CHAPTER 14

Cues and Pseudocues in Texture and Shape Perception

Michael S. Landy, Yun-Xian Ho, Sascha Serwe, Julia Trommershäuser, and Laurence T. Maloney

INTRODUCTION

In estimating properties of the world, we often use multiple sources of information. For example, in estimating the three-dimensional (3D) layout of a scene, there are many sources of information or "cues" available for the estimation of depth and shape (Kaufman, 1974). These include binocular cues (disparity, vergence), motion cues (motion parallax, the kinetic depth effect), pictorial cues (texture, linear perspective, occlusion, etc.), and more. Human observers often combine these cues in a near-optimal fashion so as to maximize the precision of their estimates of scene layout (see Chapter 1, also Landy, Maloney, Johnston, & Young, 1995).

Some cues are acquired after birth. In infancy, the motion and binocular cues appear to develop around 3–4 months of age, with use of the pictorial cues beginning some 3 months later (Kellman & Arterberry, 1998, 2006). The onset of responses to binocular disparity appears to be based on maturation of the required vergence control and cortical architecture (binocular, disparity-tuned cells). On the other hand, it is not known whether the onset of the response to pictorial depth cues is the result of maturation or learning.

If a new depth cue is to be learned, the visual system has to note a correlation between values of that cue and other indicators of the values of the environmental variable being estimated. For example, consider the pattern of texture as a cue to slant. For observers to learn how to use texture, they must associate larger texture gradients (rapid changes in the size and density of texture elements across the image) or larger values of foreshortening (in which circles on the surface appear as eccentric ellipses in the retinal image) with larger values of surface slant. An observer could do so by noting the correlation of these image features with previously learned cues to surface slant such as the gradient of binocular disparities, or by using haptic cues as the observer handles the surface manually.

Backus and colleagues (see Chapter 6; Backus & Haijiang, 2007; Haijiang, Saunders, Stone, & Backus, 2006) investigated the visual system's ability to learn new depth cues. They did so by artificially pairing different values of a new, arbitrary "cue" with depth cues on which the visual system already relies. In their experiment, observers viewed Necker cubes that rotated about a vertical axis (Fig. 14.1). A Necker cube (Necker, 1832) is a picture of a transparent cube in which only the edges are drawn. It is an ambiguous figure; every few seconds the perception of the cube "switches" so that the face that was previously seen as behind is now perceived to be in front, and vice versa. When such a figure is rotated, the perceived direction of rotation reverses along with perceived depth.

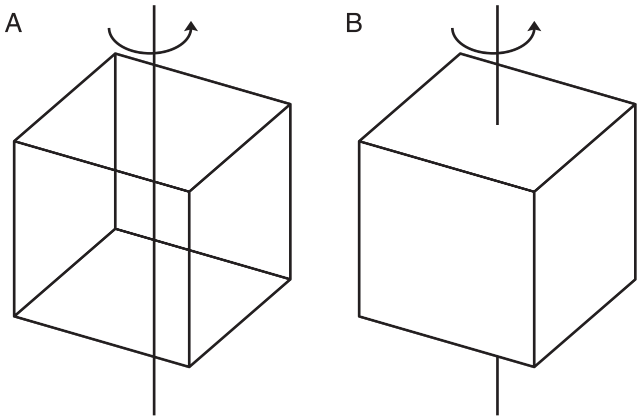

Figure 14.1 The experiments of Backus and colleagues (Chapter 6; Backus & Haijiang, 2007; Haijiang et al., 2006). (A) In these experiments, observers viewed rotating Necker cubes that spontaneously reverse in perceived depth and rotation direction. (B) In a training session, unambiguous cubes (with occlusion as shown here as well as binocular disparity) were shown, and rotation direction was paired with a new cue (e.g., position in the display). After training, perception of rotation direction of ambiguous stimuli was affected by the value of the newly recruited cue.

In their experiment, the direction of rotation was disambiguated by including disparity and occlusion cues (the cube was now opaque, so the rear face was not visible, and binocular disparities were consistent with the pictorial depth; Fig. 14.1B). In addition, another aspect of the display varied in concert with the rotation direction. For example, if the cube rotated front to the right, it was shown in the upper half of the display. If it rotated front to the left, it was shown in the lower half of the display. After a sequence of such learning trials, the experimenters presented test trials with no disparity or occlusion cues. These displays consisted of standard rotating Necker cubes, which as described previously are normally ambiguous as to rotation direction. However, after learning, the previously irrelevant "cue", that is, the position in the display (upper vs. lower), had a substantial influence on perceived rotation direction. This "cue recruitment" paradigm is analogous to conditioning; however, the conditioning results in an increased frequency of occurrence of a particular percept rather than an overt behavior.

The recruited cues explored by Backus and colleagues are all binary; the cue is either a higher or lower display location, movement in the left or right direction, and so forth. They can be used to disambiguate a rotating cube but cannot, in and of themselves, provide a cue that could be used to quantify depth. The strength of these cues can be measured experimentally, and indeed these cues can trade off against trusted cues such as binocular disparity to determine which of two possible percepts (front rightward vs. front leftward) is obtained (Backus & Haijiang, 2007). However, by themselves they do not provide a continuously valued estimate of depth that combines with other depth cues. Note that binary depth cues such as occlusion can also trade off against continuously valued depth cues (e.g., binocular disparity) in the perception of depth (Burge, Peterson, & Palmer, 2005).

In this chapter, we review several studies that provide evidence that the visual system uses additional cues beyond those traditionally discussed for the estimation of scene properties, including shape, depth, and 3D surface texture. Figure 14.2 shows several example stimuli from one of these studies (Ho, Landy, & Maloney, 2006) in which observers were asked to judge surface roughness. These computer-rendered stereograms depict bumpy, pitted surfaces that include several cues to depth and shape. Figure 14.2B depicts a larger range of depth

