

Focus cues affect perceived depth

Simon J. Watt

School of Psychology, University of Wales,
Bangor, United Kingdom



Kurt Akeley

Microsoft Research Asia, Beijing, China

Marc O. Ernst

Max-Planck Institute for Biological Cybernetics,
Tübingen, Germany



Martin S. Banks

Vision Science Program, Department of Psychology, and
Wills Neuroscience Institute, University of California,
Berkeley, CA, USA



Depth information from focus cues—accommodation and the gradient of retinal blur—is typically incorrect in three-dimensional (3-D) displays because the light comes from a planar display surface. If the visual system incorporates information from focus cues into its calculation of 3-D scene parameters, this could cause distortions in perceived depth even when the 2-D retinal images are geometrically correct. In [Experiment 1](#) we measured the direct contribution of focus cues to perceived slant by varying independently the physical slant of the display surface and the slant of a simulated surface specified by binocular disparity (binocular viewing) or perspective/texture (monocular viewing). In the binocular condition, slant estimates were unaffected by display slant. In the monocular condition, display slant had a systematic effect on slant estimates. Estimates were consistent with a weighted average of slant from focus cues and slant from disparity/texture, where the cue weights are determined by the reliability of each cue. In [Experiment 2](#), we examined whether focus cues also have an indirect effect on perceived slant via the distance estimate used in disparity scaling. We varied independently the simulated distance and the focal distance to a disparity-defined 3-D stimulus. Perceived slant was systematically affected by changes in focal distance. Accordingly, depth constancy (with respect to simulated distance) was significantly reduced when focal distance was held constant compared to when it varied appropriately with the simulated distance to the stimulus. The results of both experiments show that focus cues can contribute to estimates of 3-D scene parameters. Inappropriate focus cues in typical 3-D displays may therefore contribute to distortions in perceived space.

Keywords: accommodation, blur, cue combination, depth perception, stereoscopic displays, virtual reality

Introduction

Overview

Consider two viewing conditions: a complex real scene viewed binocularly and a computer display of the same scene. The computer display is carefully constructed so all the traditional depth cues—binocular disparity, texture gradients, occlusion, shading, etc.—are geometrically correct. Thus, the geometric patterns of stimulation striking the two eyes are the same in the two cases. Despite the fact that the stimulation patterns are the same, psychophysical research (e.g., Buckley & Frisby, [1993](#); Ellis, Smith, Grunwald, & McGreevy, [1991](#); Frisby, Buckley, & Duke, [1996](#); Frisby, Buckley, & Horsman, [1995](#); van Ee, Banks, & Backus, [1999](#)) and experience with virtual reality displays (Thompson et al., [2004](#)) leads one to expect that the perceived 3-D structure will differ in the two cases: the depth in the computer display will appear flattened relative to the real scene from which it is derived.

A plausible cause for depth flattening is the fact that computer displays present images on one surface: the phosphor grid for cathode-ray displays (CRTs), the pixel grid for liquid crystal displays (LCDs), and the projection screen for projectors. This means that depth information from *focus cues*—accommodation and the retinal blur gradient—is inconsistent with the depicted scene. Instead the information specifies the depth of the display surface. We examined whether such inappropriate focus cues contribute to distortions in perceived depth when viewing 3-D computer displays.

Combining information from multiple depth cues

The 3-D structure of a visual scene is inferred from the 2-D retinal images. The visual system does not rely arbitrarily on one depth cue or another but combines information from multiple available cues to estimate the 3-D parameters of the scene. Consider the case of recovering

the slant of a plane. The visual system's estimate of slant from a given cue can be represented by

$$\hat{S}_i = f_i(S),$$

where S is the slant being estimated and f is the operation by which the visual system does the estimation; the cue is represented by the subscript. Estimates of slant from each cue (\hat{S}_i) are subject to error. When multiple cues are available, the most likely slant can be calculated from a weighted linear combination of the slant indicated by each cue (provided that the noises associated with cue measurement are independent and Gaussian distributed, and that all slants are equally likely)

$$\hat{S} = \sum w_i \hat{S}_i, \quad (1)$$

where

$$w_i = \frac{1/\sigma_i^2}{\sum_i 1/\sigma_i^2}. \quad (2)$$

The weights (w_i) are proportional to the normalized inverse variances (σ_i^2) of the cue distributions (\hat{S}_i), so greater weight is assigned to less variable (i.e., more reliable) cues (Backus & Banks, 1999; Ernst & Banks, 2002; Ghahramani, Wolpert, & Jordan, 1997; Jacobs, 1999; Oruç, Maloney, & Landy, 2003). The variance of the combined estimate is lower than the variance of any single-cue estimate, so by combining information from several depth cues, the visual system can in principle estimate slant (or any other 3-D property) with greater precision than it can by relying on one cue alone. There are now many empirical studies showing that cue reliability is taken into account when combining sensory signals (e.g., Backus & Banks, 1999; Buckley & Frisby, 1993; Jacobs, 1999; Kording & Wolpert, 2004; van Beers, Sittig, & Denier van der Gon, 1998; van Beers, Wolpert, & Haggard, 2002). Furthermore, several studies have tested the quantitative predictions of this model by measuring the reliability of the underlying estimators when only one cue is informative and using these to predict performance when multiple cues are available (Alais & Burr, 2004; Ernst & Banks, 2002; Gepshtein & Banks, 2003; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003; Landy & Kojima, 2001). These studies show that performance is often close to that predicted by the statistically optimal model (in the sense of being the minimum variance unbiased estimate; Ghahramani et al., 1997).

Inappropriate focus cues in 3-D displays

The abovementioned research suggests that the visual system uses all available sources of information to compute 3-D scene parameters. This has important implications for 3-D computer displays because unmodeled depth cues

could affect the percept, causing it to differ from the depicted scene. In almost all computer displays, the focal distance of the light from the display is fixed because the images are presented on one surface (for counter-examples, see Akeley, Watt, Girshick, & Banks, 2004; McQuaide, Seibel, Burstein, & Furness, 2002). This provides inappropriate depth information in two ways.

First, the variation in blur in the retinal image is consistent with the fixed distance of the display surface and not with the distances in the simulated scene. With real scenes, the amount of retinal blur varies because the distance of points in the scene varies with respect to the eye's focal distance: the retinal image is sharpest for objects at the focal distance and blurred for points nearer and farther away. In computer displays, the variation in blur specifies the constant distance of the display surface and is thus a cue to flatness.

Second, accommodation provides an extra-retinal cue signaling the constant distance of the display surface. As the eye looks around a real scene, commands are sent to the ciliary muscles to change the refractive power of the crystalline lens and thereby minimize blur for the fixated part of the scene. As the eye looks around the simulated scene in a computer display, the focal distance of the light does not vary appropriately, so this again signals flatness rather than the simulated depth variation.

If blur and accommodation provide inputs to the calculation of depth, their erroneous values can in principle adversely affect percepts of 3-D scene structure.

Inappropriate motion parallax in 3-D displays

In many settings (including psychophysical experiments), the observer's head position is not strictly constrained. For a viewing distance of 28.5 cm (used in our first experiment), head movements of a few millimeters could result in a detectable signal to depth from motion parallax (Rogers & Graham, 1982). As with focus cues, residual motion parallax specifies the distance to the display rather than distances in the simulated scene. If parallax is figured into the brain's calculation of depth, its erroneous value will adversely affect 3-D percepts.

Unlike the problem of inappropriate focus cues, there are straightforward solutions to this problem: one can track head position and update the image accordingly (Welch et al., 1999) or one can immobilize the head position. Therefore, we did not explicitly examine whether residual motion parallax contributes to distortions in perceived depth when viewing 3-D displays (but see the [Isolating information from accommodation and blur section](#)).

Implications for psychophysics

Powerful 3-D computer graphics has revolutionized research on depth perception. Psychophysicists no longer

have to rely on shadow casters (Gibson, Gibson, Smith, & Flock, 1959), glass plates (Ogle, 1950), or other mechanical means to create stimuli. Using modern computer graphics, they can now create realistic 3-D images and independently manipulate depth cues. As a result, great advances have occurred in the last three decades. However, if focus cues affect perceived depth from conventional computer displays, many observations in the depth perception literature may not be representative of vision in the natural environment. Here we describe two illustrative examples from the literature: (1) the perceived depth of computer-displayed versus real ridges, and (2) the slant-contrast illusion.

Buckley and Frisby (1993) examined the perceived depth of CRT-displayed and real ridges. The stimuli depicted vertical or horizontal parabolic ridges. The authors independently manipulated the disparity- and texture-specified depths of the ridges. With CRT stimuli, they did this in conventional fashion by programming different disparity and texture signals. With the real ridges, they did it by distorting the texture on the card covering wooden forms to create the desired texture gradient viewed from the observer's eye. The data from the CRT stimuli (vertical ridges) revealed clear effects of disparity and texture: Disparity dominated when the texture-specified depth was large and texture dominated when the texture depth was small. In the framework of the cue-weight model (Equations 1 and 2), the disparity and texture weights changed depending on the texture-specified depth. The data from the real-ridge stimuli were quite different: The disparity-specified depth now dominated the percept. The important point for our purposes is that the CRT-based and real-ridge results differed dramatically.

Buckley and Frisby (1993) speculated that focus cues played an important role in the striking difference between the CRT and real results. In Appendix C, we quantify and generalize their argument by translating it into the framework of the weight model. The fact that more depth was perceived in real than in CRT-displayed ridges suggests that focus cues contributed to the depth calculation in their experiments (see also Frisby et al., 1995).

We cannot tell from the Buckley and Frisby (1993) experiments whether depth percepts were veridical once focus cues were consistent with the depth specified by disparity and texture. The reason is that responses were judged depth in cm and we cannot know whether the mapping between perceived depth and depth responses is veridical. For our purposes, the important point is that observers reported and therefore presumably saw more depth when focus cues were consistent with the depth specified by other cues.

Now consider the second example: the slant-contrast illusion (Sato & Howard, 2001; van Ee & Erkelens, 1996; Werner, 1937). In this illusion, a central object is presented that has the disparity and texture gradients of a fronto-parallel plane. It is surrounded by a surface that typically has the texture gradient of a frontoparallel plane but

the disparity gradient of a slanted plane. The presence of the surrounding plane causes the central object to appear slanted in a direction opposite to the disparity-specified slant of the surround. Interesting psychophysical effects draw researchers' attention, so several theories have been developed to explain the illusory slant. Most share the idea that disparity-encoding mechanisms have antagonistic, center-surround receptive fields for disparity (in analogy to the center-surround organization of receptive fields in the luminance domain). Such mechanisms are allegedly less responsive to zero- and first-order disparities (absolute disparity and the relative disparity associated with a slanted plane, respectively) than to second- and higher-order disparities (the disparity associated with curvature or discontinuities in depth) (Anstis, Howard, & Rogers, 1978; Brookes & Stevens, 1989; Gillam, Chambers, & Russo, 1988; Mitchison, 1993; Rogers & Graham, 1983; van Ee & Erkelens, 1996; Westheimer, 1986).

van Ee et al. (1999) measured the magnitude of the slant-contrast illusion when the stimulus was presented as a conventional computer display and as real surfaces. They observed a typically large illusion with the computer display, but no illusion at all with the real surfaces. The computer-displayed and real-surface stimuli had the same dimensions and were viewed from the same distance, so the disparity- and texture-gradient signals created by the two stimuli were identical. The fact that one produced the illusion and the other did not means that the encoding of disparity (and the texture gradient) *per se* cannot be the cause of the illusion. van Ee et al. argued that cue conflicts between geometric cues (disparity and texture) and inappropriate focus cues caused the illusion in the computer-displayed stimuli. The conflicts were eliminated in the real-surface stimulus and so the illusion was eliminated. Sato and Howard (2001) also showed that manipulating the magnitude of cue conflicts has a large effect on the slant-contrast illusion when the disparity signals are held constant. Our point is that cue conflicts between disparity, texture, and the previously unmodeled cues of blur and accommodation affect or may even cause the slant-contrast illusion. Thus, previous theories of the illusion are attempting to explain an illusion that may not occur in the natural environment, when all cues signal the same depth structure.

The potential importance of inappropriate focus cues is not restricted to stereoscopic vision. We argue in the General discussion section that investigations of any aspect of visual space perception should take the potentially confounding effects of those cues into account.

Recovering depth from blur

We define blur in the retinal image as the spread of the optical point-spread function (Westheimer, 1986). For a fixed accommodative state, the amount of blur in the image of an object is roughly proportional to the focus error

in diopters (Green & Campbell, 1965; Mather & Smith, 2000; Smith, Jacobs, & Chan, 1989). Objects at different distances are blurred by different amounts, signaling depth variations in the scene. Interpreting this signal is complicated by two factors. First, the sign of depth variation is undetermined because the retinal images of objects nearer or farther than fixation can be equally blurred. Second, the magnitude of the depth signaled is ambiguous because for a given accommodative state, blur depends not only on the distance of an object from fixation, but also on the visual system's depth of focus, which in turn depends on pupil size and the spatial frequency content of the input, neither of which is known independently (Green, Powers, & Banks, 1980). For these reasons, it seems unlikely that metric depth can be recovered directly from retinal blur. However, the continuous microfluctuations that occur in accommodation (Campbell & Westheimer, 1959) and chromatic aberration could be used to disambiguate the blur signal (Nguyen, Howard, & Allison, 2005; Pentland, 1987). Additionally, eye movements could be used to sample changes in blur dynamically as the observer focuses on different parts of the scene. The sign of depth variations could also be disambiguated by other depth cues including binocular disparity and occlusion.

Some psychophysical studies have reported a modest effect of the blur gradient on judgments of perceived depth (Marshall, Burbeck, Ariely, Rolland, & Martin, 1996; Mather, 1996, 1997; Mather & Smith, 2000, 2002; O'Shea, Govan, & Sekuler, 1997). In these studies, the blur gradient was varied artificially by blurring the displayed object in selected regions to simulate the effects of defocus, and most used brief presentations. This means that the abovementioned strategies for disambiguating the depth signaled by blur could not have been used. It is thus possible that the blur gradient is a more useful depth cue in natural viewing than previously realized.

Recovering depth from accommodation

The efferent signal to the muscles controlling the crystalline lens could be a depth cue because the magnitude of the response required to focus the retinal image depends directly on the distance from the eye to the fixated object. To be a useful depth cue, the accommodative system must respond reliably to changes in focal distance and the visual system must be able to monitor the muscle commands. Accommodation to isolated, high-contrast targets is reliably related to changes in a target's focal distance (Campbell & Westheimer, 1959; Charman & Tucker, 1977; Heath, 1956). Indeed, accommodation can occur to changes in retinal blur that are below perceptual threshold (Kotulak & Schor, 1986).

In contrast to the blur gradient (and most other depth cues), accommodation can in principle be used to recover the absolute distance to fixation. Several studies have examined distance estimates with verbal or pointing

responses based on the accommodative response to single targets and have shown that observers' estimates are correlated with target distance, but that accuracy is poor and variability is high (Baird, 1903; Biersdorf, 1966; Dixon, 1895; Fisher & Ciuffreda, 1988; Foley, 1977; Hillebrand, 1894; Künnapas, 1968; Mon-Williams & Tresilian, 2000; Peter, 1915; Swenson, 1932; Wundt, 1862). In principle, accommodation can also provide information about surface structure if estimates of relative distance are compared over successive fixations. Accommodation, like blur, could therefore be a more useful depth cue in complex scenes than the existing psychophysical data suggest.

Direct versus indirect influence of focus cues on perceived depth

The above discussion examined how blur and accommodation could be used directly in estimating depth. Accommodation could also have an indirect effect on perceived depth by interacting with stereopsis. Binocular disparity is an important and reliable depth cue. But horizontal disparities are inherently ambiguous because 3-D layout cannot be determined from them without scaling by an estimate of viewing distance (Gårding, Porrill, Mayhew, & Frisby, 1995). To perform the scaling, the brain uses the eyes' vergence and the horizontal gradient of vertical disparity (Rogers & Bradshaw, 1995). In principle, accommodation can also provide an estimate of fixation distance, which may in turn influence disparity scaling. In computer displays, the accommodative stimulus is the distance to the display screen and not the simulated distance. This erroneous information may affect depth percepts indirectly via disparity scaling.

There is a small literature on indirect effects of accommodation on perception. Fisher and Ebenholtz (1986), Mon-Williams and Tresilian (2000), and Wallach and Norris (1963) observed an influence of accommodation on depth interpretation (for a negative result, see Ritter, 1977). Heinemann, Tulving, and Nachmias (1959) and von Holst (1973) observed an influence of accommodation on perceived size.

Direct and indirect effects of focus cues were examined in Experiments 1 and 2, respectively.

Experiment 1

In the first experiment, slant specified by geometric cues (texture and binocular disparity) was varied independently from slant specified by focus cues.

Because so many reliable depth cues are available in natural viewing, focus cues should have only a small influence on the recovery of 3-D scene properties in natural conditions. The simulated scenes used in psychophysical

experiments are often impoverished in order to study individual cues and their interactions. An example is the sparse random-dot stereogram, which allows researchers to isolate binocular disparity while making all other cues uninformative or unreliable. Focus cues may have more influence under these circumstances. To examine this possibility, we measured the effect of varying focus cues on slant estimates when the stimulus was defined by only binocular disparity or by only the texture gradient.

