in the natural environment. We can also consider how to display stereo information in a way that is optimized for precision of depth perception on the display (i.e., most likely to stimulate corresponding points). Several studies have observed that traditional computer working environments are set up such that the binocular disparities created by the display surface may tend to stimulate corresponding points. Computer workstations tend to contain nearby, flat screens tilted top-back (Ankrum, Hansen, & Nemeth, 1995). This agreement between the horopter and the screen may be helpful for ensuring that the content displayed on the screen is easily fused by the eyes into a single image (Grove, Kaneko, & Ono, 2001). As stereo displays become more commonplace, it is important to make sure that disparities created by the stereo picture are not just fused, but also perceived correctly and precisely. This is particularly valuable for practical stereo applications like medical imaging and scientific visualization (Held & Banks, 2008; Held & Hui, 2011). Making sure that a stereo display surface is aligned with the horopter could help people make better use of the depth information on a stereo picture. The experiments in Chapter 2 revealed a new aspect of the shape of the vertical horopter that has not be previously described: on average, the vertical horopter is curved in depth. Combined with the curved shape of the horizontal horopter, this suggested that the optimal surface for stereo displays should not be flat (Schreiber et al., 2008). While creating curved display surfaces may not be practical, it may be possible to warp a pair of stereo images such that zero-disparity points fall roughly along the surface of the horopter. This could be useful for identifying relevant shapes and structures in stereo medical images like angiograms and mammograms (Held & Hui, 2011), or any situation in which it is desirable to have good depth precision over a large area of the visual field.

We also discussed in Chapter 2 that certain disparities are impossible in the natural environment. Depending on the capture and display of a stereo picture, one can easily create these impossible disparities. It is unclear how the visual system would interpret such disparities. For instance, if a stereo picture is presented such that the eyes' visual axes must be parallel to fuse it (i.e., eyes fixated at infinity), what is the

perceived absolute distance of a point presented behind the picture surface, with uncrossed disparity? Equation 2.5 could be used as a guideline to determine if the disparity of a given point on screen is impossible in the natural environment. Much like the guidelines presented in Chapter 3, this calculation would need to consider all aspects of displaying and viewing, such as the capturing cameras, the size of the display, the method of stereo picture presentation, and the viewing location.

4.2 Picture perception and object constancy

Pictures are useful in part because viewers can gain a faithful impression of the pictured content even when they are not positioned precisely at the COP. However, the ability to compensate for incorrect viewing position differs between being off-axis (i.e., off to the side) and at the wrong distance (i.e., too far or too near). Compensation also varies between stereo and non-stereo pictures. For stereo pictures, both off-axis and wrong-distance viewing can create considerable perceptual distortions (Banks, Held & Girshick, 2009). For non-stereo pictures, Chapter 3 described experiments demonstrating that large perceptual distortions are created by viewing at the wrong distance. What about viewing non-stereo pictures from off to the side? In this case, there is strong evidence that the visual system can in part compensate and correct for perceptual distortions. Specifically, studies have shown that the visual system estimates the slant of the picture surface relative to the oblique line of sight and corrects for the expected foreshortening (Pirenne, 1970; Vishwanath, Girshick, & Banks, 2005; Rosinksi et al., 1980). This correction can remove a large amount of the distortions in the projection of the picture to the eye. However, the visual process underlying this compensation may not be unique to pictures. The ability to correctly perceive the dimensions of objects seen from various different viewing angles is called shape constancy (Wallach & Marshall, 1986). Consider, for instance, when reaching to grab a cup on a table. The foreshortened image of the top of the cup projects to an oval on the retina, but one still perceives the cup as circular, and adjusts the grasp for the correct shape. Similarly, when a rectangular picture is viewed obliquely, the shape of the picture's surface is still correctly perceived even though the retinal image is distorted to a trapezoid. This general shape constancy mechanism which is very useful in the natural environment, also makes it possible to view a picture obliquely and still perceive the 3D content shown in the picture as relatively undistorted. However, because pictures can contain content at many orientations relative to the picture surface, studies have shown that nearly complete compensation only occurs when the 3D content being judged is roughly parallel to the picture surface; when the content is roughly perpendicular to the surface, compensation is much less complete (Goldstein, 1987; Todorović, 2008).

