

# The precision of binocular and monocular depth judgments in natural settings

Suzanne P. McKee

Smith-Kettlewell Eye Research Institute,  
San Francisco, CA, USA



Douglas G. Taylor

Smith-Kettlewell Eye Research Institute,  
San Francisco, CA, USA



We measured binocular and monocular depth thresholds for objects presented in a real environment. Observers judged the depth separating a pair of metal rods presented either in relative isolation, or surrounded by other objects, including a textured surface. In the isolated setting, binocular thresholds were greatly superior to the monocular thresholds by as much as a factor of 18. The presence of adjacent objects and textures improved the monocular thresholds somewhat, but the superiority of binocular viewing remained substantial (roughly a factor of 10). To determine whether motion parallax would improve monocular sensitivity for the textured setting, we asked observers to move their heads laterally, so that the viewing eye was displaced by 8–10 cm; this motion produced little improvement in the monocular thresholds. We also compared disparity thresholds measured with the real rods to thresholds measured with virtual images in a standard mirror stereoscope. Surprisingly, for the two naive observers, the stereoscope thresholds were far worse than the thresholds for the real rods—a finding that indicates that stereoscope measurements for unpracticed observers should be treated with caution. With practice, the stereoscope thresholds for one observer improved to almost the precision of the thresholds for the real rods.

Keywords: binocular vision, stereopsis, monocular depth cues

Citation: McKee, S. P., & Taylor, D. G. (2010). The precision of binocular and monocular depth judgments in natural settings. *Journal of Vision*, 10(10):5, 1–13, <http://www.journalofvision.org/content/10/10/5>, doi:10.1167/10.10.5.

## Introduction

Half a century ago, Gibson (1950) drew attention to the rich array of monocular depth information in the natural world. He felt that stereopsis, as a cue to depth, was overrated, noting that the apparent depth of a natural scene changes little when one closes one eye. The depth portrayed in two-dimensional media such as movies and computer graphics is compelling, providing further evidence of the strength of monocular depth cues. What exactly does stereopsis add to our perception of depth in the natural world?

In the last two decades, several studies have used real objects presented in natural surroundings to examine the human ability to judge three-dimensional shapes. Almost all have found that binocular shape estimates are more nearly veridical than monocular estimates (Allison, Gillam, & Vecellio, 2009; Buckley & Frisby, 1993; Durgin, Proffitt, Olson, & Reinke, 1995; Frisby, Buckley & Duke, 1996; Loomis, Philbeck, & Zahorik, 2002). Three-dimensional shape judgments require an estimate of the object's extent along the  $z$ -axis—the depth interval. It is likely that binocular judgments are better than monocular judgments because stereopsis provides less erroneous information about depth intervals.

As is well known, error comes in two varieties: systematic errors (bias, assessed by accuracy measurements, such as the PSE) and random errors (reliability, assessed by precision measurements, such as thresholds). Most of the studies comparing monocular and binocular judgments have focused on accuracy—on how close the shape judgment was to the actual physical shape—rather than on precision. In principle, humans should be able to compensate for systematic depth errors, particularly in performing well-practiced movements in familiar environments. A major league center fielder must be able to throw a ball accurately to second base from anywhere in the outfield, no matter what his perceived distance. Experimental evidence for these compensatory effects comes from a study by Loomis, Da Silva, Fujita, and Fukusima (1996). They asked observers to match a depth interval ( $z$ -axis) to a lateral extent ( $x$ -axis) and found that the depth intervals were generally underestimated. However, when they asked their observers to walk blindfolded across the same interval, their motor performance showed no evident bias.

Why does precision matter, if we already know that monocular depth estimates are inaccurate? It is difficult to correct for random depth errors, since these errors generally arise from inherent physiological noise, whereas we can and do correct for systematic errors. Precision is a

measure of uncertainty, which affects how rapidly we can carry out actions. If our estimate of the distance separating two objects is  $50 \text{ cm} \pm 1 \text{ cm}$ , we can move rapidly between them without damaging our bodies. However, if our estimate is  $50 \text{ cm} \pm 10 \text{ cm}$ , then we have to move slowly and update our information continuously. Precision (reliability) is also thought to determine the weights attached to various cues to depth. In contemporary Bayesian models of cue combination, cues from different modules (disparity, texture, motion parallax, etc.) are separately weighted according to their reliability (Landy, Maloney, Johnston, & Young, 1995), and then combined optimally. Many studies have shown that human observers combine cues in a way that is consistent with this optimal model (Ernst & Banks, 2002; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003; Svarverud, Gilson, & Glennerster, 2010).

Only a few studies have compared monocular and binocular precisions for judging depth in real settings. Frisby et al. (1996) measured Weber fractions for judging the length of real twigs presented in random orientations. In some of their experiments, the monocular and binocular Weber fractions were similar, while in other experiments the monocular Weber fractions were about twice the binocular ones. Because of the random orientations of the twigs, the judgments were based on  $x$ -,  $y$ -, and  $z$ -axis components of length, rather than on depth intervals *per se*. More recently, Allison et al. (2009) compared binocular and monocular judgments of the depth interval separating a metal rod from an adjacent panel. Binocular precision, estimated from the dispersion statistics, was as much as a factor of 40 better than monocular precision.