Methods

Observers

Three observers participated, aged 24–29 years. All had normal vision and stereoacuity. All were experienced psychophysical observers. One (AJW) was naive to the experimental purpose. The other two knew the general purpose but not the specifics.

Apparatus

The layout of the apparatus is schematized in Figure 1. The stimuli were displayed on a conventional 21-in. CRT (KDS VS21e) with 1600×1024 resolution. Each pixel

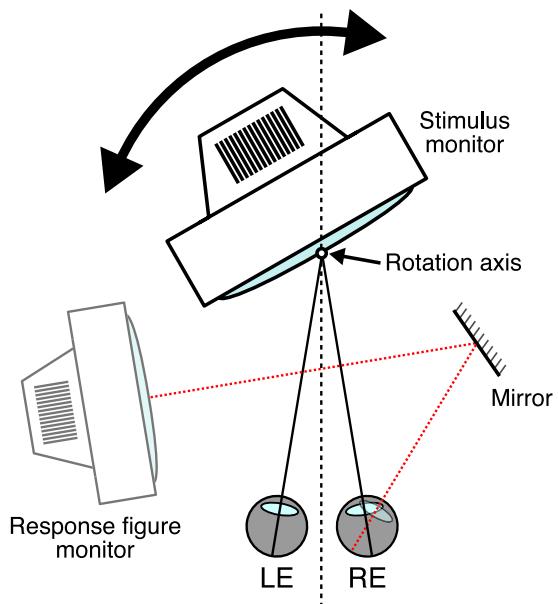


Figure 1. Layout of the apparatus for [Experiment 1](#). The stimulus monitor was straight ahead of the observer at different distances. It could be rotated about a vertical axis passing through the center of its front surface. The response monitor was to the left; observers made an eye movement to view the response figure on this monitor, which was visible only to the right eye. The response figure consisted of two line segments, one horizontal and the other variable in orientation. Observers adjusted the variable segment until the angle between it and the horizontal segment was the same as the perceived slant of the stimulus.

subtended 2.9×2.9 arcmin. To manipulate the information from focus cues, the monitor was rotated about the vertical axis passing through the center of its front surface.

Focus cues issuing from the phosphor grid specified a surface that was not exactly a plane for two reasons. (1) The surface containing the phosphor grid was slightly curved, and (2) the grid's virtual distance was affected by refraction due to the front glass plate. (We could not use a flat-panel LCD because the luminance of such displays depends strongly on viewing angle.)

Dichoptic presentation of the left- and right-eye images was achieved using CrystalEyes™ liquid crystal shutter glasses. The monitor refresh rate was 100 Hz, so each eye's image was redrawn at 50 Hz. It was crucial to have no artifactual cues to the monitor's slant, so we were careful to eliminate cross-talk through the glasses (aided by drawing the images with the red phosphor only) and to eliminate the observer's ability to see the monitor casing (accomplished by masking the casing and by periodically light-adapting the observer). We checked that observers could not determine the monitor's slant in a pilot experiment. In the monocular-viewing conditions of the main experiment, observers wore a patch over their left eye.

We used anti-aliasing to specify the position of stimulus elements to subpixel accuracy. Stimuli were rendered using OpenGL (Segal & Akeley, [2002](#)) and the associated utility library, GLUT (Kilgard, [1996](#)). Precise reproduction of visual directions was achieved using a spatial calibration technique similar to the one described by Backus, Banks, van Ee, and Crowell ([1999](#)). A wire-filament loom was placed in a known position in front of the monitor and the experimenter aligned individual dots with the loom intersections. During calibration, the experimenter's head was carefully positioned using a bite bar, which was adjusted so as to position the eyes' centers of rotation in known positions relative to the display. Two-dimensional polynomial functions were used to fit the x and y values from the loom calibration to pixel space in which the stimuli were rendered. These equations provided a continuous look-up table relating pixel space and physical screen space. When the stimulus was drawn, each stimulus element (squares or lines) was subdivided into a series of smaller polygons and the position of each vertex of these was corrected using the look-up table. This procedure corrected overall dot positions and line endpoints, and it also closely approximated the correct calibration for the outlines of the stimulus elements. Because of the calibration procedure, the geometric properties of the stimulus were matched for all monitor slants. The spatial calibration procedure was carried out separately for the left and right eyes at each monitor slant used in the experiment.

During the main experiment, the observer's head position was stabilized using a conventional chin rest. A sighting technique (Hillis & Banks, [2001](#)) was used to position the chin rest precisely. We chose this method for

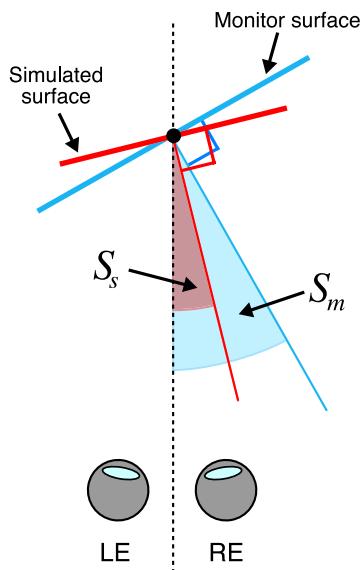
head constraint to mimic the most common practice in the psychophysical literature. As discussed previously, it is possible that motion parallax resulting from small head movements may have provided an additional cue to the physical slant of the monitor. Possible implications of this, and additional control conditions in which the head was immobilized with a bite bar, are described in the [Isolating information from accommodation and blur](#) section. A response figure was presented on a second CRT. It was viewed via a mirror so that observers could respond without making head movements ([Figure 1](#)).

Stimuli

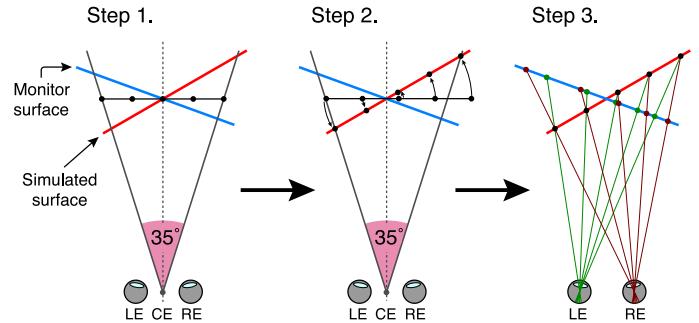
The stimuli were planes rotated about the vertical axis (tilt = 0°). We independently manipulated two cues to slant: (1) focus cues, which were manipulated by varying monitor slant, and (2) the simulated slant of the surface, which was specified by geometric information from disparity and texture cues. We refer to monitor slant as S_m and simulated slant as S_s , respectively ([Figure 2](#)).

S_s was specified either by binocular disparity (disparity condition) or by the perspective projection of a textured pattern (texture condition). For all viewing conditions and values of S_s and S_m , the stimulus width was 35° with respect to the cyclopean eye.

[Figure 3](#) shows how the stimuli were created. The stimulus generation method was used in both the disparity and texture conditions; only the right-eye's image was displayed in the latter case. The stimulus width was



[Figure 2](#). Plan view of the stimulus configuration for [Experiment 1](#). The slants S_m and S_s were defined relative to the cyclopean line of sight. Slant in both cases is the angle between the line of sight to the middle of the display monitor (dotted line) and the surface normal for each cue (red and blue lines). Positive slant (shown here) is "right side back".



[Figure 3](#). The method of stimulus generation for [Experiment 1](#). Step 1: Coordinates were defined for a homogeneous, frontoparallel pattern (randomly positioned squares or a Voronoi texture) 35° wide, measured at the cyclopean eye (CE, midway between the two eyes). Step 2: This pattern was scaled and translated in x such that after rotation by the angle S_s , it remained 35° wide, measured at the cyclopean eye. Step 3: The left- and right-eye's images were determined by projecting the pattern onto the monitor plane using each eye's position as the center of projection. The screen space was spatially calibrated (see text) so that the visual direction of each point on the stimulus was appropriate, and the retinal images at each value of S_s were geometrically equivalent at each monitor slant, S_m .

matched with respect to the cyclopean eye (midway between the two eyes). Therefore, its angular extent in the right eye (its width in the texture condition) varied slightly as a function of S_s . The angular extent of the stimulus in either eye (and all other geometric properties) was unaffected by variations in S_m . Stimulus height at the axis of rotation on average was 28°. Due to random aspects of the stimulus generation method, there were small variations in stimulus height and width from trial to trial. The distance from the cyclopean eye to the rotation axis of the stimulus was always 28.5 cm. We chose this distance because it was short enough to create discriminable changes in focal distance while being long enough to allow accurate accommodation.

In the disparity condition, S_s was specified by the difference between left- and right-eye projections (calculated for each observer's inter-ocular distance) of a pattern of randomly positioned square elements. We used squares instead of the more typical Gaussian blobs to provide a better stimulus to accommodation. The initial 2-D pattern ([Figure 3](#), Step 1) was generated by drawing x and y square positions from a uniform random distribution. The average size of each square was 1.7 × 1.7 mm (1.7 mm ≈ 0.34° at the center of the stimulus). We minimized the informativeness of the texture cue by presenting few squares—roughly 0.2 square/ deg^2 —in random positions. We also clipped the stimulus with an elliptical window (whose size and orientation varied randomly within a small range) so that the outline of the stimulus pattern did not provide a cue to S_s . The scaling process ([Figure 3](#), Step 2) stretched the entire pattern, including

the squares, so that when the stimulus was rotated, the angular width of the squares was on average constant across values of S_s . Each eye's view was calculated by finding the intersection with the monitor plane of rays through the stimulus pattern and each eye's center of rotation (Figure 3, Step 3). Using this procedure, the outline of each square was correctly projected in each eye's view. This meant that the simulated slant of each square was consistent with S_s , and the monocular texture cue in each eye (including square density) was consistent with the disparity-specified slant. We could have used the conventional method, in which stimulus elements are shifted by equal and opposite amounts in the two eyes, thereby creating a texture gradient that is consistent with a frontoparallel plane. It was preferable, however, to use correct perspective projection because the conventional method yields a texture-specified slant of zero, which would have complicated the data interpretation.

In the texture condition, S_s was defined by the perspective projection of a Voronoi pattern (de Berg, van Kreveld, Overmars, & Schwarzkopf, 2000; see also Knill, 1998) viewed with the right eye. The stimuli consisted of 320 Voronoi cells on average. To create the Voronoi patterns, the initial pattern consisted of a grid of 20×16 regularly spaced points. The x and y coordinates of each point were then perturbed by a random amount in the range ± 0.2 times the inter-point spacing (equivalent to 0.36° in the center), and the Voronoi pattern defined by these points was calculated. The resultant had ~ 0.33 Voronoi cells/ deg^2 . The stimulus was then scaled, rotated, and perspective projected into the monitor plane following the procedure in Figure 3. As with the random-dot stimulus, each line segment was correctly projected for the slant angle, S_s .

The average luminous intensity of a square or line seen through the shutter glasses was 0.9 cd/m^2 , and the background luminance was 0.01 cd/m^2 .

A new stimulus was drawn on each trial in both the disparity and texture conditions. In our experimental design, it was critical that the geometric information at a given value of S_s was equivalent for all values of S_m . We checked this empirically by viewing a simulated frontoparallel plane ($S_s = 0^\circ$) through the calibration loom. The stimulus was identical at a range of values of S_m .

Our stimuli should have been good stimuli to accommodation because they were spatially complex and therefore contained a wide range of spatial frequencies (Charman & Tucker, 1977).

Procedure

Observers reported the amount of perceived slant for each combination of monitor slant (S_m) and simulated slant (S_s): 0° , $\pm 10^\circ$, $\pm 20^\circ$, and $\pm 30^\circ$. They did so by setting the angle between two line segments to be equal to the perceived slant of the stimulus. The response figure con-

sisted of a fixed horizontal line and a rotatable oblique line, the former representing the frontoparallel plane and the latter the perceived slant of the experimental stimulus. The oblique line started at a random orientation on each trial and could be adjusted by key presses in either direction in increments as small as 0.5° . This figure was viewed by the right eye in a second monitor via a mirror by making a small eye movement (Figure 1).

Before each trial, a small fixation square ($0.35^\circ \times 0.35^\circ$) was presented in the center of the screen. The square was constructed and calibrated using the same methods as the stimulus. Its simulated slant was always frontoparallel, so it did not provide a cue to monitor slant. Each trial followed the same sequence. The fixation square first appeared for 1 s, then the stimulus for 2 s. Following stimulus offset, the response figure appeared on the second monitor and observers indicated the amount of slant they had seen. The response figure then disappeared followed by a 1-s blank display before the fixation square appeared on the main monitor for the next trial. The fixation square was not present during the stimulus presentation and observers were given no specific instructions about where to look. The observers completed six trials for each $S_m \times S_s$ combination in both the disparity and texture conditions: a total of 588 trials. Trials were blocked by monitor slant and viewing condition, both randomly ordered.

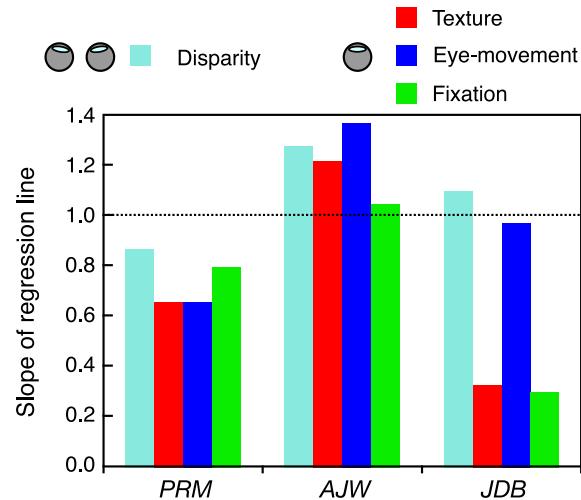


Figure 4. Effect of slant on slant settings for the cues-consistent stimuli for each observer and condition in Experiment 1. The abscissa values are different observers. Blue-gray and red bars represent the disparity and texture conditions, respectively. The dark-blue and green bars represent two additional monocular conditions, described in isolating information from accommodation and blur. The ordinate values are the slopes of the best-fitting lines relating slant to observer responses in the cues-consistent ($S_m = S_s$) subset of the data in each viewing condition. These values were used to normalize the raw responses for each observer (see the Normalization of slant estimates section).

The apparatus was concealed behind a curtain when observers entered the room, and the experiment was conducted in complete darkness. Observers were always unaware of the monitor's slant (the naïve observer was not aware that the monitor ever rotated). Between experimental blocks, observers were exposed to normal light levels to prevent dark adaptation. Before the main experiment, the observers completed two blocks of practice trials. All three observers reported a clear percept of depth in binocular and monocular conditions, and they were all readily able to do the task.

Results

Normalization of slant estimates

We cannot know the mapping between perceived slant and response setting, so we used the settings with the cues-consistent stimuli ($S_m = S_s$) to normalize the other

data. We did this by transforming the raw data as follows. For each observer and condition, a response-mapping function was derived by least-squares fitting of a line ($y = mx + c$) to the mean slant estimates from the subset of the data for which $S_m = S_s$. If it is assumed that perceived slant was veridical for these cues-consistent stimuli, the settings can then be used as a yardstick to transform the data in the other conditions. The fitted function was used to scale each response into a normalized slant estimate. These values were then used to calculate the points in the data figures. Because the data were merely scaled to make effect sizes equivalent across conditions and observers, the relative effects within each condition were unaffected. The data were in every case well fitted by a line.