So why can the visual system compensate for incorrect viewing angles but not incorrect viewing distances when interpreting the 3D layout shown in a picture? To compensate for an incorrect viewing distance, the visual system would have to estimate the correct distance from the picture's contents, not from the picture's surface. Such an estimation is possible, but is very prone to error (La Gournerie, 1859; Kubovy, 1986;

O'Brien & Farid, 2012). This type of estimation would also never be required in the natural environment, when the visual system is always viewing real 3D scenes from the correct distance by definition. On the other hand, shape constancy, which is akin to compensation for viewing angle, is highly useful in the natural environment. Thus, we argue that compensation for off-axis viewing occurs because the computations involved are useful in everyday vision. Compensation for incorrect viewing distance does not occur because the required computations are not useful in everyday vision and are prone to error. While the perception of pictures has many similarities to the perception of 3D scenes, the nature of pictures is to present images to the eyes that are often quite different from the images created by the natural environment.

References

Adams K. R. (1972). Perspective and the viewpoint. *Leonardo*, **5**, 209–217.

Amigo G. (1967). The stereoscopic frame of reference in asymmetric convergence of the eyes. *Vision Research*, **7**, 785–799.

Ankrum D. R., Hansen E. E., Nemeth K. J. (1995). The vertical horopter and the angle of view. In Gieco, A., Molteni, G., Piccoli, B., Occhipini E. (Eds.), *Work with display units '94*. Amsterdam : Elsevier.

Amigo G. (1974). A vertical horopter. *Optica Acta*, **21**, 277–292.

Ardito M. (1994). Studies of the influence of display size and picture brightness on the preferred viewing distance for HDTV programs. *SMPTE Journal*, **103**, 517–522.

Associated Press (2009). Inauguration Day coverage propels record viewership of AP's online video services [Press release]. Internet site: http://www.ap.org/pages/about/pressreleases/pr_012109a.html (Accessed January 18, 2012).

Banks M. S., Held R. T., Girshick A. R. (2009). Perception of 3-D Layout in Stereo Displays. *Information Display*, **25**(1), 12–16.

Banks M. S., Hooge I. T., Backus B. T. (2001). Perceiving slant about a horizontal axis from stereopsis. *Journal of Vision*, **1**(2):1, 55–79.

Belt A. F. (2008). *The elements of photography: Understanding and creating sophisticated images.* 2nd ed. Burlington, MA : Focal Press.

Bengston J. K., Stergios J. C., Ward J. L., Jester, R. E. (1980). Optic array determinants of apparent distance and size in pictures. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 751–759.

Birn J. (2008). Lighting & Rendering in Maya: Lights and Shadows, Internet site: http://www.3drender.com (Accessed March 11, 2010).

Blakemore C. (1970). The range and scope of binocular depth discrimination in man. *The Journal of Physiology*, **211**, 599–622.

Breitmeyer B., Battaglia F., Bridge J. (1977). Existence and implications of a tilted binocular disparity space. *Perception*, **6**, 161–164.

Burton H. E. (1945). The optics of Euclid. *Journal of the Optical Society of America*, **35**, 357–372.

Carlsson C., Walden P. (2007). Mobile TV—To live or die by content. *Proceedings of the 40th Hawaii International Conference on System Sciences*, **1**, 1530–1605.

Carroll R., Agarwala A., Agrawala M. (2010). Image warps for artistic perspective manipulation. *ACM Transactions on Graphics*, **29**(4), 1271–1279.

Choney S. (2009). Online video watching nearly doubles since '06. [Pew Internet]. Internet site: http://www.pewinternet.org/Media-Mentions/2009/Online-video-watching-nearly-doubles-since-06.aspx (Accessed September 22, 2011).

Cooper M. L., Pettigrew J. D. (1979). A neurophysiological determination of the vertical horopter in the cat and owl. *Journal of Computational Neurology*, **184**, 1–26.

Crone R. A., Everhard-Halm Y. (1975). Optically induced eye torsion: I Fusional cyclovergence. *Von Graefes Archiv für Klinische und Experimentelle Ophthalmologie*, **195**, 231–239.

Current I. (1990). Best viewing distance for photographers. *PSA Journal*, Sept, 16.

D'Amelio J. (2004). *Perspective Drawing Handbook.* New York : Dover Publications.

Goldstein E. B. (1987). Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. *Journal of Experimental Psychology*, **13**(2), 256–266.

Grove P. M., Kaneko H., Ono H. (2001). The backward inclination of a surface defined by empirical corresponding points. *Perception*, **30**, 411–429.

Hagen, M. A. (1974). Picture perception: Toward a theoretical model. *Psychological Bulletin*, **81**(8), 471–497.

Hagen M. A., Glick R., Morse B. (1978). The role of two-dimensional characteristics in pictorial depth perception. *Perceptual and Motor Skills*, **46**, 875–881.