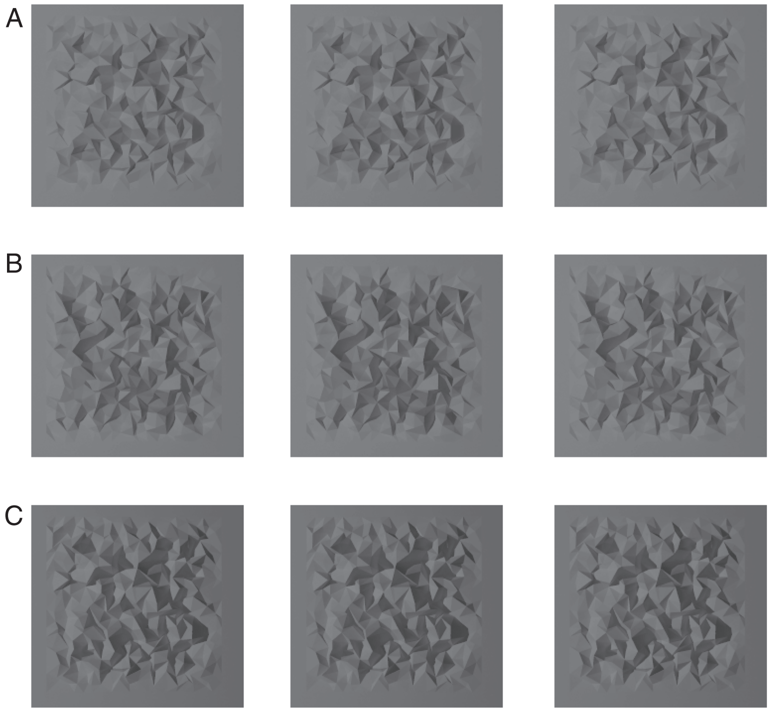

Figure 14.2 Cast shadows as a pseudocue. (A) This stereogram shows a sample stimulus from the study of Ho and colleagues (Ho et al., 2006). A random bumpy surface was computer-rendered using a combination of ambient illumination and a single point source. (B) If we increase the range of depth in the surface, the amount of the image in cast shadow increases. (C) If, instead of increasing the range of depth, we move the light source to a more glancing angle relative to the surface, the amount of cast shadows also increases. As a result, perceived depth and surface roughness increase as well, even though there was no change of rendered depth as compared to (A). (Reprinted from Ho et al., 2006. Copyright ARVO; used with permission.)

values across the bumps and valleys, and hence appears rougher than Figure 14.2A. Figure 14.2C appears rougher than Figure 14.2A as well, but in this case the parameters used to generate the bumpy surface are identical. Rather, all that has changed is the pattern of illumination. In Figure 14.2C, the point illuminant is located such that the surface is illuminated from a more glancing angle, resulting in more and deeper cast shadows. Relative to Figure 14.2A, both Figures 14.2B and 14.2C have an increase in the amount of visible cast shadow, and both appear to be rougher surfaces than Figure 14.2A.

We refer to the proportion of the surface in cast shadow (and several related image statistics) as a "pseudocue" for the estimation of both surface roughness (in Fig. 14.2) and depth (for an experiment we discuss below). We call this a pseudocue because it is a cue to an object or surface property that confounds changes in that property with changes in irrelevant properties of the surface, object, or viewing environment. The proportion of cast shadow is a pseudocue for surface roughness in the sense that an increase in the variance of depths of textural elements comprising the surface, that is, an increase in roughness, results in an increase in the amount of cast shadow given that all other variables in the scene are constant. But a change in light-source position can also increase the value of this pseudocue, with no concomitant change in physical roughness. This pseudocue would be a valid cue to roughness if all aspects of the scene except for the range of depth values were fixed, including the position of the illuminant, the location of the observer relative to the surface, and the class of surface geometry. However, observers appear to use this pseudocue as though it were a valid tool for estimating surface roughness across varying illumination and viewing conditions, resulting in failures of "roughness constancy."

This definition of *pseudocue* is related to the idea of *cue promotion* described by Landy and colleagues (1995). For example, the raw data for binocular stereopsis are the binocular disparity values for various points in the retinal images. These data do not provide estimates of depth or surface slant in and of themselves; rather, they must be *promoted* to depth estimates after scaling disparities based on additional parameters. These additional parameters describe the viewing geometry and include ocular vergence, version, and torsion. Use of the raw disparity values as a direct indication of depth results in failures of depth constancy with changes of gaze.

Suppose we were able to show that observers used the mean luminance of stimuli such as those in Figure 14.2 as an indication of the range of depth in the stimulus or of surface roughness. Certainly, as the textural elements of the surface in Figure 14.2 are stretched out in depth, more shadows appear and the overall mean luminance of the display is reduced. Yet a reduction of mean luminance can result from many changes in the scene, including a reduction in overall illumination, a change of viewpoint, a change in illuminant position, or decreased surface reflectance. Thus, a pseudocue differs from a standard cue in that information provided by a pseudocue is not invariant across viewing and environmental conditions and when used can lead to misjudgments of a given object or surface property. A standard cue is one for which the observer can estimate the environmental parameters needed to interpret the raw data, and hence the observer can compute an estimate of an object or surface property independent of extraneous viewing or environmental conditions. The distinction between a standard cue and a pseudocue is one of degree; to the extent that the required environmental parameters are not available or used by the observer, the cue is less useful for estimating object or surface properties and hence is more of a "pseudocue."

In this chapter, we review experimental evidence concerning the use of such pseudocues in the perception of 3D scene properties such as roughness and depth. In the first two experiments, observers judged surface roughness of an irregular surface in which roughness was varied by scaling the range of depths of bumps and valleys. We show that observers do indeed use pseudocues such as the amount of shadow and as a result they misperceive surface roughness. We also briefly note that a similar phenomenon occurs in the perception of surface gloss. Finally, we summarize a third study in which observers judged the depth of a single bump. In this final experiment, we investigated how pseudocues might be learned and, in particular, how observers determine how much weight to give to a pseudocue in combining it with other depth cues.

CUES AND PSEUDOCUES FOR TEXTURE ROUGHNESS

In this section we review the results of two studies (Ho et al., 2006; Ho, Maloney, & Landy, 2007) that suggest observers use pseudocues such as the amount of cast shadows as a depth cue in judgments of surface roughness. The stimuli were like those in Figure 14.2. We test whether judged roughness increases with the amount of cast shadow caused by changes in the position of the light source (Fig. 14.2C) or the position of the observer relative to the surface. Thus, we ask whether "roughness constancy" holds for this task and these viewing conditions.

Methods

The methods are summarized in Figure 14.3. Stimuli such as those shown in Figure 14.2 were generated as follows. We began with a 20×20 array of points, 19×19 cm in size (Fig. 14.3A). First, grid intersections were randomly jittered within the grid plane (Fig. 14.3B). Next, grid intersections were randomly jittered along the z-axis (orthogonal to the grid plane).