The slopes of the normalization functions for the disparity (blue-gray bars) and texture conditions (red bars) are shown in Figure 4. The observers' settings in the cues-consistent conditions were reasonably consistent

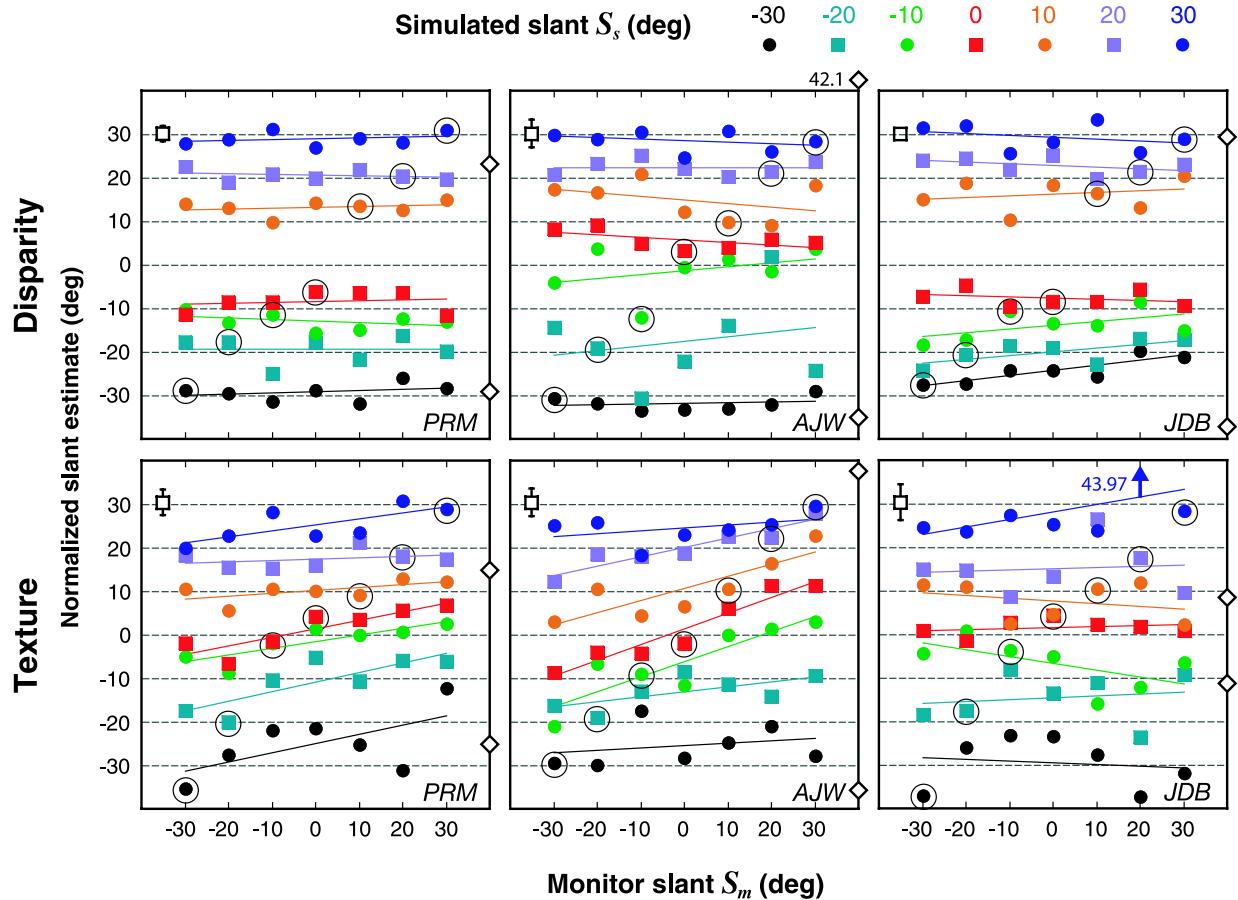


Figure 5. Average normalized slant estimates for each value of S_s as a function of S_m in Experiment 1. The upper row shows the data for the disparity condition and the lower row the data for the texture condition. The columns show the data for different observers. The horizontal dashed lines represent veridical estimates for each S_s . The colored symbols represent the data, different colors denoting different values of S_s . The circled points are the data for the cues-consistent ($S_m = S_s$) conditions. The colored lines are the best fits to the data for each S_s . The error bars in the upper left corner of each panel are \pm the average SEM. The diamonds on the right side of each panel indicate the actual response settings for the cues-consistent stimuli at $S_m = S_s = \pm 30^\circ$. The data were normalized such that the fitted settings at those points plotted at ordinate values of $\pm 30^\circ$.

across the disparity and texture conditions with the exception of observer JDB, whose settings in the texture condition were considerably smaller than in the corresponding disparity condition. Despite possible differences in the use of the response measure, it seems likely that the texture-defined planes looked less slanted than the disparity-defined planes to this observer.

Effects of monitor orientation on perceived slant

Figure 5 plots each observer's average normalized slant estimates as a function of S_m in the disparity and texture conditions. Different colors represent different values of S_s . The solid lines are the best-fitting lines for each value of S_s . The data are plotted as a function of S_m , so effects of this variable are indicated by deviations from a horizontal line. (The data were fitted with lines for simplicity, but one would expect departures from linearity because the effect of focus cues is likely to vary with S_m ; see the Evidence for reliability-based cue weighting section.) The normalization of the data is indicated here by the diamonds on the right side of each panel (see caption for explanation).

Consider the data for observer PRM. In the disparity condition, the data are clearly separated according to S_s ,

indicating that disparity was an effective slant cue. Monitor slant did not affect his judgments in this condition. For example, his slant estimates in the cues-inconsistent conditions ($S_m \neq S_s$) did not differ noticeably from estimates in cues-consistent conditions ($S_m = S_s$; circled data points). In contrast, PRM's slant estimates in the texture condition reveal a clear effect of monitor slant. Again, the data are separated according to S_s , indicating that the texture cue was effective. However, for most values of S_s , increasing or decreasing monitor slant had a systematic effect on his estimates, suggesting that focus cues affected perceived slant.

The results for observer AJW are similar. In the disparity condition, her slant estimates were less consistent than those of PRM, but there was no systematic effect of monitor slant. Her slant estimates in the texture condition varied systematically with S_m .

The results for observer JDB are more variable, but reasonably consistent with those of the other two observers. He showed no effect of monitor slant in the disparity condition and a somewhat inconsistent effect in the texture condition; in his data, the effect of monitor slant in the texture condition is most evident when one compares perceived slant when $S_m = S_s$ to perceived slant when $S_m = 0$ (see Figure 6).

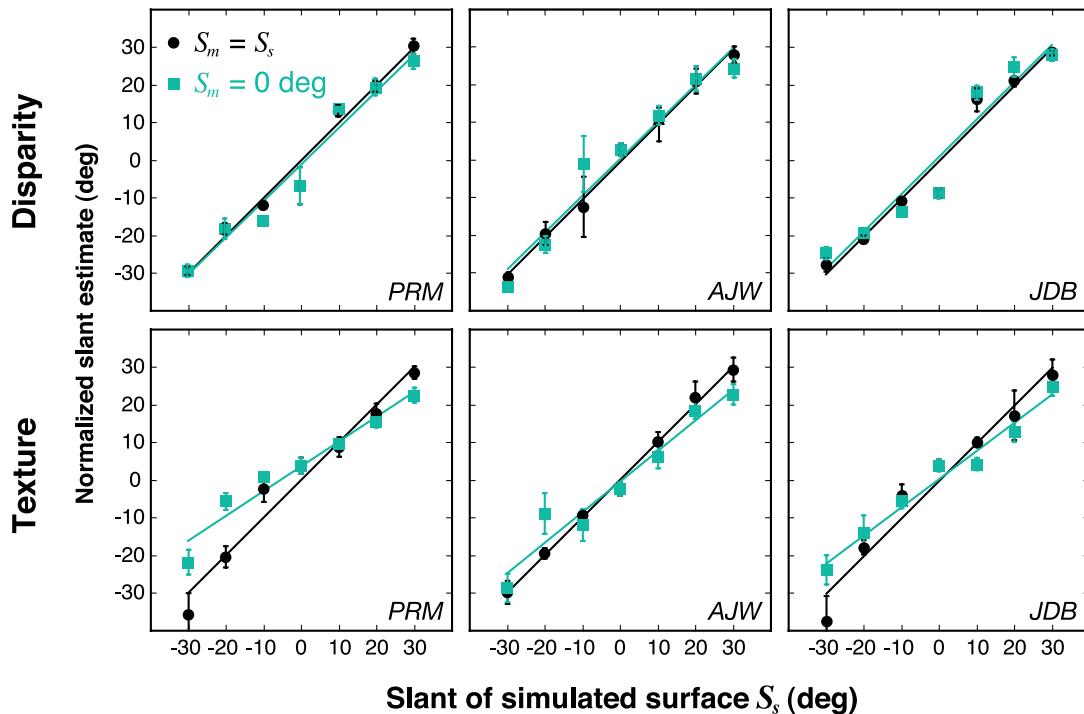


Figure 6. Average normalized slant estimates for $S_m = S_s$ and $S_m = 0$ in Experiment 1. Each panel plots the normalized estimates as a function of S_s . Each column shows the data for a different observer. The upper and lower rows show the data from the disparity and texture conditions, respectively. The black circles are the data for $S_m = S_s$ and the blue squares are the data for $S_m = 0$. The lines are best fits to the data. The slopes of the fitted lines to the cues-consistent data are constrained to be 1 as a result of the normalization process. Error bars are ± 1 SEM.

Implications of the direct effect of focus cues for 3-D displays

Figure 6 illustrates implications for viewing simulated scenes as opposed to real scenes. The figure re-plots two subsets of the data: (i) the cues-consistent conditions ($S_m = S_s$), and (ii) the $S_s = 0^\circ$ condition. Normalized slant estimates are now plotted as a function of S_s instead of S_m . The cues-consistent condition is essentially equivalent to real-world viewing in that all cues specify the same depth structure. The $S_m = 0$ condition is the typical viewing situation in psychophysics in which the display surface is frontoparallel. The lines in **Figure 6** are the best fits for each data subset. The slopes of the fitted lines to the cues-consistent data (black lines) are constrained to be 1 as a result of the normalization process. There was no systematic difference in the disparity condition between the cues-consistent and cues-inconsistent conditions. For all three observers, we calculated the difference between slant estimates in the cues-consistent and cues-inconsistent conditions

at each value of S_s (except for $S_s = 0$, where the data in the two conditions are the same). The signs of the differences were adjusted so that a negative difference always indicated less estimated slant (stimulus appeared closer to frontoparallel) in the $S_m = 0$ condition irrespective of the sign of S_s . A one-sample *t* test showed that these difference scores were not significantly different from zero, indicating that slant estimates in the disparity condition were not reliably different in the cues-consistent and cues-inconsistent conditions, $t(17) = 0.19$, $p = 0.85$. This shows again that focus cues had no direct influence on slant percepts under binocular viewing. In the texture condition, all three observers reported seeing less slant when the monitor was frontoparallel ($S_m = 0$) compared to when all cues were consistent ($S_m = S_s$). The difference score analysis described above showed that this effect was statistically significant, $t(17) = 4.18$, $p < 0.001$. This suggests again that focus cues affected slant percepts directly under monocular viewing.

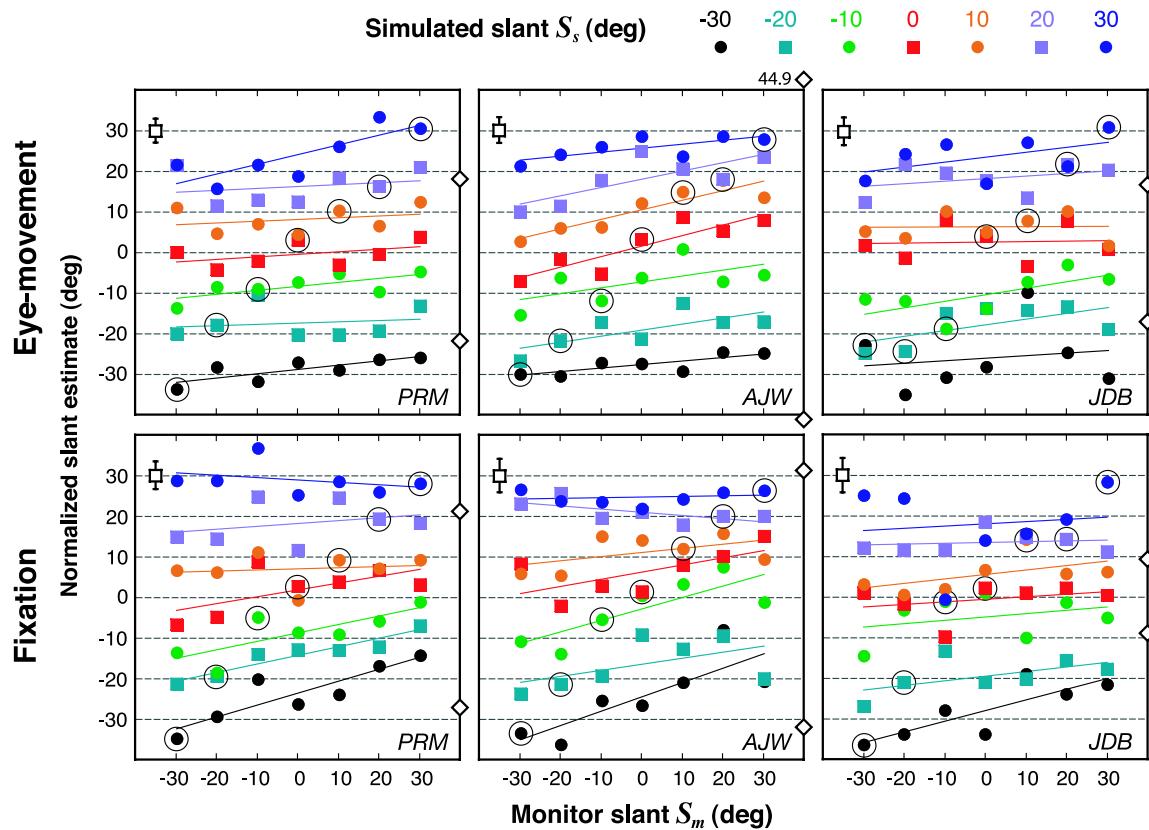


Figure 7. Average normalized slant estimates for S_s as a function of S_m in the eye movement and fixation conditions in **Experiment 1**. The upper and lower rows show the data from the eye movement and fixation conditions, respectively. Each column shows data from a different observer. The horizontal dashed lines represent veridical estimates for each S_s . The colored symbols represent the data, different colors denoting different values of S_s . The circled points are the data for the cues-consistent ($S_m = S_s$) conditions. The colored lines are the best fits to the data for each S_s . The error bar in the upper left corner of each panel represents \pm the average SEM. The diamonds on the right side of each panel indicate the actual response settings for the cues-consistent stimuli at $S_m = S_s = \pm 30^\circ$. The data were normalized such that the fitted settings at those points plotted at ordinate values of $\pm 30^\circ$.

Isolating information from accommodation and blur

To determine if residual motion parallax contributed to the monitor-slant effect, we re-ran the monocular condition with the observers' heads completely stabilized with a bite bar. To determine whether the blur gradient or accommodation made a greater contribution to the monitor-slant effect, we compared performance in two conditions: (1) the *eye movement* condition, in which observers made two horizontal eye movements from one edge of the stimulus to the other and back during the 2-s presentation, and (2) the *fixation* condition, in which observers maintained fixation on a small cross ($0.75^\circ \times 0.75^\circ$) in the center of the screen before and during stimulus presentation. Accommodation should have varied much less in the fixation than in the eye movement condition, so by comparing the data in the eye movement and fixation conditions, we could assess the contribution of accommodation. By comparing the data in these two conditions to the original data in Figures 5 and 6, we could determine the contribution of residual motion parallax.

The data were normalized using the abovementioned procedure. Figure 4 shows the slopes of the normalization functions for the eye movement (dark-blue bars) and fixation conditions (green bars). Once again, the observers' settings were consistent across conditions with the exception of observer JDB, who made very small settings in the

fixation condition (similar to those he made in the original texture condition). Again, this may be because for this observer the surfaces looked less slanted in this condition, although it is unclear why this should have been the case.

Figure 7 shows the results of the eye movement and fixation conditions in the same format as Figure 5. Results for the eye movement and the fixation conditions were quite similar to the results in the original texture condition (Figure 5, see also Figure 9). The similarity between the results in Figures 5 and 7 implies that the monitor-slant effect in the texture condition was not caused by residual motion parallax or by differential accommodation accompanying eye movements. We conclude that retinal blur was the primary cause of the effect of monitor slant under monocular viewing.