Held R. T., Banks M. S. (2008). Misperceptions in stereoscopic displays: A vision science perspective. *Proceedings of the 5th Symposium on Applied Perception in Graphics and visualization*, 23–32.

Held R. T., Cooper E. A., O'Brien J., Banks M. S. (2010). Using blur to affect perceived distance and size. *ACM Transactions on Graphics*, **29**, 1–16.

Held R. T., Hui T. T. (2011). A guide to stereoscopic 3D displays in medicine. *Academic Radiology*, **18**, 1035–1048.

Helmholtz H. (1925). *Treatise on physiological optics.* In Southall J. P. C. (Ed.), (vol. III, pp.1–688). New York : Dover. (Original work published 1866).

Hibbard P. B., Bouzit S. (2005). Stereoscopic correspondence for ambiguous targets is affected by elevation and fixation distance. *Spatial Vision*, **18**, 399–411.

Hillis J. M., Banks M. S. (2001). Are corresponding points fixed? *Vision Research*, **41**, 2457– 2473.

Howard I. P., Ohmi M., Sun L. (1993). Cyclovergence: A comparison of objective and psychophysical measurements. *Experimental Brain Research*, **97**, 349–355.

Howard I. P., Rogers B. J. (2002). *Seeing in depth. Depth perception.* Ontario, Canada : I Porteuous Thornhill.

Julesz B. (1964). Binocular depth perception without familiarity cues. *Science*, **145**, 356–362.

Kingslake R. (1992). *Optics in photography.* Bellingham, WA : SPIE.

Kleffner D., Ramachandran V. (1992). On the perception of shape from shading. *Perception & Psychophysics,* **52***,* 18–36.

Knoche H., Papaleo M., Sasse M. A., Vanelli-Coralli A. (2007). The kindest cut: Enhancing the user experience of mobile TV through adequate zooming. *Proceedings of ACM Multimedia*, 123–29.

Knoche H., Sasse M. A. (2008). The sweet spot: How people trade off size and definition on mobile devices. *Proceedings of ACM Multimedia*, 21–30.

Koenderink J. J., van Doorn A. J., Kappers A. M. L. (1994). On so-called paradoxical monocular stereoscopy. *Perception*, **23**, 583–594.

Kubovy M. (1986). *The psychology of perspective and Renaissance art.* New York : Cambridge University Press.

La Gournerie J. D. (1859). *Traité de perspective linéare contenant les tracés pour les tableaux, plans et courbes, les bas-reliefs et les décorations théatrales, avec une théorie des effets de perspective.* Paris : Dalmont et Dunod.

Langer M. S., Bülthoff H. H. (2001). A prior for global convexity in local shape-from-shading. *Perception*, **30**, 403–410.

Lau E. (1921). Neue untersuchungen über das tiefen- und ebenensehen. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane Abteilung*, **53**, 1–35.

Ledgeway T., Rogers B. J. (1999). The effects of eccentricity and vergence angle upon the relative tilt of corresponding vertical and horizontal meridian revealed using the minimum motion paradigm. *Perception*, **28**, 143–153.

Liu Y., Bovik A. C., Cormack L. K. (2008). Disparity statistics in natural scenes. *Journal of Vision*, **8**(11):19, 1–14.

Liu B., Todd J. T. (2004). Perceptual biases in the interpretation of 3D shape from shading. *Vision Research*, **44**, 2135–2145.

London B., Stone J., Upton J. (2010). *Photography.* 10th ed. New Jersey : Prentice Hill.

Lumsden E. A. (1983). Perception of radial distance as a function of magnification and truncation of depicted spatial layout. *Perception & Psychophysics*, **33**(2), 177–182.

Lund A. M. (1993). The influence of video image size and resolution on viewing-distance preferences. *SMPTE Journal*, **102**, 406–415.

Modrak R., Anthes B. (2011). *Reframing photography*. New York : Routledge.

Nakayama K. (1977). Human depth perception. *Society of Photo-Optical Instrumentation Engineers Journal,* **120**, 2–9.

O'Brien J. F., Farid H. (2012). Exposing photo manipulation with inconsistent reflections. *ACM Transactions on Graphics*, **31**(1), 1–11.

Ogle K. N., Ellerbrock V. J. (1946). Cyclofusional movements. *AMA Archives of Ophthalmology*, **36**, 700–735.

Ogle K. N. (1950). *Researches in binocular vision*. Philadelphia : Saunders.