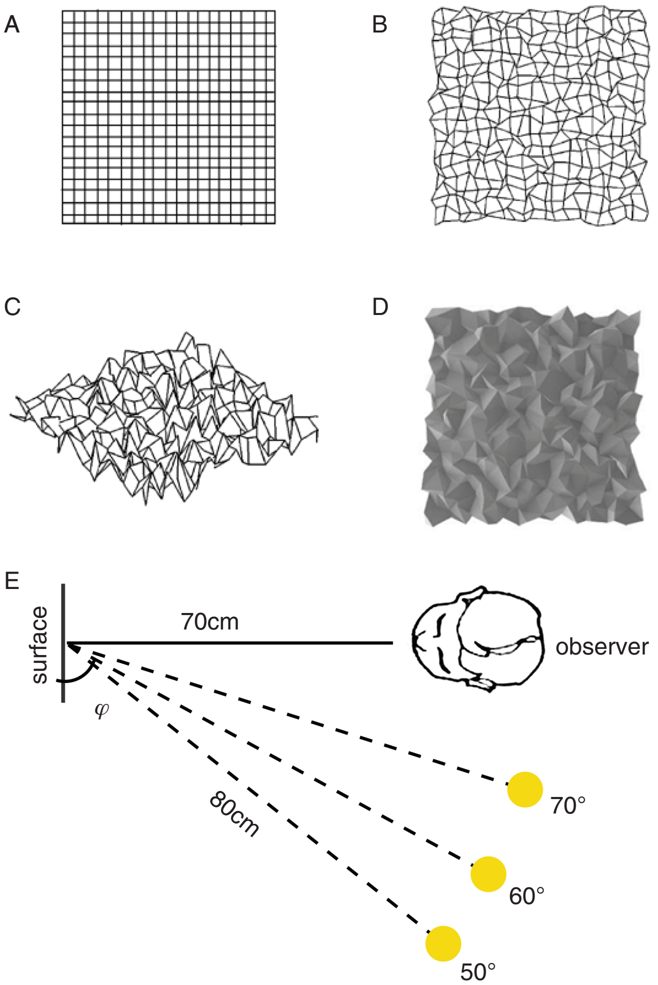

Figure 14.3 Stimulus construction. (A) Stimuli were constructed based on a 20 × 20 grid of points. (B) Grid intersections were randomly jittered in the x and y directions (within the grid plane) and then (C) in the z direction. The pair of points at the ends of a randomly chosen diagonal were connected (for each grid square) and then (D) the stimulus was rendered using a combination of ambient and point-source illuminants. (E) The point source was located to the left of the observer, at one of three possible angles relative to the surface. (Reprinted from Ho et al., 2006. Copyright ARVO; used with permission.)

The z-values were drawn from a uniform distribution. The roughness level r of a given stimulus corresponded to the range of this distribution. Each grid "square" was then split into two triangles by randomly connecting one of the two "diagonals" (Fig. 14.3C). The resulting triangulation was then rendered using the RADIANCE rendering software (Larson & Shakespeare, 1996; Ward, 1994) resulting in an image like that shown in Figure 14.3D. Eight values of r were used, resulting in surfaces with depth ranges varying from ±0.6 to ±40 mm, as viewed from a distance of 70 cm.

For the first study we review here, the illumination environment consisted of an ambient illuminant as well as a single-point illuminant located to the left of the observer at one of three angles φ relative to the surface (Fig. 14.3E).

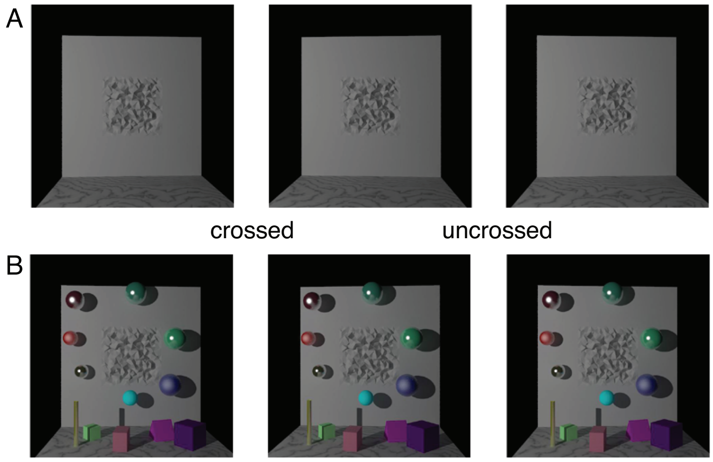

Figure 14.4 Sample stimuli. (A) A sample stereogram for the condition with no additional context. (B) A sample stereogram for the condition in which additional objects were placed in the scene to provide cues to the position of the point source. (Reprinted from Ho et al., 2006. Copyright ARVO; used with permission.)

The bumpy surface was embedded smoothly into a wall attached to a two-dimensional (2D) textured ground surface (Fig. 14.4A). In one condition, additional context was added to the scene, that is, matte and glossy objects placed randomly about in the scene, so as to provide additional image cues to the position of the point light source (Fig. 14.4B). Stimuli were viewed on a custom stereoscope with the left and right eye's images projected onto separate cathode ray tube (CRT) displays located to the left and right of the observer.

The observer's task was a two-interval forced-choice roughness discrimination. On each trial, two stimuli were displayed in sequence. In one of the two temporal intervals, chosen randomly, a "test" stimulus was shown with a particular combination of roughness level and point-illuminant position. In the other interval a "match" stimulus was shown. The match stimulus had a different point-illuminant position and its roughness level was adjusted across trials by a staircase procedure. The observer indicated which of the two stimuli appeared to be rougher. There were interleaved staircases corresponding to several test patch roughness levels and combinations of test-match illuminant positions.

Results: Use of Pseudocues

Figure 14.5 illustrates the analysis of the resulting data. In Figure 14.5A, the probability that an observer chose the match stimulus as appearing rougher is plotted as a function of the match stimulus roughness level for a test stimulus with roughness of $r = 2.25$ cm. The greater the rendered roughness was of the match stimulus, the more often it was perceived as rougher than the test stimulus. We fit a Weibull distribution to each such psychometric function and determined the "point of subjective equality" (PSE), the point at which we estimated the observer would judge the match stimulus to be rougher than the test stimulus 50% of the time. In other words, the PSE is an estimate of the test and match roughness levels of two surfaces under different illumination conditions that would be judged to be equally rough.