Figure 8 re-plots two subsets of the data: the cues-consistent conditions ($S_m = S_s$), and the $S_m = 0^\circ$ condition, in the same format as Figure 6. The abscissa is S_s . The lines are the best fits for each data subset. The slopes of the fitted lines to the cues-consistent data (black lines) are constrained to be 1 as a result of the normalization process. All three observers in both conditions reported less slant when $S_m = 0$ than when $S_m = S_s$ (cues consistent), with the exception of AJW in the eye movement condition (she, however, showed a consistent effect of monitor slant overall; Figure 7). One-sample *t* tests were carried out on the differences between slant estimates

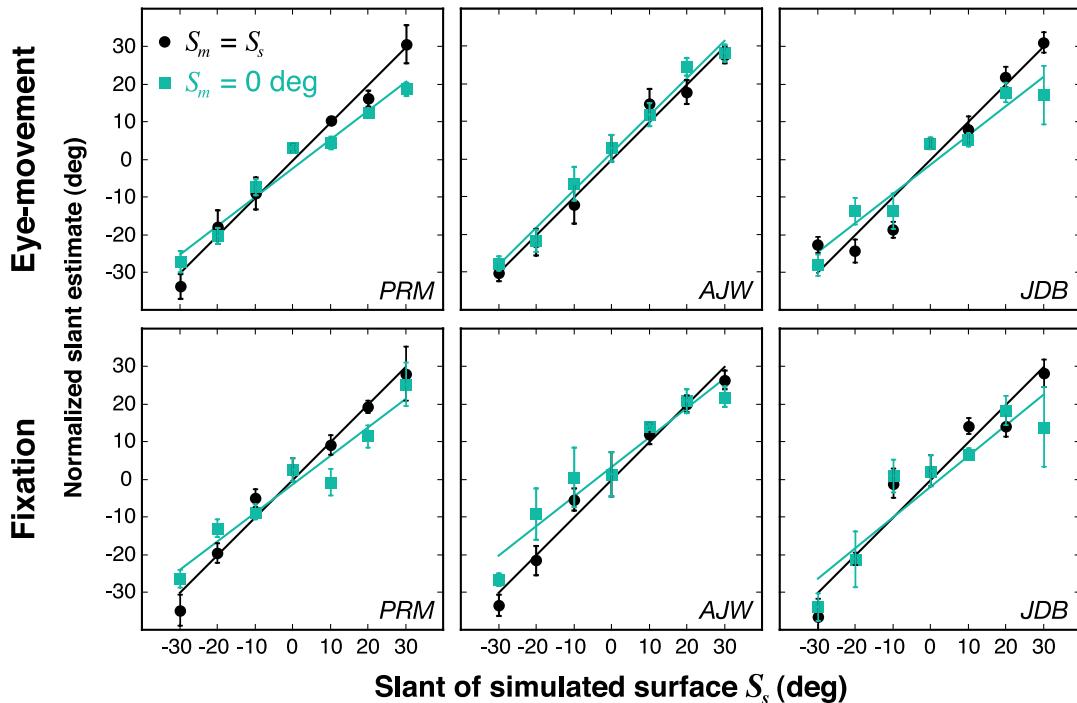


Figure 8. Average normalized slant estimates for $S_m = S_s$ and $S_m = 0$ in the eye movement and fixation conditions. Each panel plots the normalized estimates as a function of S_s . Each column shows the data for a different observer. The upper and lower rows show the data from the eye movement and fixation conditions, respectively. Both of those conditions were conducted with monocular viewing. The black circles are the data for $S_m = S_s$ and the blue squares are the data for $S_m = 0$. The lines are best fits to the data. The slopes of the fitted lines to the cues-consistent data are constrained to be 1 as a result of the normalization process. Error bars are ± 1 SEM.

for $S_m = S_s$ and $S_m = 0^\circ$. The difference in reported slant was statistically significant: observers reported less slant when $S_m = 0$ than when $S_m = S_s$ in the eye movement condition, $t(17) = 2.63$, $p < 0.05$, and fixation condition, $t(17) = 3.46$, $p < 0.01$. This suggests again that focus cues affected slant percepts directly under monocular viewing.

The effects of monitor slant are summarized in Figure 9. The normalized data from EACH observer in each condition were entered into a multiple regression analysis with S_m and S_s as factors. The figure plots the regression weight for S_m separately for each condition and observer, as well as an average weight for each condition. The regression weights are the average weights given to monitor slant across all values of S_m and S_s . Regression weights greater than 0 indicate an effect of monitor slant. No effect was observed in the disparity condition. A consistent effect was observed in the texture condition and it persisted in the eye movement and fixation conditions where head position was fixed. Thus, residual motion parallax with chin rest constraint had no discernible effect, perhaps because the head movements were small. The fact that the effect persisted in the fixation condition, where observers held fixation on one point in the stimulus, suggests also that accommodation accompanying 3-D eye movements had no effect.

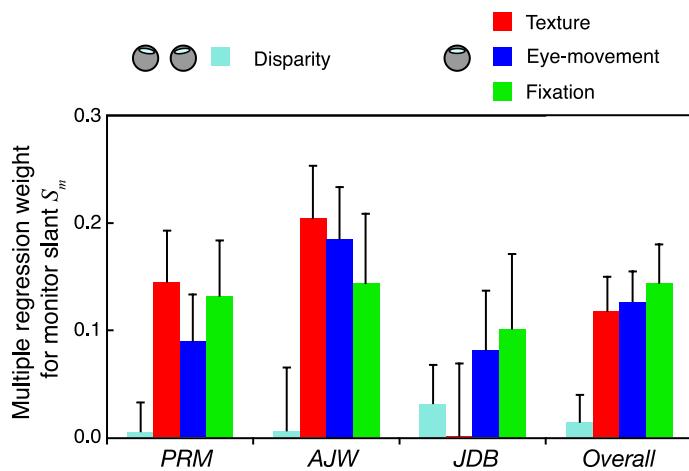


Figure 9. Regression weights for S_m in Experiment 1. The abscissa values are the three observers and an overall summary. Different colors represent different viewing conditions. The ordinate values are the multiple regression weights for S_m , obtained by entering the slant estimates in each case into a multiple regression analysis with S_m and S_s as factors. The overall weights were calculated by entering the data from all three observers into a single analysis. The regression weights are equivalent to the weights given to S_m in each condition, averaged across all values of S_m and S_s . Error bars are +95% confidence intervals for the regression weights.

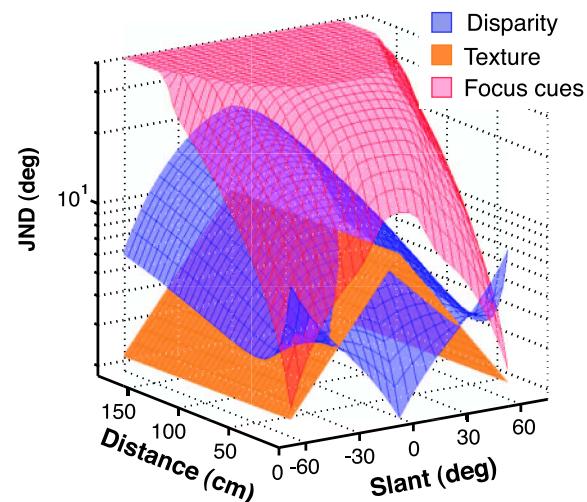


Figure 10. JND estimates for slant from disparity, texture, and focus as a function of slant and viewing distance. The different colored surfaces represent JNDs based on the individual cues. The disparity and texture JNDs were estimated from the data of Hillis et al. (2004) (see Appendix A). The focus JNDs were estimated by calculations described in Appendix A. The calculations determined how much slant would be required for the difference in defocus at the nearest and farthest points in the stimulus plane to exceed the visual system's depth of focus. The estimated JNDs from focus cues become very large at far distances and small slants, so the top portion of the focus surface has been clipped at 40° .

Discussion

Summary of results

With monocular viewing, observers' slant estimates were systematically affected by the orientation of the monitor surface (S_m). Observers reported seeing more slant when $S_m = S_s$ (as occurs with real stimuli) than when $S_m = 0$ (as usually occurs in psychophysical experiments and with most 3-D displays). The effect for the conditions of our experiment was small but quite consistent. These results show that information from focus cues (specifically, retinal blur) can, under monocular viewing, contribute directly to the visual system's estimate of 3-D surface orientation.

Evidence for reliability-based cue weighting

We next asked whether our findings are consistent with reliability-based cue weighting (Equations 1 and 2). To answer this, we first estimated the reliabilities of focus cues as well as texture and disparity cues for our stimuli. We then used those reliabilities to predict perceived slant for each combination of S_m and S_s in our experiment. Although we did not determine the single-cue reliabilities by our own experimental measurements, the exercise is useful for understanding the data.

According to [Equation 2](#), the reliability of each cue is the normalized reciprocal variance of the underlying estimator for that cue. To estimate this variance for the disparity and texture cues, we used previous slant discrimination measurements for each cue in isolation. To estimate this variance for focus cues, we simulated slant discrimination from blur using previous measurements of the visual system's depth of focus. [Figure 10](#) plots estimates of the JNDs for slant from disparity, texture, and focus cues as a function of surface slant (tilt = 0) and distance. Details of the calculations are provided in [Appendix A](#).

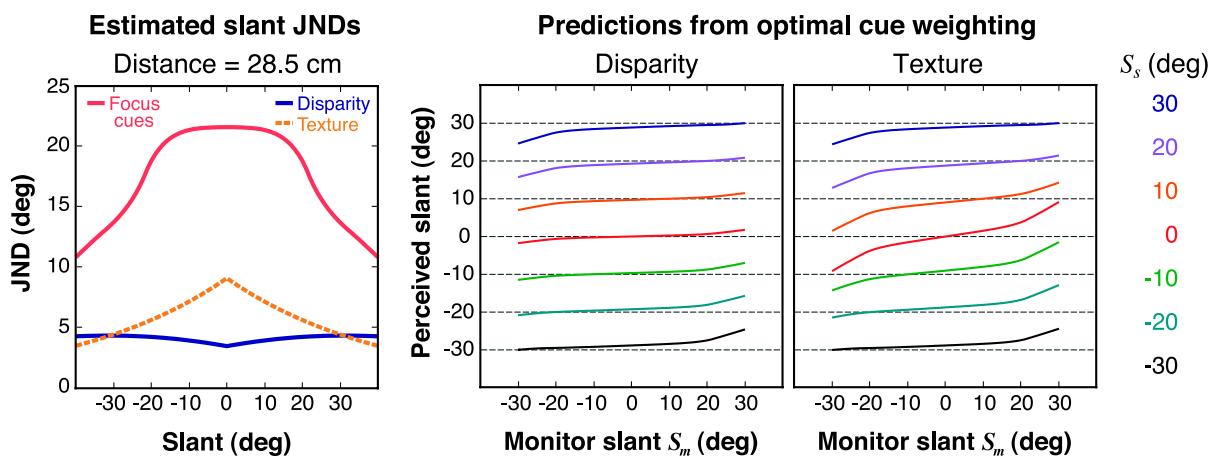
The texture JNDs were estimated from measurements made by Hillis et al. ([2004](#)) for monocularly viewed Voronoi patterns (see [Appendix A](#)). They are represented by the orange surface in [Figure 10](#). Texture JNDs decrease with increasing slant because the image changes associated with a given change in slant increase (Blake, Bültlhoff, & Steinberg, [1993](#); Knill, [1998](#)). Texture JNDs do not change with distance because doubling the size of a given textured surface and viewing it from twice the distance leaves the retinal image unchanged.

The disparity JNDs were derived from discrimination thresholds for slant from disparity alone (Hillis et al., [2004](#)), measured using sparse random-dot stereograms (see [Appendix A](#)). They are represented by the blue surface in [Figure 10](#). Disparity JNDs increase with viewing distance because the magnitude of binocular disparities for a given depth difference decreases with increasing viewing distance (Howard & Rogers, [2002](#); Ogle, [1950](#)). Disparity JNDs also vary with slant, which is expected from the viewing geometry (Hillis et al., [2004](#)). The variation is distance dependent: JNDs increase with slant at long viewing distances and decrease with slant at short ones (see also Banks, Hooge, & Backus,

[2001](#); Knill & Saunders, [2003](#)). The steep rise at large slant and short viewing distance probably reflects the influence of the disparity-gradient limit. In that situation, the horizontal disparity gradient increases significantly, and the two retinal images are difficult to fuse (Banks, Gepshtain, & Landy, [2004](#); Burt & Julesz, [1980](#); Hillis et al., [2004](#)).

We could not measure thresholds for slant from blur independent of other slant cues, but we can make a rough estimate of JNDs by considering how much slant would be required for the difference in defocus at the nearest and farthest points in the stimulus plane to exceed the visual system's depth of focus. We did this calculation for each combination of slant and distance in [Figure 10](#), using the same stimulus viewing frustum as [Experiment 1](#) (see the [Methods](#) section). Details are provided in [Appendix A](#). For our dim viewing conditions, pupil size was 5–7 mm (Wyszecki & Stiles, [1982](#)), so depth of focus was approximately ± 0.33 diopters (Campbell, [1957](#); Charman & Whitefoot, [1977](#); Green & Campbell, [1965](#); Green et al., [1980](#)). The red surface in [Figure 10](#) represents the estimated focus JNDs as a function of slant and distance.

The focus JNDs are generally larger than the disparity and texture JNDs, but the differences depend on slant and viewing distance. Specifically, focus-cue JNDs increase with increasing distance and decrease with increasing slant. The optimal cue-combination scheme ([Equations 1](#) and [2](#)) predicts therefore that focus cues should have little effect on 3-D percepts for many viewing situations. At short distances and large slants, however, focus JNDs can be equal to or less than those for disparity and texture. In these cases, optimal combination predicts a noticeable effect of focus cues.



[Figure 11](#). Estimated slant JNDs and predicted results for [Experiment 1](#). Left: Estimated JNDs for slant from disparity, texture, and focus cues, plotted as a function of slant at the 28.5 cm viewing distance used in [Experiment 1](#). The curves are a slice through the contours of [Figure 10](#). Middle: Predicted perceived slant for the disparity-defined stimulus. Right: Predicted perceived slant for the texture-defined stimulus. The format of the middle and right panels is the same as [Figures 5](#) and [7](#). The curves are plotted as a function of S_m ; each color represents a different value of S_s . The variance of each cue's slant estimate was calculated from the estimated JNDs in the left panel. The predicted perceived slants were calculated using those variances and the cue-combination scheme described by [Equations 1](#) and [2](#).

We can use the estimated JNDs to derive predictions of the effect of focus cues in the conditions of our experiment. The left panel in [Figure 11](#) plots the estimated JNDs for slant from disparity, texture, and focus cues for the range of slants ($\pm 30^\circ$) and the viewing distance (28.5 cm) used in [Experiment 1](#). From those JNDs, we estimated the standard deviations of the estimators associated with disparity, texture, and focus cues. Then using [Equations 1](#) and [2](#), we calculated the slants an observer would perceive if he weighted the three cues optimally. The middle and right panels show those predicted perceived slants, plotted in the same format as [Figures 5](#) and [7](#). In the disparity condition, the optimal cue combination predicts a small effect of monitor slant because the standard deviation of the disparity estimator is generally small relative to that of focus cues. In the texture condition, the model predicts a more systematic effect of monitor slant because in many cases the standard deviation associated with the competing cue—the texture gradient—does not differ very much from the standard deviation associated with focus cues.

Our empirical findings ([Figures 5](#) and [7](#)) are generally quite similar to these predictions. The data exhibit a small but consistent effect of monitor slant in the texture condition; that effect is similar in magnitude to the predicted effect. The data reveal no effect of monitor slant in the disparity condition, while a very small effect is predicted. From a multiple regression analysis of the predictions and data, we find that the average predicted weights given to focus cues in the disparity and texture conditions were 0.07 and 0.15, respectively, and that empirical weights were 0.01 and 0.12 ([Figure 9](#)).

Despite this general similarity, the model does not capture the details of the empirical findings. In particular, in the monocular viewing conditions we found a significant difference between slant estimates in the cues-consistent ($S_m = S_s$) and the $S_m = 0^\circ$ conditions. The model predicts only small differences between these conditions because focus-cue JNDs are large when $S_m = 0$. It is important to note that our predictions are based on a simple and untested model of how the visual system discriminates changes in slant from focus cues ([Appendix A](#)). We do not know how the brain actually computes slant from those cues. Therefore, it is quite possible that the discrepancy between the predictions and observed effects of focus cues resulted from inadequacies in our model. Furthermore, the reliability of slant estimates from focus cues surely depends on several factors including the spatial frequency, luminance, and contrast of the stimulus, as well as on fixation patterns and pupil size. Thus, a proper analysis would require empirical measurement of slant from focus for the stimuli used in the main experiment. Nonetheless, our analysis yields insight into the informativeness of focus cues as a function of slant and viewing distance, and the relationship to the informativeness of texture and disparity. Under reasonable assumptions, the pattern of effects across conditions in our empirical data was generally consistent with reliability-based cue weighting.

We examined only two conventional depth cues, disparity and texture, so it remains to be determined whether inappropriate focus cues also contribute to perceived depth for stimuli defined by other conventional cues.

Experiment 2

Overview and background

[Experiment 1](#) revealed that focus cues can have a *direct* effect on 3-D percepts. Accommodation could also affect perceived depth *indirectly* through the process of disparity scaling. The disparity (δ) created by two points in space is related to viewing distance as follows:

$$\delta \approx \frac{I\Delta D}{D^2}, \quad (3)$$

where ΔD is depth, D is viewing distance, and I is inter-pupillary distance (Howard & Rogers, [2002](#)). To recover ΔD from δ , D must be estimated. We know that viewing distance is estimated from the eyes' vergence and the horizontal gradient of vertical disparity (Rogers & Bradshaw, [1993](#), [1995](#)). In principle, it could also be estimated from accommodation. In computer displays, the focal distance to the display surface is fixed and often quite different from the simulated distances in the virtual scene. If the stimulus to accommodation (the focal distance of the display surface) affects the estimate of viewing distance, the distance to simulated points nearer than the display surface will be overestimated and the distance to points farther than the display surface will be underestimated. Such estimation errors might affect disparity scaling and hence the depth interpretation.

There have been many studies of disparity scaling (e.g., Bradshaw, Glennerster, & Rogers, [1996](#); Glennerster, Rogers, & Bradshaw, [1996](#); Johnston, Cumming, & Parker, [1993](#); O'Leary & Wallach, [1980](#); Rogers & Bradshaw, [1993](#), [1995](#); van Damme & Brenner, [1997](#)), but only one (Ritter, [1977](#)) examined the contribution of focal distance directly, and he observed no effect.