O'Shea J. P., Agrawala M., Banks M. S. (2010). The influence of shape cues on the perception of lighting direction. *Journal of Vision*, **10**(12):21, 1–21.

Pansell T., Schworm U., Ygge J. (2003). Torsional and vertical eye movements during head tilt dynamic characteristics. *Investigative Opthalmology & Visual Science*, **44**, 2986–2990.

Perona P. (2007). A new perspective on portraiture. *Journal of Vision*, **7**(9):992.

Pirenne M. (1970). *Optics, painting, & photography*. Cambridge : Cambridge University Press.

Potetz B., Lee T. S. (2003). Statistical correlations between 2D images and 3D structures in natural scenes. *Journal of the Optical Society of America*, **20**, 1292–1303.

Ray S. F. (2000). The geometry of image formation. In Jacobson R. E., Ray S. F., Atteridge G. G., Axford N. R. (Eds.), *The manual of photography: Photographic and digital imaging*, (pp. 39–57). Oxford : Focal Press.

Rosinski R. R., Mulholland T., Degelman D., Farber J. (1980). Picture perception: An analysis of visual compensation. *Perception & Psychophysics*, **28**, 521–526.

Samonds J. M., Potetz B. R., Lee T. S. (2012). Relative luminance and binocular disparity preferences are correlated in macaque primary visual cortex, matching natural scene statistics. *Proceedings of the National Academy of Sciences*, **109**(16), 6313–6318.

Schreiber K. M., Hillis J. M., Filippini H. R., Schor C. M., Banks M. S. (2008). The surface of the empirical horopter. *Journal of Vision*, **8**(3):7, 1–20.

Sedgwick H. A. (1991). The effects of viewpoint on the virtual space of pictures. In Ellis S. R., Kaiser M. K., Grunwald A. C. (Eds.), *Pictorial communication in virtual and real environments*, (pp. 460–479). London : Taylor & Francis.

Sedgwick H. A., Nicholls A. L. (1994). Distortions of pictorial space produced by optical minification. *Investigative Ophthalmology & Visual Science*, **35**, 2111.

Sedgwick H. A., Nicholls A. L., Brehaut J. (1995). Perceptual interaction of surface and depth in optically minified pictures. *Investigative Ophthalmology & Visual Science*, **36**, 667.

Shaffer J. (2011). The Facebook Blog: Bigger, faster photos. Internet Site: http://blog.facebook.com/blog.php?post=10150262684247131 (Accessed September 22, 2011).

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

Smith O. W., Gruber H. (1958). Perception of depth in photographs. *Perceptual and Motor Skills*, **8**, 307–313.

Song W., Tjondronegoro D. W., Wang S., Docherty M. J. (2010). Impact of zooming and enhancing region of interests for optimizing user experience on mobile sports video. *Proceedings of the International Conference on Multimedia*, **1**, 321–329.

Sun J., Perona P. (1998). Where is the sun? *Nature Neuroscience*, **1**, 183–184.

Take H. (2003) Market and technology trends in LCD TVs. *Sharp Technical Journal*, **4**,1–4.

Todorović D. (2005). Geometric and perceptual effects of the location of the observer vantage point for linear-perspective images. *Perception*, **34**, 521–544.

Todorović D. (2008). Is pictorial perception robust? The effect of the observer vantage point on the perceived depth structure of linear-perspective images. *Perception*, **37**, 106–125.

Todorović D. (2009). The effect of observer vantage point on perceived distortions in linear perspective images. *Attention, Perception, & Psychophysics*, **71**(1), 183–193.

Torralba A., Oliva A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, **14**, 391–412.

Torralba A. (2009). How many pixels make an image? *Visual Neuroscience*, **26**(1), 123–131.

Vishwanath D., Girshick A. R., Banks M. S. (2005). Why pictures look right when viewed from the wrong place. *Nature Neuroscience*, **8**, 1401–1410.

von Liebermann K. (1910). Beiträge zur Lehre von der binocularen Tiefenlokalization. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, **44**, 428–443.

Wallach H., Marshall F. J. (1986). Shape constancy in pictorial representation. *Perception & Psychophysics*, **39**, 233–235.

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

Wichmann F. A., Hill N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, **63**, 1293–1313.

Yang T., Kubovy M. (1999). Weakening the robustness of perspective: Evidence for a modified theory of compensation in picture perception. *Perception & Psychophysics*, **61**, 456–467.

Yang Z., Purves D. (2003). A statistical explanation of visual space. *Nature Neuroscience*, **6**, 632–640.