Figure 14.5B shows PSEs for one observer and one combination of test and match point-source locations ($\varphi_{\text{test}} = 70°$ and $\varphi_{\text{match}} = 50°$). For each test-patch roughness level we plot the PSE, that is, the match roughness level that, when illuminated by φ_{match}, appears as rough as the test roughness illuminated by φ_{test}. If the change in illumination position had no effect on perceived roughness, that is, if the observer

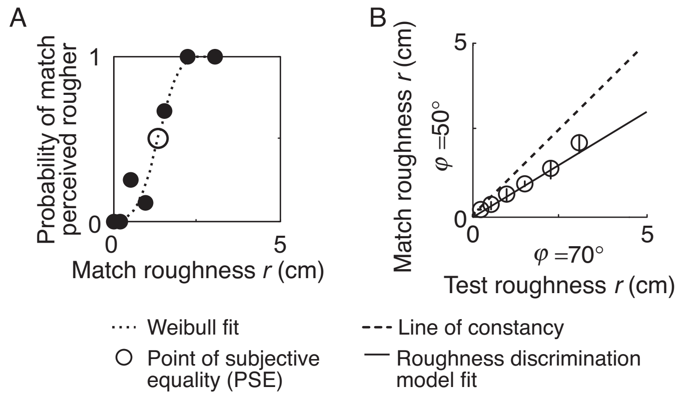

Figure 14.5 Sample data. (A) Psychometric function for subject JG for a test patch with roughness 2.25 cm under point source $\varphi_{\text{test}} = 70°$ and match stimuli under point source $\varphi_{\text{match}} = 50°$. The open circle indicates the point of subjective equality (PSE) derived from a Weibull fit to the data (dotted line). (B) PSEs for this subject and combination of test and match illumination conditions. The dashed line indicates the line of constancy, that is, expected results for a roughness-constant observer. The solid line was fit to the data under the constraint that it pass through the origin. The slope of this line (the "PSE slope") was used to summarize each data set. (Reprinted from Ho et al., 2006. Copyright ARVO; used with permission.)

were "roughness constant" across this change in viewing conditions, then PSEs should fall roughly on the dashed identity line. Clearly, roughness constancy was not demonstrated because many PSEs fell significantly below the identity line, and a line fit to the PSEs and forced to pass through the origin had a slope significantly less than one (see Ho et al., 2006 for details on fitting).

Figure 14.6 summarizes all of our results by showing the slope of the fit PSE lines (as in Fig. 14.5B) for seven subjects and all three possible illuminant-position comparisons. The abscissa plots the PSE slope for the no-context condition (Fig. 14.4A) and the ordinate plots the corresponding PSE slope for the context condition (Fig. 14.4B). The dotted lines indicate the values of PSE slope corresponding to roughness constancy (i.e., a slope of one). Nearly all conditions resulted in slopes less than one. This means that a decrease of illuminant angle φ, corresponding to a more glancing point-source illumination and hence more cast shadows, led observers to perceive the stimulus as rougher than the identical stimulus lit with a more direct illuminant. Second, nearly all points are close to the identity line. This indicates that the additional context provided by objects in the scene was ineffective in helping subjects maintain a more constant representation of roughness. These additional objects provided cast shadows (on the ground plane) and highlights which could have provided the observer with cues to illuminant location. We hypothesized that observers would make better use of the shadow pseudocue if they had a better estimate of the pattern of illumination. Instead, the addition of these contextual cues had little effect on judgment and did not improve the degree of roughness constancy.

This suggests that the estimation of surface properties is primarily based on cues that are spatially local to the object being judged. This is consistent with the results of Hillis and colleagues (Hillis, Watt, Landy, & Banks, 2004). They examined combinations of texture and disparity cues to surface slant. The relative reliability of these two cues depends on a number of scene factors, including absolute distance to the object and the actual value of slant. As a result, for large surfaces, the weight given to each cue is determined locally and can change across the surface. They provided evidence that cue weights can indeed vary across a large

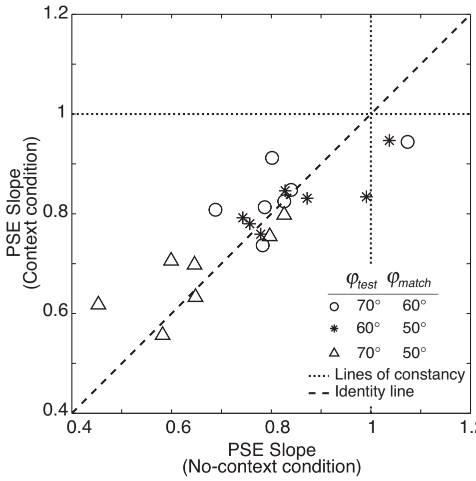

Figure 14.6 Data from the first experiment. For each subject and illuminant comparison, we plot the point of subjective equality (PSE) slope from the no-context versus the context conditions. Different symbols indicate the illuminant comparison. Data are shown for seven subjects. The dotted lines indicate PSE slopes of one, that is, the expected result for a roughness-constant observer. The diagonal dashed identity indicates results for which the addition of context objects had no effect. (Reprinted from Ho et al., 2006. Copyright ARVO; used with permission.)

extended surface. As a result, a large planar surface with conflicting texture and disparity cues to depth appears curved as the local estimate approaches that from disparity or from texture in different regions of the stimulus.

In a subsequent study (Ho et al., 2007), we extended these results by varying the observer's viewing position relative to the rendered textured surface. The methods were similar to the study we just described. However, rather than comparing pairs of point-source locations, we compared pairs of observer viewpoints φ_v. The point-source location was fixed in the scene at $\varphi_p = 60°$. Pairs of test and match viewpoints were chosen from among those illustrated in Figure 14.7A, resulting in stimuli like those shown in Figure 14.7B.

Figure 14.7C shows a summary of the data for the four observers and the subset of viewpoint comparisons for which both viewpoints were located to the right of the point-source location (i.e., both test and match viewpoints $\varphi_v > 60°$). For this subset of viewpoints, the larger the value of φ_v the more cast shadows are present in the image; hence, if the results of the previous experiment apply, surfaces viewed from larger viewing angles as defined here should appear rougher. Indeed, for most subjects and viewpoint pairs, PSE slopes were less than one.

Discussion

In the two studies we have reviewed, observers displayed significant failures of roughness constancy. That is, the roughness of a surface appeared to vary with changes in the viewing conditions (change in the position of a point light source or the observer), even though there was no change in the physical rendering of the surface being judged. In all cases, apparent roughness increased with increases in the amount of visible shadow. Does this mean that the amount of visible shadow, or some related image statistic, is used by observers as a cue to 3D surface properties like roughness that contain variations in depth or, more generally, to depth itself?