Frisby et al. ([1996](#)) observed veridical disparity scaling with real stimuli. The general consensus is that disparity scaling is most accurate when multiple cues are available and consistent with one another (e.g., vergence, vertical disparity, familiar size), but that scaling is usually non-veridical. At near viewing distances, the visual system behaves as if distance is overestimated, and at far distances, as if distance is underestimated (Collett, Schwarz, & Sobel, [1991](#); Foley, [1980](#); Glennerster et al., [1996](#); Johnston, [1991](#); Johnston et al., [1993](#); Rogers & Bradshaw, [1995](#); Wallach & Zuckerman, [1963](#)). Although many of these studies varied display distance and simulated distance concordantly, this pattern of results is also generally what

one would expect if focal distance (of a fixed display) affects the distance used for disparity scaling.

In [Experiment 2](#) we examined the contribution of accommodation to the estimate of the distance used to scale horizontal disparities. In particular, we examined the indirect influence of focal distance on disparity scaling by independently manipulating vergence distance (by varying absolute disparity) and focal distance, referred to hereafter as accommodative distance (by varying the distance to the display).

Methods

Observers

The experiment required observers to decouple vergence and accommodation, which many people find difficult ([Judge & Miles, 1985](#); [Wann & Mon-Williams, 1997](#)). We piloted the experiment on 12 observers and chose the four who could fuse the stimulus in all the conditions. They were 24, 25, 28, and 40 years old. Two had normal uncorrected vision and two wore their normal corrective lenses during the experiment. All four had normal stereoacuity and were experienced psychophysical observers. All were naïve to the specific purposes of the experiment.

Apparatus

We used the same apparatus and stimulus-rendering techniques as in [Experiment 1](#) except that only the stimulus display monitor was used. Monitor slant was always zero. The observer's head position was restrained using the same bite-bar apparatus. Distance to the display was varied by moving the bite bar relative to the monitor. Spatial calibration was done for each eye at each of the three viewing distances. To ensure precise, repeatable positioning of the observers, the table holding the bite bar was fixed with drilled holes in the floor.

Stimuli and task

We used a task similar to the apparently circular cylinder task ([Johnston, 1991](#)). The stimuli were concave hinges: two planes slanted about the vertical axis (tilt = 0°) and joined at their far point to form an “open book”. The dihedral angle between the two planes is the *hinge angle*. Observers indicated whether the perceived hinge angle was larger or smaller than 90°. With a related stimulus, [Johnston, Cumming, and Landy \(1994\)](#) showed that observers perceive the 3-D shape veridically when depth cues (disparity, texture gradient, and motion parallax) are consistent. The surfaces were defined by sparse randomly positioned squares. Square size and density were constant at 0.18° and 1.6 squares/deg², respectively. The width and height of the stimuli measured from the cyclopean eye were on average constant across hinge angles at 8.5° and

2°, respectively. A small random perturbation was added to both on each trial. The stimulus was clipped by an elliptical aperture to make the outline shape uninformative. To create the stimulus, we first rendered two frontoparallel grids of regularly spaced squares, one for each plane of the hinge. Each square's position was then jittered by a random amount horizontally and vertically in the range ±1.25 times the inter-square separation. Each plane was then rotated about the vertical axis by the appropriate amount for the desired hinge angle. Overlapping squares at the intersection of the hinge were deleted. The position of each square on the display surface was determined separately for the left- and right-eye's images by calculating where projections from the each eye's position intersected the monitor plane (calculated for each observer's interpupillary distance); the texture gradient was always appropriate for the slants of the two planes and the individual squares had disparities consistent with the simulated surface slant.

We were careful not to introduce uncontrolled cues into our stimulus that could confound the measurements. Previous studies have often used dense random-dot stereograms in which disparity was introduced by shifting dots horizontally in each eye's view. In such stimuli, the monocular texture gradient at each eye specifies a frontoparallel surface, which could cause objects to appear flatter than if disparity were the only informative cue. Moreover, the reliability of the disparity cue decreases with increasing distance while the reliability of the texture cue does not ([Hillis et al., 2004](#)), so the texture cue would likely be given more weight with increasing viewing distance, causing the stimulus to appear increasingly flattened. This is the same pattern of results that would be produced by misestimates of the distance for disparity scaling ([Hillis et al., 2004](#)). To minimize the probability of this bias, perspective projection of the stimulus was correct in each eye's view (the texture cue was consistent with disparity). We also attempted to minimize the contribution of the texture cue by presenting few stimulus elements. A pilot experiment confirmed that observers could not do the task with monocular information. Although vertical disparities were correct for the simulated distance, we minimized their influence by using short stimuli ([Backus et al., 1999](#); [Rogers & Bradshaw, 1995](#)). In this way we could isolate the effects of vergence and accommodation. Of course, our stimulus contained a blur gradient consistent with a flat surface, but the failure to observe an effect of blur gradient with disparity-defined stimuli in [Experiment 1](#) implies that the gradient did not affect performance in the direct sense in [Experiment 2](#).

On each trial, the entire hinge stimulus was rotated around the vertical axis (at the intersection of the hinge) by an angle chosen randomly in the range ±10°, so observers could not do the task by estimating the slant of one plane. A fixation cross was drawn in the center of the stimulus at the same depth as the intersection of the hinge planes. The cross was a good stimulus to accommodation,

although it was slightly blurred by anti-aliasing. Because the cross was displayed at the intersection of the hinge, its vergence- and accommodation-specified distances were unaffected by changes in hinge angle. The room was dark except for the stimulus. The frame of the monitor was not visible.

We were concerned that observer biases would affect the settings and would therefore affect the interpretation of the data. So we ran a pretest with a real hinge in which all cues were available and consistent and compared those data with the data from the computer-displayed stimuli. The real hinge consisted of two plywood planes ($\text{tilt} = 0$) joined at their inside edges. The visible surfaces were covered with paper on which a Voronoi pattern like the one in [Experiment 1](#) was printed. Pattern size was changed at each viewing distance so that the average angular size of the Voronoi cells and the line thickness were constant. The stimulus was illuminated by a diffuse light positioned so that there was no detectable variation in shading with changes in hinge angle. Observers viewed the stimulus through circular apertures, one for each eye, so that the visible portion of the stimulus had a diameter of 20° . At this size vertical disparities could provide reliable distance information (Rogers & Bradshaw, [1995](#)). Head position was restrained using a chin-and-forehead rest. Nothing was visible except the hinge stimulus itself. The hinge was moved up or down behind the apertures after each trial so that different parts of the Voronoi patterns were presented on each trial. The entire hinge assembly was rotated from trial to trial in the range $\pm 10^\circ$ (as with the simulated hinge). Viewing distance was varied by positioning the observer at different distances from the apparatus.

In the simulated surface conditions, each combination of accommodation distance (D_a) and vergence distance (D_v)—28.5, 57.0, and 85.5 cm—was presented for a total of nine conditions. In the real-surface condition, the same set of distances was presented for a total of three conditions.

The hinge task has significant advantages over other methods. For instance, it avoids the problem of response quantization (which can occur when observers are asked to report numerical estimates), and the problem of also having to estimate width in the apparently circular cylinder task (Johnston, [1991](#); Johnston et al., [1994](#)).

Procedure

On each trial the fixation cross was presented for 2 s, followed by the hinge stimulus and the fixation cross for another 2 s. Observers were told to fixate the cross throughout each presentation. If they were unable to fuse the cross before the stimulus appeared, the trial was discarded; this rarely occurred. Observers indicated on each trial whether the hinge angle was larger or smaller than a right angle.

We used 2-down/1-up and 1-down/2-up staircase procedures to vary the hinge angle. Each staircase was terminated after 12 reversals, resulting in 120–150 trials per condition. The responses were used to construct psychometric functions (percentage of responses that the angle was larger than 90° as a function of the specified angle). The 50% point was estimated by a maximum-likelihood procedure (Wichmann & Hill, [2001](#)). That point served as the estimate of the dihedral angle that was perceived as 90° ; we refer to this angle as the PSE. Our method allowed rapid measurement, which was important because vergence accommodation dissociations can cause response adaptation and fatigue (Schor & Tsuetaki, [1987](#)).

Before each trial in the real-surface condition, observers' eyes were closed as the experimenter set the hinge to the appropriate angle for the next presentation. They then opened their eyes and indicated whether the hinge angle was larger or smaller than 90° . Then they closed their eyes in preparation for the next trial. Viewing time was not strictly controlled but was usually 2–3 s. The hinge angle was again varied according to 2-down/1-up and 1-down/2-up staircases.

The simulated surface conditions were run in blocks in which D_v was varied and D_a was constant. There were three values of D_v , so a block consisted of three randomly interleaved staircases. The two staircase rules were run in different blocks. D_a was varied across blocks. Thus, there were six blocks in the simulated surface condition: three values of D_a and two staircase rules. The real-surface conditions were run with one distance and one staircase rule in each block. There were again six blocks, but the blocks were briefer. Each observer completed all 12 blocks in random order, over several days.

It is important to distinguish the stimulus to vergence from the vergence response, and the stimulus to accommodation from the accommodative response. We manipulated the stimuli to vergence and accommodation, which presumably produced changes in the responses, but we do not know how well correlated the responses were with changes in the stimuli because we did not measure the responses per se. Therefore, as we discuss the means by which vergence and accommodation contribute to distance estimation in disparity scaling, we will refer to the stimuli—vergence-specified distance (D_v) and accommodation-specified distance (D_a)—and not to the responses.

Results

Equivalent distance

The PSE in different conditions—the hinge angle perceived on average as a right angle—was used to determine the equivalent distance in the various conditions. The method for calculating the equivalent distance is schematized in [Figure 12](#). Each hinge angle PSE corresponds to a pattern of horizontal disparities that was perceived as

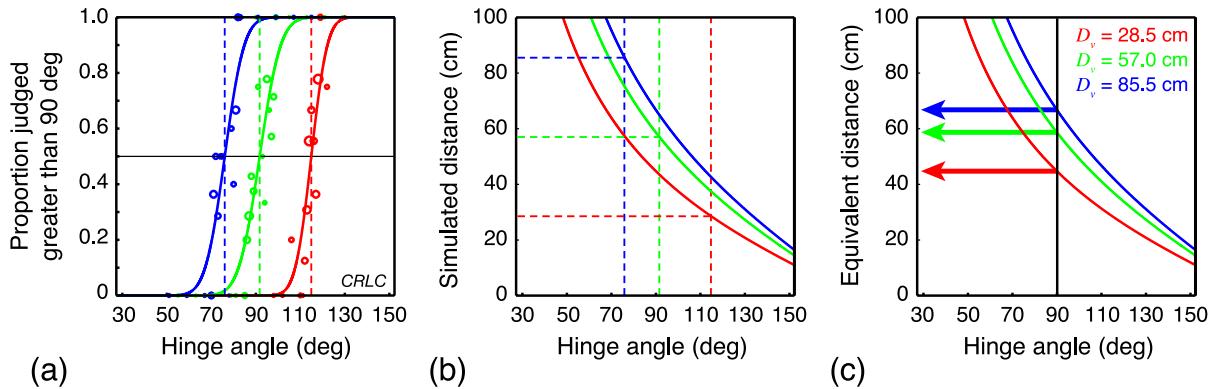


Figure 12. The calculation of equivalent distance for Experiment 2. (a) Example psychometric functions for observer CRLC at an accommodative distance (D_a) of 57 cm. The proportion of trials in which he responded “greater than 90° ” is plotted as a function of the hinge angle. Different colors represent different vergence distances (D_v). The vertical lines indicate the PSE, the hinge angle that was judged as greater than 90° 50% of the time. The size of the data points is proportional to the number of trials at that point. (b) Each curve is an “iso-disparity” line showing different hinge angles/distances consistent with the pattern of disparities defined by each of the PSEs from panel a. The horizontal lines show the simulated distance in each case. (c) The curves are the same as in panel b. The arrows show the distance at which the pattern of disparities associated with each PSE are actually consistent with a 90° hinge angle (using the relationship in Equation 4). This is the equivalent distance.

a right angle. The disparity pattern can be expressed as the horizontal size ratio (HSR)—the ratio of the widths of a small surface patch in the left- and right-eye’s images (Rogers & Bradshaw, 1993)—for both planes of the hinge. For straight-ahead viewing, slant is

$$S = \arctan\left(\frac{2-D(\text{HSR}-1)}{I(\text{HSR}+1)}\right), \quad (4)$$

where D is viewing distance and I is inter-pupillary distance (Howard & Rogers, 2002; Ogle, 1950). Equation 4 shows that a given HSR is consistent with many slants and therefore many hinge angles depending on what the distance is.

Figure 12a plots example psychometric functions at an accommodation distance (D_a) of 57 cm for each of three vergence distances (D_v). Figure 12b shows the range of hinge angles that is consistent with the disparity pattern specified by the PSE (calculated by rearranging Equation 4, and considering the two planes of the hinge separately). The range of hinge angles consistent with the disparity pattern is large because the disparities specify different angles at different distances. Assuming that the observer’s internal standard for 90° is unbiased (and that the visual system measures disparities without bias), the disparity pattern associated with the observer’s setting would specify a right angle at some distance. This is equivalent distance shown in Figure 12c.

Figure 12a shows that for the near vergence distance (28.5 cm, red lines and symbols) an angle larger than 90° looked like a right angle. The distance at which this disparity pattern specifies a 90° hinge angle is 44.8 cm (red curve and arrow in Figure 12c). This suggests that

viewing distance, which was 28.5 cm according to vergence, was overestimated. The blue lines and symbols denote the data for the far vergence distance ($D_v = 85.5$ cm); they show the converse pattern, consistent with an underestimate of viewing distance. Performance at the middle distance (green) was close to veridical. These data show a pattern of underconstancy with respect to changes in vergence-specified distance, in which near distances are overestimated and far distances are underestimated.

Figure 13 shows the equivalent distances for the simulated and real surfaces as a function of the vergence-specified distance (D_v). The black points and lines represent the data from the real-surface measurements in which all cues were consistent and the colored points and lines represent the data from the simulated surface measurements, each color representing a different accommodative distance (D_a). The circled points are the cues-consistent data: the subset of simulated surface data for which $D_a = D_v$.

If the visual system did not adjust its distance estimate for disparity scaling with changes in the vergence- or accommodation-specified distances, the data would lie on a horizontal line. If the system scaled by using only the vergence-specified distance, the data would lie on the same diagonal for all accommodation-specified distances. An effect of accommodation distance would cause the data to separate vertically. The simulated surface data reveal a clear effect of accommodation-specified distance, which suggests that focal distance affected the visual system’s estimate of viewing distance. The effect was most systematic for observers JMA and AKB: For both of them, an increase in D_a produced

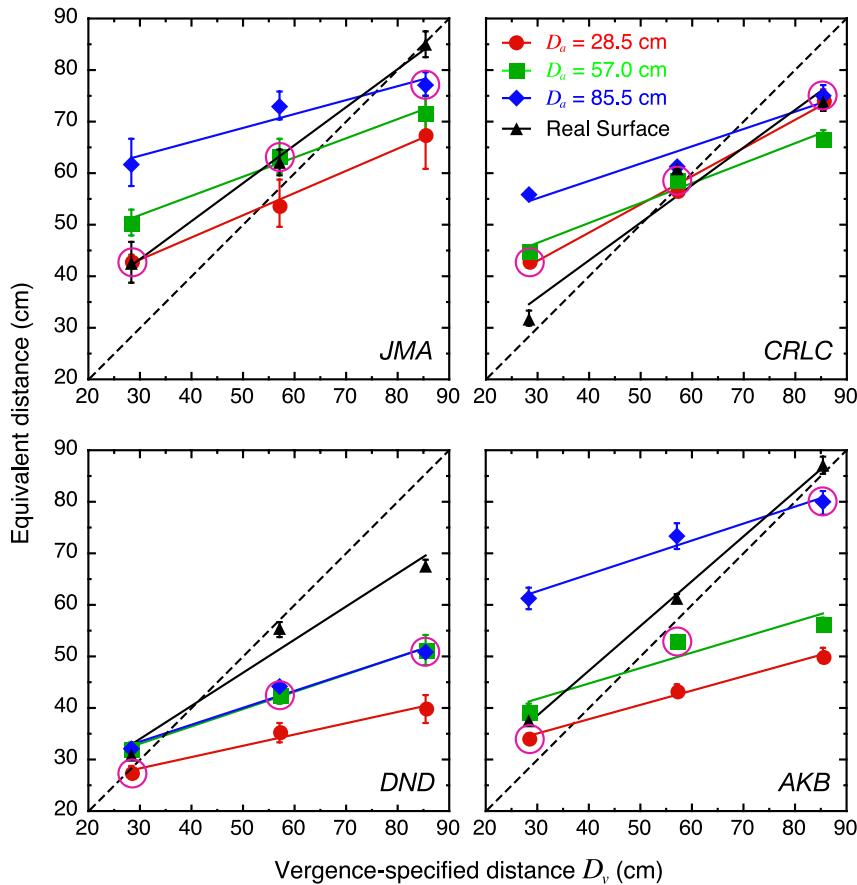


Figure 13. Equivalent distance as a function of vergence-specified distance in Experiment 2. The panels show data from different observers. The dotted diagonal lines represent veridical performance with respect to changes in D_v . The red, green, and blue symbols represent the simulated surface data for $D_v = 28.5$, 57.0 , and 85.5 cm, respectively. The colored lines are the best fits to these data. The data points for the cues-consistent conditions ($D_a = D_v$) are circled. The black symbols represent the real-surface data. Error bars represent ± 1 SE of the equivalent distance estimate. They were derived by taking ± 1 SE of each PSE, and then computing the equivalent distance for these points in the same manner as the average data points.

a very consistent increase in equivalent distance. The effect of D_a was somewhat less consistent in the other two observers (for discussion of individual differences, see Appendix B).

The cues-consistent (circled points) and real-hinge settings were very similar for all but observer DND. Thus, when accommodation- and vergence-specified distances are the same in a computer display, changes in viewing distance are taken into account as well as they are with real stimuli.

It is interesting to note that the smallest effects of accommodative distance were manifest by the two oldest observers, DND and CRLC. At 40 and 28 years, respectively, they may have been less able to accommodate accurately due to emerging presbyopia (see the section Does accommodation provide a distance signal?).

Depth constancy

We can describe the degree to which the visual system took vergence distance changes into account from the

slopes of the fitted lines in Figure 13. A value of 1 means complete depth constancy based on vergence distance and 0 means no constancy. The amount of vergence-based depth constancy for each observer in each condition is shown in Figure 14. Clearly some constancy occurred in all observers and conditions, which shows that D_v contributed significantly to disparity scaling in all cases. However, D_a also affected depth constancy. Constancy was least when D_a was fixed and was consistently greater when $D_a = D_v$. Indeed, the amount of depth constancy in the cues-consistent simulated condition approached the amount in the real-surface condition.

We also wondered whether conflict between vergence- and accommodation-specified distances would reduce sensitivity to changes in hinge angle. We examined this by comparing the slopes of the psychometric functions in the cues-consistent and cues-inconsistent conditions. There was no systematic effect, so conflict did not seem to reduce sensitivity. Perhaps error in measuring the disparities and/or imprecision in the observers' internal standard for 90° were the limits to sensitivity.

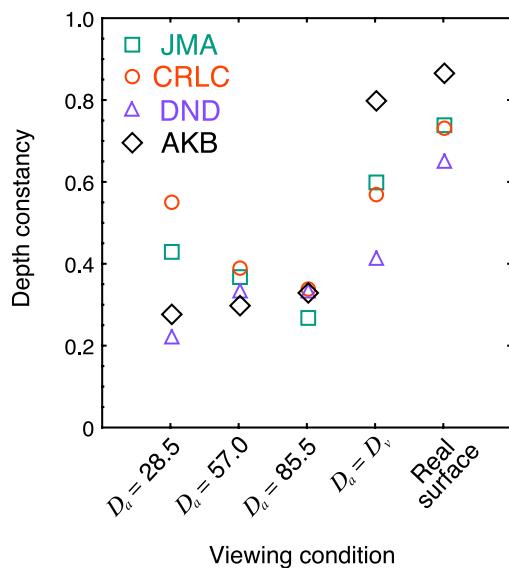


Figure 14. The amount of stereoscopic depth constancy with respect to changes in vergence-specified distance (D_v) for each observer in Experiment 2. The abscissa values are the different conditions (including the subset of cues-consistent data for which $D_a = D_v$). The ordinate values are the slopes of the lines relating equivalent distance to D_v in each case. The different symbols indicate different observers.

Discussion

Summary of the results

Perceived depth from disparity was consistently affected by variations in the physical distance to the display surface. Specifically, stereoscopic depth constancy was greater when the vergence- and accommodation-specified distances were equal ($D_a = D_v$) than when the accommodation-specified distance was constant. This finding is consistent with previous reports that perceived distance is a compromise when vergence and accommodation specify different distances (Lipson, 2001; Ono & Comerford, 1977; Ono, Mitson, & Seabrook, 1971; Swenson, 1932). We conclude that focus cues contribute to the visual system's estimate of distance and thereby exert an indirect influence on perceived depth through the process of disparity scaling.

Does accommodation provide a distance signal?

For two observers—JMA (24 years old) and AKB (25 years)—the effect of varying D_a was very systematic, and the equivalent distances were consistent with a weighted average of D_a and D_v . The other two—DND (40 years) and CRLC (29 years)—exhibited clear but less systematic effects of accommodative distance. The fact that the youngest two observers, who were presumably most able to accommodate to the whole stimulus range, exhibited the most systematic effects suggests that

accommodative range could explain these differences. We examined the relationship between the effect of accommodative distance in Experiment 2 and clinical measurements of (1) accommodative range, (2) the ratio of convergence accommodation over convergence (CA/C), and (3) the ratio of accommodative convergence over accommodation (AC/A). Those results are provided in Appendix B.

The main result of Experiment 2 is that disparity scaling is affected by accommodative distance, so focus cues influence perceived depth in an indirect fashion. We cannot prove that the accommodative *response* rather than the accommodative *stimulus* is the key variable because we did not measure accommodative responses during the experiment. For our purposes, we have shown that variations in focal distance (the stimulus to accommodation) affect the distance estimate used in disparity scaling and thereby affect depth perception.

Depth constancy for real versus simulated surfaces

We observed significant depth constancy with cues-consistent simulated surfaces, essentially as much as with real surfaces. The similarity is somewhat surprising because the real surfaces in principle provided better distance and slant information than the simulated ones for a handful of reasons: (1) higher luminance in the real-surface condition, (2) sharper edges in the real stimulus (which would presumably provide a better stimulus to accommodation, (3) the real surfaces were larger horizontally and vertically than the simulated surfaces (which should have improved the reliability of the vertical-disparity signal to viewing distance), and (4) richer texture in the real stimulus. Apparently, none of these differences between the real hinge and the simulated cue-consistent hinge influenced depth constancy significantly.

We wondered why we failed to observe complete constancy with the real-hinge stimulus while others have observed veridicality with such stimuli (Frisby et al., 1996). One cause could be the lack of co-variation between distance and projected size. In our experiment, the angular sizes of the hinge stimulus (and of the texture pattern) were constant despite a 3-fold change in distance. In the study of Frisby et al. (1996), projected size covaried in normal fashion with distance. Collett et al. (1991) reported that equivalent distance is influenced by the angular size of the stimulus. In their experiments, depth constancy was consistently greater when the angular size and texture density of the stimulus varied appropriately with viewing distance compared to when the projected sizes were held constant. Collett et al. suggested that constant angular size is interpreted as specifying a constant viewing distance, which leads to a bias in disparity scaling.

General discussion

Implications for psychophysical research

The fact that focus cues can affect 3-D percepts has significant implications for psychophysical research on visual space perception. Here we consider some of those implications.

Retinal image motion caused by relative movement between the observer and an object can create a vivid 3-D impression, but the judged depth of the object is often less than the simulated depth (Braunstein, Liter, & Tittle, 1993; Caudek & Proffitt, 1993; Domini and Caudek, 1999; Loomis & Eby, 1988; Todd & Bressan, 1990). Here we consider a series of elegant studies by Hogervorst and Eagle (1998, 2000) because they exemplify the potential significance of not including focus cues in the interpretation of the data.

Hogervorst and Eagle (1998, 2000) presented monocular structure-from-motion displays simulating two hinged planes, much like the stimulus in our [Experiment 2](#). Observers indicated the perceived angle between the planes. The authors examined the effects of perspective projection (versus orthographic) and field of view. The most important data for our purposes are from those stimuli in which 3-D shape was best specified: perspective projection and large field of view (Figures 2a and 2c in Hogervorst & Eagle, 1998; Figure 4a in Hogervorst & Eagle, 2000). The results revealed large overestimations of the hinge angle (underestimations of the depth) across a variety of conditions. The results were well predicted by a Bayesian model incorporating all flow measurements and their associated noises. However, depth underestimation would also be expected if focus cues affected the observers' percepts because they would have signaled the constant depth of the computer display rather than the varying distance of the hinge. Hogervorst and Eagle were concerned about this possibility (Hogervorst & Eagle, 1998, p. 1589; Hogervorst & Eagle, 2000, p. 945), so they conducted a control experiment to assess the contribution of focus cues. One observer repeated the main experiment while viewing through a pinhole. The authors reasoned that if focus cues had caused perceptual flattening, using a pinhole should have rendered focus cues uninformative and thereby yielded percepts of greater depth. The results showed that pinhole viewing had no effect (Hogervorst & Eagle, 1998) or caused a slight decrease in perceived depth (Hogervorst & Eagle, 2000), both results being inconsistent with their expectation if focus cues had affected the original data. We argue in the [Minimizing the contribution of focus cues](#) section and [Appendix C](#) that this is an inappropriate control: using pinholes does not necessarily render focus cues uninformative but rather may cause the blur gradient and accommodative stimulus to be interpreted as specifying flatness.

In summary, by not including signals that may well have affected observers' percepts, Hogervorst and Eagle's (1998, 2000) theory may be incorrect because it may be based on data contaminated by an unmodeled cue. The same criticism potentially applies to other data and theories in the literature. The construction of an appropriate theory requires data, whether they exhibit veridicality of perception or not, that are uncontaminated by unmodeled variables. One cannot determine with certainty which results have been affected and which have not, but we offer suggestions in the [Minimizing the contribution of focus cues](#) section.

Minimizing the contribution of focus cues

Some researchers have argued that running experiments with and without pinholes in front of the eye(s) provides an adequate test for the influence of focus cues (e.g., Frisby et al., 1995; Hogervorst & Eagle, 2000). The argument is that a pinhole removes those cues as sources of depth information, so if accommodation and/or blur had contributed to the observed depth estimation—usually depth underestimation—the results with pinholes present should differ from the results without pinholes. In their studies of depth perception with computer-displayed stimuli, Hogervorst and Eagle (2000) and Frisby et al. (1995) observed no effect of using pinholes, so they concluded that accommodation and blur had not contributed to the depth underestimation they observed.

Results from Frisby et al. (1995) make clear that the above reasoning concerning pinhole usage is erroneous. The authors had observers view computer-displayed and real ridges binocularly with and without pinholes. They found that using pinholes had no effect on the perceived depth of computer-displayed ridges. However, using pinholes caused flattening of the perceived depth associated with real ridges. They pointed out that viewing through pinholes renders the blur in the retinal image similar for a wide range of distances and causes the eye to adopt a fixed focal length. This they argued causes no change in the signals arising from the computer-displayed stimuli, so perceived depth was unaffected. In contrast, the increased depth of focus changes the signals arising from real 3-D objects—the focus cues now signal "flat" as they do with computer displays—so percepts became flatter. This shows that using pinholes is not an adequate method for eliminating the influence of focus cues.

In most experiments on depth perception, the depth cues of interest are placed in conflict with one another so the experimenter can determine their relative contributions. Let us refer to the cues of interest as the experimental depth cues. There are experimental manipulations that should minimize the contaminating influence of focus cues.

1. Increase the availability and reliability of the experimental depth cues. One can see from [Equations 1](#) and

- 2 that adding reliable depth information should cause a decrease in the relative weight assigned to focus cues.
2. Making focus cues as consistent with the depth specified by the experimental cues as possible. Again from Equations 1 and 2 one can see that minimizing the conflict between experimental cues and focus cues will reduce biases caused by the contribution of focus cues. Experiment 1 showed that making the slant specified by focus cues consistent with the slant specified by the experimental cue of texture yielded a significant increase in perceived slant when compared to the conventional situation, in which focus cues specified a slant of zero (Figures 6 and 8). Experiment 2 showed that making the physical distance to the display the same as the simulated distance yielded a significant increase in depth constancy (Figures 13 and 14).
 3. Increasing the distance to the display. The informativeness of many depth cues does not decrease as rapidly with distance as the informativeness of focus cues, so increasing the viewing distance should decrease the influence of focus cues (Figure 10, Appendix A).

We hasten to point out that the abovementioned points are qualitative guidelines; we cannot currently delineate the precise set of viewing conditions for which inappropriate focus cues from the display adversely affect 3-D percepts.

3-D displays with correct focus cues

3-D computer graphics has become increasingly important for many applications including operation of remote devices, scientific visualization, education, training, computer-assisted design and virtual prototyping, and entertainment (e.g., Hunter & Sackier, 1993; Wickens, Merwin, & Lin, 1994). It is important in these applications for the computer-graphic image to create a realistic impression of the 3-D structure of the object or scene being portrayed. Consider, for example, telesurgery (Rassweiler, Binder, & Frede, 2001; Stanberry, 2000). A surgeon at a remote site views the patient's tissue on a digital display. Our research suggests that the depth the surgeon perceives may not be correct because focus cues signal the distance to the display rather than the distances to the tissue. In addition, the decoupling of vergence and accommodation required by conventional displays causes discomfort (Wöpking, 1995), binocular stress (Mon-Williams, Wann, & Rushton, 1993; Wann, Rushton, & Mon-Williams, 1995), and difficulty fusing the images of a stereo pair (Akeley et al., 2004; Wann et al., 1995). It is not surprising that several researchers and companies are developing new display technologies that are meant to minimize the adverse effects of inappropriate focus cues.

One solution has been to fix the focal distance at infinity by collimating the light from the display (North & Wooding, 1970) or positioning the display surface far from the viewer.

This provides a good approximation to reality when all points in the scene are far from the observer (when looking out from aircraft or spacecraft windows, for example). Unfortunately, such systems fail for virtual objects close to the viewer and hence cannot work for general settings.

Autostereoscopic volumetric displays present the scene as a volume of light sources (Downing, Hesselink, Ralson, & Macfarlane, 1996; Favalora et al., 2002; Perlin, Paxia, & Kollin, 2000; Suyama, Date, & Takada, 2000; Suyama, Takada, Uehira, & Sakai, 2000; Suyama, Takada, Uehira, & Sakai, 2001). Such displays provide focus cues that are consistent with the geometric depth cues, but they do not create realistic images because view-dependent effects such as occlusions, specularities, and reflections cannot be simulated properly.

Non-volumetric approaches to correcting focus cues include displays that adjust the focal distance of the entire image to match the viewer's accommodation, which must be estimated by tracking the direction of gaze (Omura, Shiwa, & Kishino, 1996). They also include displays that adjust focal distance for regions of the image, with the ultimate goal of pixel by pixel adjustment (McQuaide et al., 2002; Silverman, Schowengerdt, Kelly, & Seibel, 2003). These techniques are limited by their inability to present multiple focal distances in a given visual direction.

Following a concept discussed by Rolland, Krueger, and Goon (1999), Akeley et al. (2004) developed a fixed viewpoint, volumetric display that creates nearly correct focus cues for distances of ~28–65 cm. The display enables view-dependent lighting effects such as occlusion, specularity, and reflection. To demonstrate the utility of the display, Akeley et al. compared the time required to fuse a stereogram when the accommodation-specified distance was fixed (as in a conventional display) to that required when accommodation-specified distance was consistent with disparity-specified depth. Fusion time was significantly reduced when the cues were consistent. They also observed informally that viewer discomfort was reduced, and that the 3-dimensionality of simulated scenes was more convincing.

We hope that the research reported here will provide further motivation to develop display technologies for basic research and for applied settings that minimize the adverse effects of inappropriate focus cues.

Conclusions

We performed two experiments to examine whether information from focus cues contributes to perceived depth. The results of Experiment 1 showed that focus cues can contribute directly to estimates of perceived 3-D scene properties under some circumstances, and that this can be mediated by changes in the blur gradient alone. The finding that focus cues affected slant estimates for texture-defined stimuli but not for disparity-defined stimuli is

consistent with reliability-based cue weighting. We also showed in [Experiment 2](#) that focus cues contribute indirectly to 3-D percepts by influencing the process of disparity scaling.

Because blur and accommodation affect 3-D percepts, inappropriate focus cues in typical displays can contribute to biases in perceived 3-D scene structure under some conditions.

Appendix A: Informativeness of blur, disparity, and texture cues to slant

The disparity and texture JNDs plotted in [Figure 10](#) were derived from subject JMH's slant discrimination data presented by Hillis et al. ([2004](#)). Hillis et al. measured slant discrimination thresholds based on either disparity (sparse random-dot stereograms) or texture (monocularly viewed Voronoi patterns) alone as a function of base slant ($\pm 60^\circ$) at three distances (19.1, 57.3, and 171.9 cm). The experiment used a 2-IFC procedure, so we divided the threshold values (84% correct) by $\sqrt{2}$ to obtain the standard deviations of the underlying estimator in each condition (Ernst & Banks, [2002](#)). To generate the surface plots in [Figure 10](#), we fitted the following function with the parameters p_1 and p_2 to the derived standard deviations using the maximum-likelihood method:

$$\text{JND} = p_1 e^{p_2 S}.$$

The variable S was slant for the texture cue and HSR for the disparity cue.