To answer this question, we modeled the results as a combination of appropriate, veridical cues (disparity, contour, etc.) and putative pseudocues. The possible pseudocues we investigated included the proportion of the image in cast shadow, the mean luminance of the nonshadowed regions, the standard deviation of luminance of the nonshadowed regions, and *texture contrast* (Pont & Koenderink, 2005). These image statistics were computed for each stimulus image (resulting in the values R_p, R_m, R_s and R_c, respectively). In addition, the veridical roughness (i.e., depth range in these experiments) was presumably signaled to the observer by disparity and other cues (R_d). We assumed that perceived roughness was a linear combination of the veridical cues and pseudocues:

$$R = w_d R_d + w_p R_p + w_m R_m + w_s R_s + w_c R_c, \tag{14.1}$$

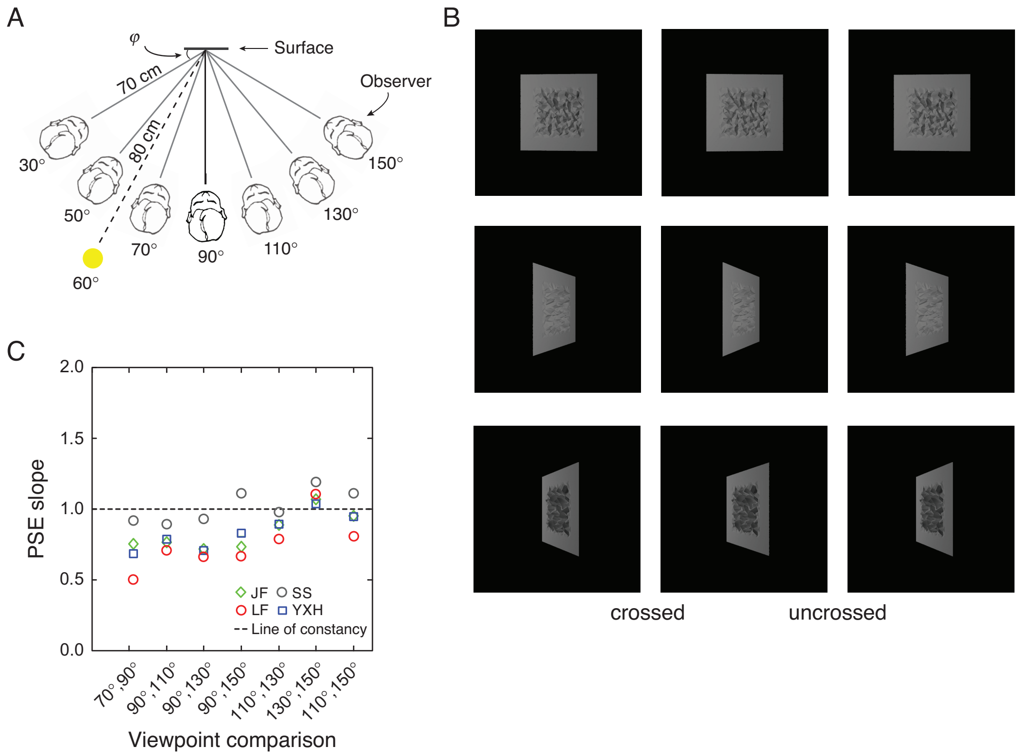

Figure 14.7 Effect of viewpoint. (A) In this experiment, the light source position φ_p was fixed at 60° and the viewpoint position φ_v was varied. (B) Example stereogram stimuli. (C) Point of subjective equality (PSE) slopes for four observers and all viewpoint comparisons in which test and match viewpoints were to the right of the light source. The dashed line indicates a PSE slope of one, that is, the expected result for a roughness-constant observer. (Reprinted from Ho et al., 2007. Copyright ARVO; used with permission.)

where the values w_i combine both scale factors and weights and need not sum to one. At the PSE, a second stimulus (viewed under a different lighting condition or viewpoint) has identical perceived roughness:

$$R = w_d R'_d + w_p R'_p + w_m R'_m + w_s R'_s + w_c R'_c. \tag{14.2}$$

Rearranging terms, we find

$$\Delta R_d = a_p \Delta R_p + a_m \Delta R_m + a_s \Delta R_s + a_c \Delta R_c, \tag{14.3}$$

where $a_i = -w_i/w_d$ for each pseudocue i and $\Delta R_i = R_i - R'_i$. Thus, the model consisted of a regression that predicted failures of roughness constancy (ΔR_d, i.e., displacements of PSEs away from the lines of constancy in Fig. 14.4B) as a linear combination of differences of the values of the pseudocues at the PSE. For almost all observers, a significant percentage of the variance in PSEs was accounted for by a subset of these pseudocues. For example, in the first study between 33% and 82% of the variance in PSEs was accounted for by a linear combination of the four candidate pseudocues (Ho et al., 2006), depending on the observer. In other words, observers did indeed appear to treat a stimulus that contained more shadows and was darker and higher in contrast as being rougher, even though that change in pseudocue resulted from changes in viewing conditions alone.

As an aside, in a third study we also identified other extraneous cues that can potentially act as pseudocues to judgments of surface properties.

In this study, observers compared pairs of surfaces that varied in bumpiness and/or degree of gloss (Ho, Landy, & Maloney, 2008). We found that the two surface properties interacted in the sense that an increase in surface gloss made the surface appear to be bumpier, and an increase in surface bumpiness made the surface appear glossier. This is another example of a failure of constancy where the percept of one property is altered when another independent property is varied. Although we did not directly investigate what particular image statistics correlated with observers' judgments in this study, results suggest that pseudocues may play a role in the perception of gloss and bumpiness.

REWEIGHTING CANDIDATE PSEUDOCUES TO DEPTH

In the previous section, we reviewed several experiments demonstrating failures of roughness constancy across changes in viewing conditions. In particular, observers perceived stimuli as rougher when the viewpoint or pattern of illumination resulted in an image containing more cast shadows. We referred to image statistics such as the amount of cast shadow as pseudocues because such statistics vary with changes in roughness itself but also change with extraneous changes in viewing conditions or surface characteristics.

To acquire a new cue (or pseudocue), the visual system presumably needs to experience the correlation of that cue with the scene property being estimated, possibly via a "trusted" cue the visual system already uses to make that judgment. In the case of shadow or other image pseudocues, haptic input might be such a trusted cue. Arguably, for the estimation of surface roughness, manual exploration of a surface provides a more reliable estimate of surface roughness than vision and thus might be expected to dominate visual cues to roughness (Klatzky, Lederman, & Matula, 1991, 1993; Lederman & Abbott, 1981; Lederman & Klatzky, 1997). As a result, one might expect that "touch educates vision" (Berkeley, 1709).

There are at least two ways an observer might respond to changes in the relationship between two cues: recalibration and reweighting. Recalibration is needed if the estimate based on a cue becomes biased due to growth or other changes in the visual apparatus. For example, to estimate depth, binocular disparity must be scaled by an estimate of the viewing distance (i.e., based on the vergence angle) and other aspects of the viewing geometry (including the distance between the eyes: the pupillary distance, or PD). The PD increases by approximately 50% (from 4 to 6 cm) across the first 16 years of life (MacLachlan & Howland, 2002) requiring a substantial recalibration of binocular stereopsis. Reweighting should occur, on the other hand, as the relative reliability of cues changes (see Chapter 1 and Landy et al., 1995). Evidence for reweighting in response to changes in cue correlation has been found for combinations of haptic and visual depth cues (Atkins, Fiser, & Jacobs, 2001), haptic and visual cues to surface slant (Ernst, Banks, & Bülthoff, 2000), and pairs of visual cues to depth (Jacobs & Fine, 1999).