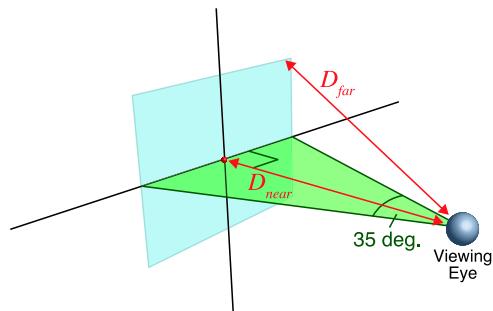
To our knowledge, there have been no measurements of slant discrimination thresholds when focus cues are the only available cue. Thus, we cannot determine from empirical data what slant JNDs from focus cues are and how they vary with slant and viewing distance. We decided therefore to estimate the likely thresholds in a simulation. To do the simulation, we naturally had to make assumptions about the underlying processes.

We simulated a 2-IFC procedure. Two planes were presented, differing in slant ($\text{tilt} = 0$). The simulated planes were those of [Experiment 1](#) except that we varied them over a larger range of slants and distances. The viewing frustum defined the visible portion of the stimulus. The horizontal angular subtense of the frustum was always 35° . The vertical angular subtense was always 28° at the axis of rotation of the stimulus; this defined the height of the plane, which at a given distance was constant at all slants (see [Figure A1](#)).

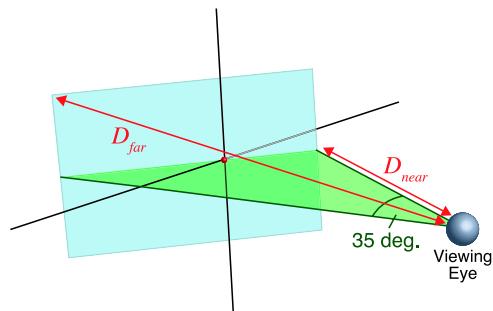
From the data in [Experiment 1](#), we concluded that the blur gradient rather than accommodation was the

critical aspect of the focus cue for slant. So we based our analysis on blur. For each of the two planes in a trial, we calculated the nearest and farthest points as shown in [Figure A1](#). The distances to those points were expressed in diopters, the reciprocals of the distances were expressed in meters. For each plane, we calculated the difference between nearest and farthest dioptric distances to yield ΔD_i , where i is the stimulus interval (1 or 2). We then computed the absolute value of the difference between those two values: $\Delta D = |\Delta D_1 - \Delta D_2|$ to obtain a measure of how different the blur gradient in the two stimuli was. To find threshold, we varied the slant increment (or decrement) relative to the base slant to find the value that yielded $\Delta D = 0.66$ (the assumed depthof focus for our viewing situation was ± 0.33 diopters; Campbell, [1957](#); Charman & Whitefoot, [1977](#)). We then averaged the increment and decrement thresholds to derive one threshold value for each simulated viewing condition, and these are plotted as JNDs in [Figures 10](#) and [11](#).

Frontoparallel plane ($S_m = 0$ deg)



Negative slant ($S_m < 0$ deg)



[Figure A1](#). Determination of nearest and farthest points for calculation of slant-from-focus JNDs. For each of the two stimuli in the 2-IFC presentation, we calculated the nearest and farthest point from the viewing eye. The farthest point was always in the far upper (or lower) corner of the plane where it intersected the viewing frustum. The nearest point was sometimes along a surface normal (upper panel), and sometimes on the near edge of the plane at the horizontal midline where the plane intersected the viewing frustum (lower panel). We calculated the distances to the nearest and farthest points and expressed them in diopters (the reciprocal of the distance in meters).

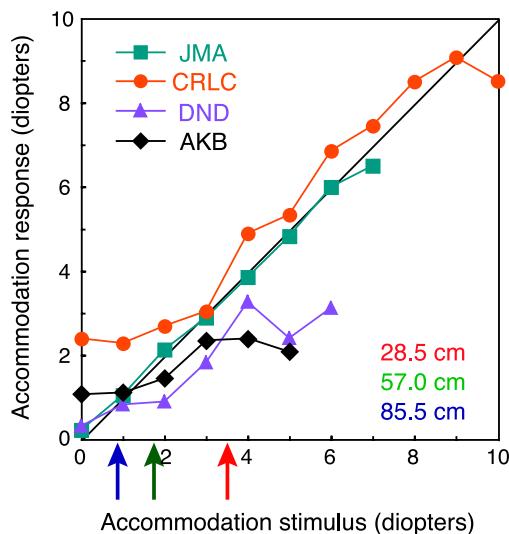


Figure B1. Accommodation stimulus–response functions for each observer in Experiment 2. Accommodative state was measured subjectively using a Badal optometer while observers viewed monocularly the hinge stimulus from Experiment 2. Focal distance was varied using trial lenses. The red, green, and blue arrows on the abscissa indicate the focal distance to the stimulus in the $D_a = 28.5$, 57.0, and 114 cm conditions, respectively.

Appendix B: Individual differences

We observed somewhat different behavior from different observers in Experiment 2. Observers AKB and JMA exhibited very consistent effects of changes in accommodative distance while observers DND and CRLC exhibited somewhat less consistent effects. Because JMA (24 years old) and AKB (25 years old) were younger than CRLC (29 years old) and DND (40 years old), we wondered whether emergent presbyopia in the older observers contributed to a less consistent effect. To examine this, we made three measurements associated with accommodation and convergence using standard clinical techniques: (1) accommodation to a range of distances, (2) the ratio of convergent accommodation over convergence (CA/C), and (3) the ratio of accommodative convergence over accommodation (AC/A).

We measured each observer's accommodation stimulus–response function using a Badal optometer while they viewed the hinge stimulus from Experiment 2. Viewing conditions were the same as during the main experiment except the stimulus was viewed monocularly and accommodation was stimulated using trial lenses. The results for each observer are plotted in Figure B1. The colored arrows on the abscissa indicate the three accommodative distances (D_a) used in Experiment 2 (3.51, 1.75, and 0.88 diopters). JMA and CRLC exhibited

the greatest range of accommodative response. DND and AKB exhibited smaller ranges. Thus, differences in accommodation stimulus–response functions do not directly predict the individual differences we observed in Experiment 2.

We also measured CA/C and AC/A ratios in each observer. CA/C is the amount of accommodation elicited by changes in convergence when there is no defocus stimulus to accommodation. AC/A is the amount of convergence elicited by changes in accommodation when there is no disparity stimulus to vergence. JMA and AKB had CA/C ratios of 0.78 and 0.92 (diopters/meter angles), whereas CRLC and DND had higher CA/C ratios of 2.08 and 1.60. JMA and AKB had AC/A ratios of 5.70 and 3.76 (prism diopters/diopters), whereas CRLC and DND had higher ratios of 7.04 and 6.56. The lower CA/C and AC/A ratios mean that JMA's and AKB's vergence and accommodation responses are less strongly cross-linked than CRLC's and DND's. We speculate that reduced cross linkage allows more independent estimates of distance from vergence and accommodation and this in turn leads to more systematic effects of accommodative distance on disparity scaling.

Appendix C: The results of Buckley and Frisby

Analysis of Buckley and Frisby (1993) and Frisby et al. (1995)

Here we analyze results from two papers by Frisby et al. in the framework of the cue-combination model (Equations 1 and 2). They obtained the largest effects with vertical as opposed to horizontal ridges, so we restrict our analysis to the vertical ridges.

We start with the real-ridge experiment of Buckley and Frisby (1993); the data are presented in their Figure 9a. There were three depth cues: disparity, texture, and focus cues. In the framework of the cue-combination model, perceived depth is based on contributions from all available depth cues, each weighted according to its statistical reliability:

$$\hat{D} = w_d D_d + w_t D_t + w_f D_f, \quad (C1)$$

$$w_d + w_t + w_f = 1,$$

where the subscripts refer to the cue (d = disparity, t = texture, and f = focus), \hat{D} is the combined depth estimate, D_i are depth estimates from individual cues, and w_i are the weights. The actual shape of the ridge was always consistent with the disparity-specified shape, so

the depth specified by focus cues was equal to the depth specified by disparity: $D_f = D_d$. Thus, [Equation C1](#) becomes:

$$\hat{D} = (w_d + w_f)D_d + w_t D_t.$$

The texture cue D_t had a constant value k for each curve in their data figure (their Figure 9a), so

$$\hat{D} = (w_d + w_f)D_d + (1 - w_d - w_f)k. \quad (\text{C2})$$

Therefore, when the results are plotted as a function of disparity-specified depth (D_d), the slope corresponds to the sum of the weights given disparity and focus cues: $w_d + w_f$. The slope was ~ 0.95 , so we can conclude that the texture weight w_t was small in the real-ridge experiment. We obviously cannot determine the individual weights w_d and w_f .

Turning to the CRT-based experiment, focus cues always signaled a flat surface ($D_f = 0$), so

$$\begin{aligned} \hat{D} &= w_d D_d + w_t D_t, \\ \hat{D} &= w_d D_d + (1 - w_d - w_f)k. \end{aligned} \quad (\text{C3})$$

Thus, the slope of the data (their Figure 4a) now corresponds to w_d rather $w_d + w_f$. The observed slopes in the CRT data were smaller than in the real-ridge data, which means that $w_f > 0$.

Interestingly, slopes in the CRT data varied substantially with the texture-specified depth: when D_t was 9 cm, the slope was 0.74 (suggesting large w_d), and when D_t was 3 cm, the slope was ~ 0 ($w_d = \sim 0$). In the framework of the model, the weights changed as a function of the texture-specified depth. Buckley and Frisby (1993) pointed out that this makes sense: When the texture-specified depth was small, the cue signaled a depth close to that signaled by focus cues. The consistency between the two cues may have lead to down weighting of the inconsistent cue—disparity depth—as in robust estimation (Landy et al., 1995). When the texture depth was large, the cue was inconsistent with focus cues, so the down weighting of disparity did not occur.

In a second set of experiments, Frisby et al. (1995) examined how pinhole viewing affects perceived depth in CRT-based stimuli and real ridges. A pinhole creates a large depth of focus (Charman & Whitefoot, 1977; Green et al., 1980), which in turn reduces the variation in focus cues with distance. As a consequence, a pinhole minimizes the blur gradient and minimizes the stimulus to accommodation. We can analyze this experiment using the cue-combination model. When viewing real ridges through a pinhole, there are three depth cues that could in principle contribute to the percept: disparity, texture, and focus cues. The focus cues now specify a constant depth value because the blur gradient has been reduced to zero or

nearly zero. With real ridges, the estimated depth should therefore be:

$$\hat{D} = w_d D_d + w_t k + w_f 0,$$

where, as before, k is the constant depth specified by texture. Using the property that the weights add to 1,

$$\hat{D} = w_d D_d + (1 - w_d - w_f)k,$$

which is the same as [Equation C3](#). Thus, the cue-combination model predicts that real ridges, viewed through a pinhole, should yield data like CRT-based stimuli viewed normally. (A caveat: If the variances of the depth-cue estimators were calculated from on-going stimulus information, the variance of the focus-cue estimator would be larger with pinhole viewing, so the weights w_f would not be the same for CRTs with normal viewing and real stimuli with pinhole viewing.) This is what Frisby et al. observed. The real-ridge data with pinholes (Figure 3a, Frisby et al., 1995) were very similar to the CRT-based data with normal viewing (Figure 6a, Frisby et al., 1995; Figure 4a, Buckley & Frisby, 1993).

When viewing CRT-based stimuli through a pinhole, the estimated depth is given by [Equation C3](#) (because focus cues again indicate zero depth). Thus, the cue-combination model predicts that the perceived depth in CRT-based stimuli viewed through a pinhole should be similar to the perceived depth when viewed normally. Again this is what Frisby et al. (1995) observed. The CRT-based data with and without pinholes were similar (compare Figures 6a and 6c; Frisby et al., 1995).

Comparison of our results and those of Buckley and Frisby

In our first set of experiments, we looked for a direct effect of focus cues on perceived slant. We found a small but consistent effect with monocular viewing, but no effect with binocular viewing. It is useful to consider whether our findings are consistent with the results of Frisby et al. We did not observe an effect of focus cues with binocular viewing, so we cannot claim from our findings that inappropriate focus cues explain the differing percepts Frisby et al. reported for binocularly viewing virtual as opposed to real stimuli (e.g., Buckley & Frisby, 1993; Frisby et al., 1995). For example, Buckley and Frisby (1993) found that the perceived depth of real parabolic ridges was greater and more closely tied to the disparity-specified depth than the perceived depth of computer-displayed ridges. They concluded that the critical difference was focus cues, which specified the same curvature as the disparity signal for the real ridges and specified flatness for the computer-displayed ridges. Our data are inconsistent with this conclusion because we observed no direct effect of focus cues with binocular viewing.

Were there critical differences between their experiment and ours? There are at least four possibilities.

1. Buckley and Frisby's viewing distance was twice ours at 57 cm, but this should have caused similar increases in the standard deviations of the disparity and focus-cue estimators (Figure 10 and Appendix A), so the difference in viewing distance is not a likely cause of the discrepancy.
2. Their experiment involved judgments of depth for curved surfaces while ours involved judgments of slant for planar surfaces. But curvature is just the change in slant across a surface, so we see no reason why their task would have promoted a greater effect of focus cues than ours.
3. Perhaps focus cues were more informative with Buckley and Frisby's real surface than with our CRT stimuli because we had to use anti-aliasing, which added blur and may have reduced the informativeness of the blur gradient.
4. It seems, however, that the most important cause of the difference between our results and theirs is the means by which Buckley and Frisby manipulated focus cues (see also Frisby et al., 1995; van Ee et al., 1999). They used real surfaces, so any signals from the surface (e.g., focus cues, graininess of the surface itself, shading) would have been consistent with the disparity signal. We used computer-displayed stimuli with the monitor rotated so the disparity and focus signals were consistent, but other cues like shading were not. The results of a control experiment in Buckley and Frisby provide partial support for the idea that this is the key difference between their study and ours. Specifically, in Buckley and Frisby's experiments, variation of the actual depth would cause variation in the possible extraneous cues of shading and graininess, while variation of the texture-specified depth would not because that cue is based on the projected shapes of the texture elements at the retina. Table 1 in their paper shows that the average reported depth increased with an increase in actual depth (although the effect was not statistically significant). This finding suggests that unmodeled cues like shading and graininess may have contributed to the depth responses in Buckley and Frisby.

Acknowledgments

This research was supported by NIH Research Grant EY14194 (MSB). Thanks to Shrikant Bharadwaj and Allicia Beach for measurements of observers' accommodation stimulus-response functions, and AC/A and CA/C ratios. Thanks to Andrew Falth for assistance in produc-

ing the stimuli. Parts of the data were previously presented at the Vision Sciences Society Annual meetings in 2002 and 2003.

Commercial relationships: none.

Corresponding author: Simon J. Watt.

Email: s.watt@bangor.ac.uk.

Address: School of Psychology, Adeilad Brigantia, Penrallt Road, Bangor, Gwynedd LL57 2AS, UK.