Logically, for the visual system to incorporate a new cue (or pseudocue) in its estimation of a scene property, both reweighting and recalibration must be involved. The new cue should receive a weight proportional to its reliability, which can be estimated using the degree it correlates with trusted cues. Calibration is required as well—for a new cue it is not yet *re*calibration—so that the raw cue measurement can be transformed to provide an estimate that is unbiased.

We next describe an experiment that illustrates how pseudocues, like more typical visual cues, undergo reweighting (Ho, Serwe, Trommershäuser, Maloney, & Landy, 2009). We created a correlation between a pseudocue (the amount of cast shadow) and haptic cues to depth. To provide the haptic cues we used a PHANToM 3D Touch interface (SensAble Technologies, Woburn, MA). This device does not provide realistic haptic sensory input for fine-grained surface material properties such as roughness. Therefore, we instead asked subjects to judge a larger scale scene property that potentially involves the use of similar pseudocues: the depth of an object. We determine whether this pseudocue is given greater weight in depth judgments made using vision alone after exposure to stimuli with haptic depth correlated with the shadow pseudocue.

Methods

Participants saw and/or felt the displays in a virtual environment. The displays portrayed a gaze-normal plane with a portion of a vertically oriented circular cylinder projecting out of the plane (Fig. 14.8). In haptic conditions, observers placed their right index finger into a thimble attached to the PHANToM, which provided force feedback to the participant as the participant moved the index finger back and forth across the virtual object. The PHANToM was programmed to simulate a cylinder made of a hard, rubber-like material. The visual display consisted of a virtual image of the object (reflected in a mirror positioned above their hands), rendered so as to appear in the same location as the haptic display. The visual display was viewed binocularly with appropriate disparities using CrystalEyes 3 (Stereographics, Beverly Hills, CA) liquid-crystal shutter glasses. A small virtual sphere was displayed and updated in real time to represent the current 3D location of the index finger, although this sphere vanished whenever the observer's finger was positioned in the region of the stimulus so that the cursor itself was never a cue to object depth.

The rendered object was a 4 cm wide portion of a vertical circular cylinder and looked like a bump. The depth of the bump was determined by the curvature (the radius of the circular cylinder, only a portion of which was visible emerging from the 4×4 cm window in the background plane). The background plane and cylinder were covered in 2D texture so as to provide strong binocular disparity and texture cues to depth. In addition, the objects were rendered with a combination of an ambient and point source illuminant. The point source was located above the observer; for test stimuli the point source was located at an angle $\varphi_{\text{test}} = 45°$ above the background plane, and for match stimuli, $\varphi_{\text{match}} = 30°$. As a result, the shapes cast a curved shadow at the bottom of the cylinder (Fig. 14.8) whose size was larger for cylinders with greater depth but also for match stimuli due to the more grazing angle of illumination.

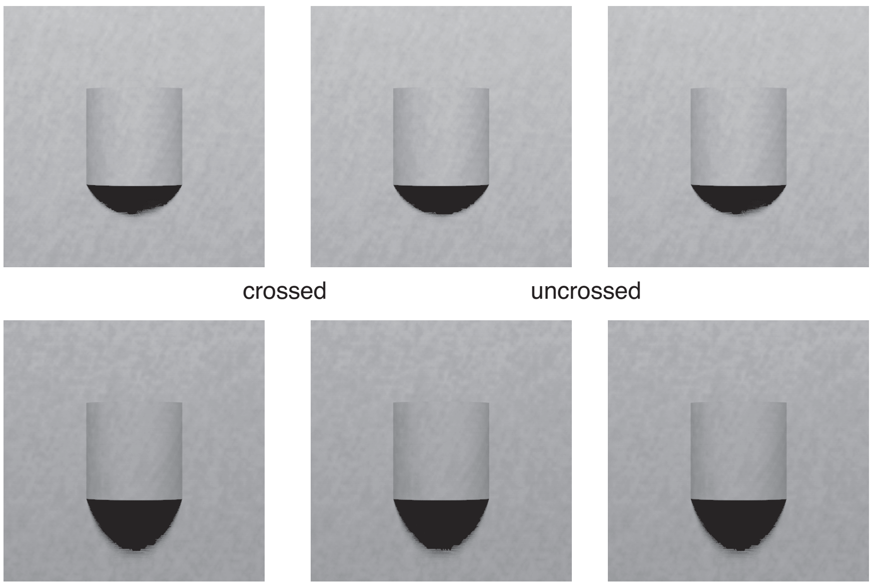

Figure 14.8 Example stereogram stimuli from the visuohaptic experiment. (Top) Test stimulus with depth = 12 mm and point-illuminant direction $\varphi_{\text{test}} = 45°$. (Bottom) Match stimulus with the same depth and $\varphi_{\text{match}} = 30°$. (Reprinted from Ho et al., 2009; used with permission.)

A similar task and procedure was used as in the roughness experiments, although for this experiment the participant was asked to judge whether the test or match stimulus had greater depth. Each session consisted of 200 trials: 40 for each of 5 test bump depth levels. Match bump depth levels were controlled by interleaved staircases.

There were three types of sessions: visual-only, haptic-only, and visuohaptic training. In the visual-only session, objects were visible but could not be felt. These sessions allowed us to determine the extent to which a participant used the pseudocue (the size of the cast shadow). In the haptic-only session, participants could feel the objects but they were not displayed visually. These sessions were primarily used to familiarize the participants with the virtual haptic experience.

The key experimental manipulation in this study was the introduction of visuohaptic training. In visuohaptic training sessions, the rendered haptic cue was altered so that it was perfectly correlated with the shadow pseudocue. Figure 14.9 shows the proportion of the image corresponding to the cast shadow as a function of bump depth for both the test and match stimuli. For each bump depth level, we computed the average proportion of cast shadow (averaged across the test and match stimuli) and fit the resulting data with a smooth curve (dashed line). This function was then used as a lookup table to determine the amount of depth portrayed by the haptic stimuli in visuohaptic training sessions. Thus, in these sessions, a test and match stimulus were shown in succession, but the rendered haptic depth was determined by the amount of shadow in each stimulus, rather than the veridical value of depth corresponding to the disparity and texture cues in the stimulus.

The experiment was carried out over the course of 4 days, with at most a 2-day break between successive sessions. The sessions were laid out as follows (numbers in parentheses are identifying session numbers used in Fig. 14.10):

Day 1: Visual-only practice session, two visual-only sessions (1–2)
Day 2: Two visual-only sessions (3–4), one haptic-only session
Day 3: One visuohaptic training session (5), one visual-only session (6)
Day 4: One visuohaptic training session (7), one visual-only session (8).