References

- Akeley, K., Watt, S. J., Girshick, A. R., & Banks, M. S. (2004). A stereo display prototype with multiple focal distances. *ACM Transactions on Graphics*, 23, 1804–1813. [\[Article\]](#)
- Alais, D., & Burr, D. (2004). The ventriloquism effect results from near-optimal bimodal integration. *Current Biology*, 14, 257–262. [\[PubMed\]](#)
- Anstis, S. M., Howard, I. P., & Rogers, B. J. (1978). A Craik-Cornsweet illusion for visual depth. *Vision Research*, 18, 213–217. [\[PubMed\]](#)
- Backus, B. T., & Banks, M. S. (1999). Estimator reliability and distance scaling in stereoscopic slant perception. *Perception*, 28, 217–242. [\[PubMed\]](#)
- Backus, B. T., Banks, M. S., van Ee, R., & Crowell, J. A. (1999). Horizontal and vertical disparity, eye position, and stereoscopic slant perception. *Vision Research*, 39, 1143–1170. [\[PubMed\]](#)
- Baird, J. W. (1903). The influence of accommodation and convergence upon the perception of depth. *American Journal of Psychology*, 14, 150–200.
- Banks, M. S., Gepshtain, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? *Journal of Neuroscience*, 24, 2077–2089. [\[PubMed\]](#)
- Banks, M. S., Hooge, I. T. C., & Backus, B. T. (2001). Perceiving slant about a horizontal axis from stereopsis. *Journal of Vision*, 1(2), 55–79, <http://journalofvision.org/1/2/1/>, doi:10.1167/1.2.1. [\[PubMed\]](#) [\[Article\]](#)
- Biersdorf, W. R. (1966). Convergence and apparent distance as correlates of size judgments at near distances. *Journal of General Psychology*, 75, 249–264. [\[PubMed\]](#)
- Blake, A., Bülthoff, H. H., & Steinberg, D. (1993). Shape from texture: Ideal observers and human psychophysics. *Vision Research*, 33, 1723–1737. [\[PubMed\]](#)
- Bradshaw, M. F., Glennerster, A., & Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research*, 36, 1255–1264. [\[PubMed\]](#)

- Braunstein, M. L., Liter, C. J., & Tittle, J. S. (1993). Recovering three-dimensional shape from perspective translations and orthographic rotations. *Journal of Experimental Psychology. Human Perception and Performance*, 19, 598–614. [PubMed]
- Brookes, A., & Stevens, K. A. (1989). The analogy between stereo depth and brightness. *Perception*, 18, 601–614. [PubMed]
- Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research*, 33, 919–933. [PubMed]
- Burt, P., & Julesz, B. (1980). A disparity gradient limit for binocular fusion. *Science*, 208, 615–617. [PubMed]
- Campbell, F. W. (1957). The depth of field of the human eye. *Optica Acta*, 4, 157–164.
- Campbell, F. W., & Westheimer, G. (1959). Factors influencing accommodation responses of the human eye. *Journal of the Optical Society of America*, 49, 568–571. [PubMed]
- Caudek, C., & Proffitt, D. R. (1993). Depth perception in motion parallax and stereokinesis. *Journal of Experimental Psychology. Human Perception and Performance*, 19, 32–47. [PubMed]
- Charman, W. N., & Tucker, J. (1977). Dependence of the accommodation response on the spatial frequency spectrum of the observed object. *Vision Research*, 27, 129–139. [PubMed]
- Charman, W. N., & Whitefoot, H. (1977). Pupil diameter and the depth-of-field of the human eye as measured by laser speckle. *Optica Acta*, 24, 1211–1216.
- Collett, T. S., Schwarz, U., & Sobel, E. C. (1991). The interaction of oculomotor cues and stimulus size in stereoscopic depth constancy. *Perception*, 20, 733–754. [PubMed]
- de Berg, M., van Kreveld, M., Overmars, M., & Schwarzkopf, O. (2000). *Computational geometry: Algorithms and applications* (2nd Ed.). New York: Springer-Verlag.
- Domini, F., & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology. Human Perception and Performance*, 25, 426–444. [PubMed]
- Downing, E., Hesselink, L., Ralston, J., & Macfarlane, R. (1996). A three-color, solid-state, three-dimensional display. *Science*, 273, 1185–1189.
- Dixon, E. T. (1895). On the relation between accommodation and convergence to our sense of depth. *Mind*, 4, 195–212.
- Ellis, S. R., Smith, S. R., Grunwald, A. J., & McGreevy, M. W. (1991). Direction judgement error in computer generated displays and actual scenes. *Pictorial communication in virtual and real environments* (pp. 504–524). London: Taylor and Francis.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415, 429–433. [PubMed]
- Favalora, G. E., Napoli, J., Hall, D. M., Dorval, R. K., Giovinco, M. G., Richmond, M. J., et al. (2002). 100 million-voxel volumetric display. *Proceedings of the SPIE*, 712, 300–312.
- Fisher, S. K., & Ciuffreda, K. J. (1988). Accommodation and apparent distance. *Perception*, 17, 609–621. [PubMed]
- Fisher, S. K., & Ebenholtz, S. M. (1986). Does perceptual adaptation to telestereoscopically enhanced depth depend on the recalibration of binocular disparity? *Perception and Psychophysics*, 40, 101–109. [PubMed]
- Foley, J. M. (1977). Effect of distance information and range on two indices of visually perceived distance. *Perception*, 6, 449–460. [PubMed]
- Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, 87, 411–434. [PubMed]
- Frisby, J. P., Buckley, D., & Duke, P. A. (1996). Evidence for good recovery of lengths of real objects seen with natural stereo viewing. *Perception*, 25, 129–154. [PubMed]
- Frisby, J. P., Buckley, D., & Horsman, J. M. (1995). Integration of stereo, texture and outline cues during pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception*, 24, 181–198. [PubMed]
- Gårding, J., Porritt, J., Mayhew, J. E. W., & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Research*, 35, 703–722. [PubMed]
- Gepshtain, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, 13, 6, 483–488. [PubMed]
- Ghahramani, Z., Wolpert, D. M., & Jordon, M. I. (1997). Computational models of sensorimotor integration. In P. G. Morasso & V. Sanguineti (Eds.) *Self-organization, computational maps, and motor control* (pp. 117–147). Amsterdam: North-Holland.
- Gibson, E. J., Gibson, J. J., Smith, O. W., & Flock, A. (1959). Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology*, 54, 40–51. [PubMed]
- Gillam, B., Chambers, D., & Russo, T. (1988). Postfusional latency in slant perception and the primitives of stereopsis. *Journal of Experimental Psychology. Human Perception and Performance*, 14, 163–175. [PubMed]

- Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research*, 36, 3441–3456. [PubMed]
- Green, D. G., & Campbell, F. W. (1965). Effect of focus on the visual response to a sinusoidally modulated spatial stimulus. *Journal of the Optical Society of America*, 55, 1154–1157.
- Green, D. G., Powers, M. K., & Banks, M. S. (1980). Depth of focus, eye size, and visual acuity. *Vision Research*, 29, 827–835. [PubMed]
- Heath, G. G. (1956). The influence of visual acuity on accommodative responses of the eye. *American Journal of Optometry and Archives of American Academy of Optometry*, 33, 513–524. [PubMed]
- Heinemann, E. G., Tulving, E., & Nachmias, J. (1959). The effect of oculomotor adjustments on apparent size. *American Journal of Psychology*, 72, 32–45.
- Hillebrand, F. (1894). Das Verhältnis von Akkommodation und Konvergenz zur Tiefenlokalisierung. *Zeitschrift für Psychologie*, 7, 97–151.
- Hillis, J. M., & Banks, M. S. (2001). Are corresponding points fixed? *Vision Research*, 41, 2457–2473. [PubMed]
- 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(12), 967–992, <http://journalofvision.org/4/12/1/>, doi:10.1167/4.12.1. [PubMed] [Article]
- Hogervorst, M. A., & Eagle, R. A. (1998). Biases in three-dimensional structure-from-motion arise from noise in early the visual system. *Proceedings of the Royal Society of London. B*, 265, 1587–1593. [PubMed]
- Hogervorst, M. A., & Eagle, R. A. (2000). The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion. *Journal of Experimental Psychology. Human Perception and Performance*, 26, 934–955. [PubMed]
- Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth: Volume 2. Depth perception*. Toronto: I Porteous.
- Hunter, J. G., & Sackier, J. M. (1993). *Minimally invasive surgery*. New York: McGraw-Hill.
- Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, 39, 3621–3629. [PubMed]
- Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, 31, 1351–1360. [PubMed]
- Johnston, E. B., Cumming, B. G., & Parker, A. J. (1993). Integration of depth modules: Stereopsis and texture. *Vision Research*, 33, 813–826. [PubMed]
- Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research*, 34, 2259–2275. [PubMed]
- Judge, S. J., & Miles, F. A. (1985). Changes in the coupling between accommodation and vergence eye movements induced in human subjects by altering the effective interocular separation. *Perception*, 14, 617–629. [PubMed]
- Kilgard, M. J. (1996). *OpenGL Programming for the X Window System*. New York: Addison-Wesley.
- Knill, D. C. (1998). Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Research*, 38, 1683–1711. [PubMed]
- 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. [PubMed]
- Kording, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, 427, 244–247. [PubMed]
- Kotulak, J. C., & Schor, C. M. (1986). The dissociability of accommodation from vergence in the dark. *Investigative Ophthalmology and Visual Science*, 27, 544–551. [PubMed]
- Künnapas, T. (1968). Distance perception as a function of available visual cues. *Journal of Experimental Psychology*, 77, 523–529. [PubMed]
- Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America. A*, 18, 2307–2320. [PubMed]
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. J. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35, 389–412. [PubMed]
- Lipson, M. M. (2001). *Depth constancy: The role of vergence, accommodation and vertical disparity*. Unpublished doctoral thesis, Oxford University, U.K.
- Loomis, J. M., & Eby, D. W. (1988). Perceiving structure from motion: Failure of shape constancy. *Proceedings of IEEE Second International Conference on Computer Vision*, 383–391.
- Marshall, J. A., Burbeck, C. A., Ariely, D., Rolland, J. P., & Martin, K. E. (1996). Occlusion edge blur: A cue to relative visual depth. *Journal of the Optical Society of America. A*, 13, 681–688. [PubMed]
- Mather, G. (1996). Image blur as a pictorial depth cue. *Proceedings of the Royal Society of London. B*, 263, 169–171. [PubMed]
- Mather, G. (1997). The use of image blur as a depth cue. *Perception*, 26, 1147–1158. [PubMed]

- Mather, G., & Smith, D. R. (2000). Depth cue integration: Stereopsis and image blur. *Vision Research*, 40, 3501–3506. [PubMed]
- Mather, G., & Smith, D. R. (2002). Blur discrimination and its relation to blur-mediated depth perception. *Perception*, 31, 1211–1219. [PubMed]
- McQuaide, S. C., Seibel, E. J., Burstein, R., & Furness III, T. A. (2002). Three-dimensional virtual retinal display system using a deformable membrane mirror. *SID International Symposium Digest of Technical Papers*, 33, 1324–1327.
- Mitchison, G. (1993). The neural representation of stereoscopic depth contrast. *Perception*, 22, 1415–1426. [PubMed]
- Mon-Williams, M., & Tresilian, J. R. (2000). Ordinal depth information from accommodation? *Ergonomics*, 43, 391–404. [PubMed]
- Mon-Williams, M., Wann, J. P., & Rushton, S. (1993). Binocular vision in a virtual world: Visual deficits following the wearing of a head-mounted display. *Ophthalmic & Physiological Optics*, 13, 387–391. [PubMed]
- Nguyen, V. A., Howard, I. P., & Allison, R. S. (2005). Detection of the depth order of defocused images. *Vision Research*, 45, 1003–1011. [PubMed]
- North, W. J., & Woodling, C. H. (1970). Apollo crew procedures, simulation, and flight planning. *Astronautics & Aeronautics*, 8, 56–62. [Article]
- Ogle, K. N. (1950). *Researches in binocular vision*. Philadelphia: W. B. Saunders.
- O'Leary, A., & Wallach, H. (1980). Familiar size and linear perspective as distance cues in stereoscopic depth constancy. *Perception & Psychophysics*, 27, 131–135.
- Omura, K., Shiwa, S., & Kishino, F. (1996). 3-D display with accommodative compensation (3DDAC) employing real-time gaze detection. *SID 96 Digest*, 889–892.
- Ono, H., & Comerford, J. (1977). Stereoscopic depth constancy. In W. Epstein (Ed.), *Stability and constancy in visual perception* (pp. 91–128). Toronto: Wiley.
- Ono, H., Mitson, L., & Seabrook, K. (1971). Change in convergence and retinal disparities as an explanation for the wallpaper phenomenon. *Journal of Experimental Psychology*, 91, 1–10. [PubMed]
- Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, 43, 2451–2468. [PubMed]
- O'Shea, R. P., Govan, D. G., & Sekuler, R. (1997). Blur and contrast as pictorial depth cues. *Perception*, 26, 599–612. [PubMed]
- Pentland, A. P. (1987). A new sense for depth of field. *IEEE Transactions for Pattern Analysis and Machine Intelligence*, 9, 523–531.
- Perlin, K., Paxia, S., & Kollin, J. S. (2000). An autostereoscopic display. In *Proceedings of ACM SIGGRAPH 2000*, 319–326. [Article]
- Peter, R. (1915). Untersuchungen über die Beziehungen zwische primären und sekundären Faktoren der Tiefenwahrnehmung. *Archiv für die Gesamte Psychologie*, 34, 515–564.
- Rassweiler, J., Binder, J., & Frede, T. (2001). Robotic and telesurgery: Will they change our future? *Current Opinions in Urology*, 11, 309–320. [PubMed]
- Ritter, M. (1977). Effect of disparity and viewing distance on perceived depth. *Perception & Psychophysics*, 22, 400–407.
- Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature*, 361, 253–255. [PubMed]
- Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception*, 24, 155–179. [PubMed]
- Rogers, B. J., & Graham, M. E. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, 22, 261–270. [PubMed]
- Rogers, B. J., & Graham, M. E. (1983). Anisotropies in the perception of three-dimensional surfaces. *Science*, 221, 1409–1411. [PubMed]
- Rolland, J. P., Krueger, M. W., & Goon, A. A. (1999). Dynamic focusing in head-mounted displays. *SPIE*, 3639, 463–470.
- Sato, M., & Howard, I. P. (2001). Effects of disparity-perspective cue conflict on depth contrast. *Vision Research*, 41, 415–426. [PubMed]
- Schor, C. M., & Tsuetaki, T. K. (1987). Fatigue of accommodation and vergence modifies their mutual interactions. *Investigative Ophthalmology and Visual Science*, 28, 1250–1259. [PubMed]
- Segal, M., & Akeley, K. (2002). *The OpenGL Graphics System: A Specification (Version 1.4)*. OpenGL Architecture Review Board. [Article]
- Silverman, N. L., Schowengerdt, B. T., Kelly, J. P., & Seibel, E. J. (2003). Late-news paper: Engineering a retinal scanning laser display with integrated accommodative depth cues. *SID Symposium Digest* 34, 1538–1541. [Article]
- Smith, G., Jacobs, R. J., & Chan, C. D. (1989). Effect of defocus on visual acuity as measured by source and observer methods. *Optometry & Vision Science*, 66, 430–435. [PubMed]

- Stanberry, B. (2000). Telemedicine: Barriers and opportunities in the 21st century. *Journal of Internal Medicine*, 247, 615–628. [PubMed]
- Suyama, S., Date, M., & Takada, H. (2000). Three-dimensional display system with dual-frequency liquid-crystal varifocal lens. *Japanese Journal of Applied Physics*, 39, 480–484.
- Suyama, S., Takada, H., Uehira, K., & Sakai, S. (2000). A novel direct-vision 3-D display using luminance-modulated two 2-D images displayed at different depths. *SID 00 Digest*, 1208–1211.
- Suyama, S., Takada, H., Uehira, K., & Sakai, S. (2001). A new method for protruding apparent 3-D images in the DFD (Depth-Fused 3-D) display. *SID Symposium Digest* 32, 1300–1301. [Article]
- Swenson, H. A. (1932). The relative influence of accommodation and convergence in the judgment of distance. *Journal of General Psychology*, 7, 360–380.
- Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S. H., Loomis, J. M., & Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments. *Presence*, 13, 560–571.
- Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception and Psychophysics*, 48, 419–430. [PubMed]
- van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1998). The precision of proprioceptive position sense. *Experimental Brain Research*, 122, 367–377. [PubMed]
- van Beers, R. J., Wolpert, D. M., & Haggard, P. (2002). When feeling is more important than seeing in sensorimotor adaptation. *Current Biology*, 12, 834–837. [PubMed]
- van Damme, W., & Brenner, E. (1997). The distance used for scaling disparity is the same as the one used for scaling retinal size. *Vision Research*, 37, 757–764. [PubMed]
- van Ee, R., Banks, M. S., & Backus, B. T. (1999). Perceived visual direction near an occluder. *Vision Research*, 39, 4085–4097. [PubMed]
- van Ee, R., & Erkelens, C. J. (1996). Stability of binocular depth perception with moving head and eyes. *Vision Research*, 36, 3827–3842. [PubMed]
- von Holst, E. (1973). The participation of convergence and accommodation in perceived size constancy. In *The behavioural physiology of animals and man*. London: Butler and Tanner. English translation by R. Martin. (1969). *Zur Verhaltensphysiologie bei Tieren und Menschen: Gesammelte Abhandlungen*. Bd. I. Munich: R. Piper.
- Wallach, H., & Norris, C. M. (1963). Accommodation as a distance-cue. *American Journal of Psychology*, 76, 659–664. [PubMed]
- Wallach, H., & Zuckerman, C. (1963). The constancy of stereoscopic depth. *American Journal of Psychology*, 76, 404–412. [PubMed]
- Wann, J. P., & Mon-Williams, M. (1997). Health issues with virtual reality displays: What we know and what we don't. *Computer Graphics, May 1997*, 53–57.
- Wann, J. P., Rushton, S., & Mon-Williams, M. (1995). Natural problems for stereoscopic depth perception in virtual environments. *Vision Research*, 35, 2731–2736. [PubMed]
- Welch, G., Bishop, G., Vicci, L., Brumback, S., Keller, K., & Colucci, D. (1999). The HiBall tracker: High-performance wide-area tracking for virtual and augmented environments. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST 99)*, 1–10. [Article]
- Werner, H. (1937). Dynamics in binocular depth perception. *Psychological Monographs*, 49, 1–120.
- Westheimer, G. (1986). Spatial interaction in the domain of disparity signals in human stereoscopic vision. *Journal of Physiology*, 370, 619–629. [PubMed]
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception & Psychophysics*, 63, 1293–1313. [PubMed]
- Wickens, C. D., Merwin, D. H., & Lin, E. L. (1994). Implications of graphics enhancements for the visualization of scientific data: Dimensional integrality, stereopsis, motion, and mesh. *Human Factors*, 36, 44–61. [PubMed]
- Wöpkling, M. (1995). Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus. *Journal of the SID*, 3, 101–103.
- Wundt, W. (1862). *Beiträge zur Theorie der Sinneswahrnehmung*. Leipzig: Winter.
- Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae*. New York: Wiley.