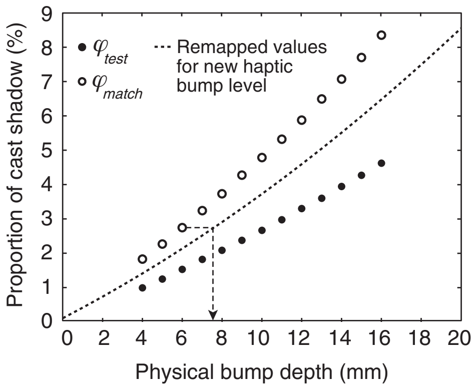

Figure 14.9 Method of artificially correlating the haptic cue and shadow pseudocue. Plotted is the percentage of the stimulus in cast shadow as a function of bump depth (open circles: match stimuli; filled circles: test stimuli). The dashed line shows a polynomial fit to the average percentage of cast shadow for each bump depth. In visuohaptic training trials, this curve was used as a lookup table to determine the haptically rendered depth (arrows). (Reprinted from Ho et al., 2009; used with permission.)

Results: Pseudocue Weights Can Change with Experience

The data from this experiment were analyzed in a manner identical to that used for the roughness experiments (Fig. 14.5), resulting in an estimated PSE slope for each session. To determine whether a subject's initial percept of depth changed with varying illumination, that is, failed to be constant across illumination conditions, we took the average of the four PSE slopes obtained for the visual-only sessions prior to visuohaptic training and compared it to a slope of one (i.e., shape constancy). Of the 12 subjects, 6 had average PSE slopes that were significantly less than one. In other words, for these subjects the pseudocue of shadow size was effective in producing a failure of size constancy across changes in the pattern of illumination.

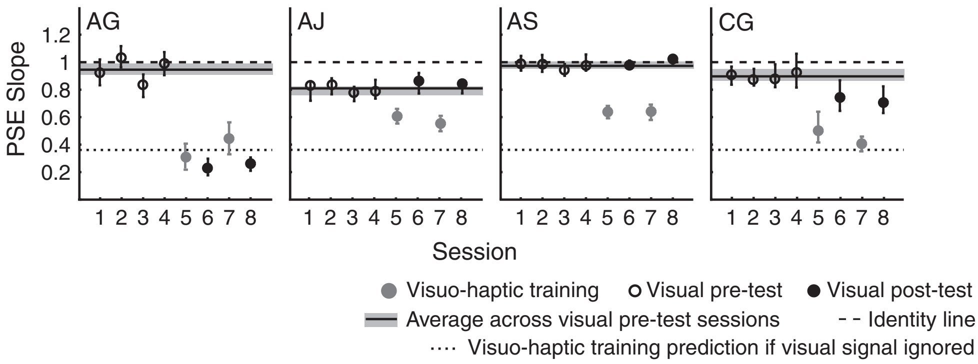

Figure 14.10 Results of visuohaptic training. Point of subjective equality (PSE) slopes are plotted for four observers (out of the 12 observers in the study) for each session. The dashed line indicates a PSE slope of one, that is, the expected result for a roughness-constant observer. The dotted line indicates the PSE slope expected if only the haptic and/or pseudocues were used. The first four visual-only sessions (open symbols) were pretraining. The solid horizontal line and corresponding shaded 95% confidence region reflect the average PSE slopes for these first four sessions. Two of these four observers (AJ and CG) showed significant failures of depth constancy (PSE slopes < 1). All four subjects gave substantial weight to the haptic cue during visuohaptic training sessions (gray filled symbols). In subsequent visual-only sessions, some subjects (AJ and AS) reverted to their previous visual-only behavior while others (AG and CG) showed significantly less depth constancy, consistent with giving an increased weight to the shadow pseudocue as a result of visuohaptic training. (Reprinted from Ho et al., 2009; used with permission.)

For most subjects, PSE slopes were significantly smaller during visuohaptic training sessions (for example subjects, see Fig. 14.10). If a subject ignored the visual displays and responded only based on the haptic input during these sessions, this would have resulted in a PSE slope of 0.36 (dotted horizontal line in Fig. 14.10) due to the mismatch of rendered haptic depth for test and match stimuli (Fig. 14.9). Thus, most subjects gave substantial weight to the haptic cue during the visuohaptic training sessions.

The key question is whether visuohaptic training sessions resulted in an increase in the weight observers gave to the pseudocue of shadow size. An increase in this weight would imply less depth-constant performance, that is, a reduction of PSE slope. There was substantial variation in performance across subjects. Five out of 12 subjects showed a significant reduction in PSE slope after training compared to pretraining slopes (e.g., subjects AG and CG), while other subjects returned to pretraining behavior (e.g., subjects AJ and AS).

We found substantial individual differences in response to these conditions. Of the six subjects who initially showed a significant use of pseudocues (i.e., significant failure of depth constancy with pretraining PSE slopes significantly less than one), two of them increased their reliance on pseudocues after visuohaptic training. The other half of the subjects were roughly depth constant before training, and three out of these six subjects showed a significant response to pseudocues after training.

Discussion

Individual differences are frequent when cue weights are measured in depth-cue-integration tasks. In some cases, this outcome is consistent with optimal cue integration with the individual differences stemming from differences in cue reliability across individuals (Hillis, Ernst, Banks, & Landy, 2002; Hillis et al., 2004; Knill & Saunders, 2003). In other cases, individual differences may be an indication of differences in cue correlation or suboptimal cue integration (Oruç, Maloney, & Landy, 2003). Thus, the variation across subjects of the use of pseudocues was to be expected.

Although the effect occurred in only a subset of our subjects, this experiment demonstrates that a correlation between a pseudocue and a trusted cue can result in an increase in the weight given to the pseudocue in subsequent estimates. From the optimal-integration standpoint, it is as though the correlation with the trusted cue is taken as evidence that the reliability of the pseudocue has increased. To test such a claim would require experiments in which the pseudocue is paired with cues with a range of reliabilities so as to estimate the reliability of the pseudocue based on the weight it was given (assuming optimal cue combination; see Chapter 1, Eqs. 1.1–1.2).

CONCLUSIONS

We have provided evidence that human observers use pseudocues in the estimation of surface roughness and shape, and that the weight given to pseudocues in estimating 3D scene properties can be altered by experience, in particular if the pseudocue is highly correlated with a trusted cue (in this case, haptic cues).

The pseudocues we consider here are simple image statistics that vary with the scene property being estimated but also vary with other, irrelevant aspects of viewing conditions. We do not assume that pseudocues must be drawn from the class of image statistics. When they are, though, the use of such pseudocues can be regarded as a visual-system heuristic or trick: A correlation between a simple image statistic and a scene property is noted and then recruited as a cue to that property. These pseudocues are related to other image statistics used to estimate scene properties without solving the ill-posed inverse-optics problem. Other examples of such statistics are the use of luminance histogram skew as a cue to surface gloss (Motoyoshi, Nishida, Sharan, & Adelson, 2007) and luminance histogram moments and percentile statistics as a cue to surface lightness (Sharan, Li, Motoyoshi, Nishida, & Adelson, 2008).

Recent work indicates that the visual system combines noisy cues to form estimates that have lower variance than any of the individual cues (see Chapter 1, also Landy et al., 1995). The reduction in variance sometimes approaches the theoretical minimum for uncorrelated cues (Ernst & Banks, 2002; Landy & Kojima, 2001).

The reliability of cues changes from scene to scene (due to changes in scene content, viewing geometry, pattern of illumination, etc.) and across the life span (as the sensory apparatus matures). As a consequence, an ideal sensory system must react to changes in cue reliability by adjusting its rule of combination, giving greater or less weight to sensory cues as they become more or less reliable.

In this chapter we are concerned with how cue weights are determined. But, we are also interested in how new cues are acquired and possible errors made by the visual system while acquiring new cues. In the experiments reviewed here, it appears that the sensory systems give weight to pseudocues that result in failures of object constancy. Thus, when a pseudocue was artificially correlated with a trusted cue (haptic input), a subset of the participants apparently increased the weight of that pseudocue, resulting in more pronounced failures of depth constancy. The weighting of pseudocues need not be fixed. Rather, evidence suggests that it is dynamic and shifts with recent experiences in a similar manner to the weighting of standard depth cues when they are artificially covaried with other cues (Atkins, Jacobs, & Knill, 2003; Ernst et al., 2000; Jacobs & Fine, 1999).

Although we have only noted a handful of possible pseudocues here, there are likely many more such pseudocues that are used by the visual system to make judgments of surface and object properties. An understanding of how and when pseudocues are used could provide useful insight to how a constant representation of the visual world is—or is not—acquired.

ACKNOWLEDGMENTS

Thanks to Sabrina Schmidt, Tim Schönwetter, and Natalie Wahl for help with data collection. This research was supported in part by National Institutes of Health grants EY16165 and

EY08266 and by the Deutsche Forschungsgemeinschaft (DFG, Emmy-Noether-Programm, grant TR 528/1-2; 1-3).

REFERENCES

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research*, *41*, 449–461.

Atkins, J. E., Jacobs, R. A., & Knill, D. C. (2003). Experience-dependent visual cue recalibration based on discrepancies between visual and haptic percepts. *Vision Research*, *43*, 2603–2613.

Backus, B. T., & Haijiang, Q. (2007). Competition between newly recruited and pre-existing visual cues during the construction of visual appearance. *Vision Research*, *47*, 919–924.

Berkeley, G. (1709). An essay towards a new theory of vision. In C. M. Turbayne (Ed.), *Works on vision*. Indianapolis, IN: Bobbs-Merrill.

Burge, J., Peterson, M. A., & Palmer, S. E. (2005). Ordinal configural cues combine with metric disparity in depth perception. *Journal of Vision*, *5*, 534–542.

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

Ernst, M. O., Banks, M. S., & Bülthoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, *3*, 69–73.

Haijiang, Q., Saunders, J. A., Stone, R. W., & Backus, B. T. (2006). Demonstration of cue recruitment: Change in visual appearance by means of Pavlovian conditioning. *Proceedings of the National Academy of Sciences USA*, *103*, 483–488.

Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, *298*, 1627–1630.

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

Ho, Y. X., Landy, M. S., & Maloney, L. T. (2006). How direction of illumination affects visually perceived surface roughness. *Journal of Vision*, *6*, 634–648.

Ho, Y. X., Landy, M. S., & Maloney, L. T. (2008). Conjoint measurement of gloss and surface texture. *Psychological Science*, *19*, 196–204.

Ho, Y. X., Maloney, L. T., & Landy, M. S. (2007). The effect of viewpoint on perceived visual roughness. *Journal of Vision*, *7*(1):1, 1–16.

Ho, Y. X., Serwe, S., Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2009). The role of visuohaptic experience in visually perceived depth. *Journal of Neurophysiology*, *101*, 2789–2801.

Jacobs, R. A., & Fine, I. (1999). Experience-dependent integration of texture and motion cues to depth. *Vision Research*, *39*, 4062–4075.

Kaufman, L. (1974). *Sight and mind*. New York, NY: Oxford University Press.

Kellman, P. J., & Arterberry, M. E. (1998). *The cradle of knowledge: Development of perception in infancy*. Cambridge, MA: MIT Press.

Kellman, P. J., & Arterberry, M. E. (2006). Infant visual perception. In D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology, Volume 2. Cognition, perception, and language*. Hoboken, NJ: Wiley.

Klatzky, R. L., Lederman, S. J., & Matula, D. E. (1991). Imagined haptic exploration in judgments of object properties. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 314–322.

Klatzky, R. L., Lederman, S. J., & Matula, D. E. (1993). Haptic exploration in the presence of vision. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 726–743.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*, 2539–2558.

Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, *18*, 2307–2320.

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

Larson, G. W., & Shakespeare, R. (1996). *Rendering with radiance: The art and science of lighting and visualization*. San Francisco, CA: Morgan Kaufmann Publishers, Inc.

Lederman, S. J., & Abbott, S. G. (1981). Texture perception: Studies of intersensory organization using a discrepancy paradigm, and visual versus tactual psychophysics. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 902–915.

Lederman, S. J., & Klatzky, R. L. (1997). Relative availability of surface and object properties during early haptic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1680–1707.

MacLachlan, C., & Howland, H. C. (2002). Normal values and standard deviations for pupil diameter and interpupillary distance in subjects aged 1 month to 19 years. *Ophthalmic and Physiological Optics*, *22*, 175–182.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature*, *447*, 206–209.

Necker, L. A. (1832). Observations on some remarkable optical phaenomena seen in Switzerland; and on an optical phaenomenon which occurs on viewing a figure of a crystal or geometrical solid. *The London and Edinburgh Philosophical Magazine and Journal of Science*, *1*(5), 329–337.

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

Pont, S. C., & Koenderink, J. J. (2005). Bidirectional texture contrast function. *International Journal of Computer Vision*, *62*, 17–34.

Sharan, L., Li, Y. Z., Motoyoshi, I., Nishida, S., & Adelson, E. H. (2008). Image statistics for surface reflectance perception. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, *25*, 846–865.

Ward, G. J. (1994). The RADIANCE lighting simulation and rendering system. *Computer Graphics*, *28*, 459–472.