Although Allison et al. (2009) and Frisby et al. (1996) presented objects in real settings, the immediate surroundings of their test objects were fairly austere, e.g., a covered empty space, which minimized the monocular cues to depth. In particular, Allison et al. (2009) went to some trouble to minimize monocular cues; their observers viewed the test stimuli through an aperture that obscured the immediate surroundings, as well as the ends of the test rod and reference panel, and their lighting was uniform to obscure shadows. In a full cue, unrestricted setting, monocular cues might provide more precise depth information than this study suggests.

In the present study, we will measure depth interval thresholds for a pair of real rods, viewed monocularly or binocularly, in two different well-lighted indoor settings. First, we will present the rods in a relatively austere setting that contains many monocular cues (shape from shading, shadows, changes in lateral separation and angular subtense, etc.). Then, we shall enrich the setting with facsimiles of the usual clutter that surrounds objects in most indoor scenes, including an adjacent textured surface and occluded items—additions that incorporate potent monocular depth cues. Ultimately, we will also introduce motion parallax, so that the whole array of normal monocular cues is available for judging depth

intervals. Do all these cues, consistently presented, lead to monocular depth judgments comparable in precision to binocular judgments of depth?

To answer this question, we will make systematic measurements of depth interval thresholds as a function of the  $z$ -axis distance between the test rods. Our results show that, even in a highly enriched natural environment, binocular depth estimates are far more precise than monocular estimates over a substantial range of depth intervals.

## Methods

### Procedure and equipment

We compared binocular and monocular depth sensitivities for objects presented on a well-illuminated laboratory table. Observers judged the relative depth separating two metal rods—a depth interval judgment. One of the rods remained in a fixed position, while the other test rod was presented in one of four positions chosen at random from trial to trial. The base size of the depth interval was varied parametrically from 0 to 8 cm in separate blocks of trials. When the depth interval was zero, the observer judged whether the test rod was in front or behind the other fixed rod, i.e., a standard stereoacuity task. When the depth interval separating the two rods was non-zero, e.g., 4 cm, the observer judged the relative size of the incremental changes in the interval, e.g., whether the depth separating the rods was smaller or larger than 4 cm.

In Signal Detection terminology, our procedure was a “Yes–No” task. Observers categorized the test position with one of two labels: “front” or “back” for the zero pedestal condition; “large” or “small” for non-zero pedestals. They were given 5 practice trials at the beginning of each block to establish the test range and category boundary. Our previous work has shown that this number of practice trials is sufficient for observers to estimate the mean and range of a test set consisting of four test intervals (Morgan, Watamaniuk, & McKee, 2000). Feedback was given in the form of a beep if the observer judged an interval incorrectly. Observers were also given some practice with the task during preliminary blocks (50–100 trials) taken to establish the appropriate threshold range.

We plotted the percentage of trials that the observer labeled the position of the test rod “front” for the zero interval measurement or “large” for the non-zero intervals, and fitted a psychometric function to the data using probit analysis. We estimated thresholds from blocks of fifty trials; the threshold criterion was  $d' = 0.67$ . Each plotted data point is based on a minimum of 4 blocks of 50 trials each (200 trials total). The error bar on each point shows the standard error of the mean of the thresholds estimated from each of the 50 trial blocks.

We mounted the test rod on a micro-stage (National Aperture Model MM-4M-EX140). Its physical position could be varied in increments of less than a millimeter with a reliability of 0.001 mm, permitting measurements of fine stereoacuity. Approximately 6.5 cm of the test rod was visible above the black baseboard that concealed the micro-stage. The fixed reference rod was identical in width (0.4 cm) to the test rod and was mounted below the test rod so that it was 1.8 cm shorter than the test rod, when viewed from the observer's perspective (see Figure 2). The rods were separated laterally by 2.5 cm (1.3 deg). For most experiments, the observers viewed the two rods through computer-controlled shutter goggles that covered the eyes during the intertrial interval when the micro-stage was shifting position. For the monocular measurements, only the right eye's shutter was opened during a trial.

Black felt drapes, located 0.7 m behind the metal rods, were the background for all objects on the table. Overhead fluorescent lights provided strong continuous illumination of the room, adjacent laboratory furniture, and the test objects (see Figure 1). The lighting fixture immediately over the rods contained only one bulb unlike the other three fixtures, so it produced gentle shadows. The viewing distance for all measurements with the metal test rods was 1.12 m.

## Austere setting

We used two different environments for our measurements of depth sensitivity in natural settings. In the "Austere setting," only the rods, the black baseboard, and the black felt curtains were centrally visible through the shutter goggles (see Figure 2), although many of the

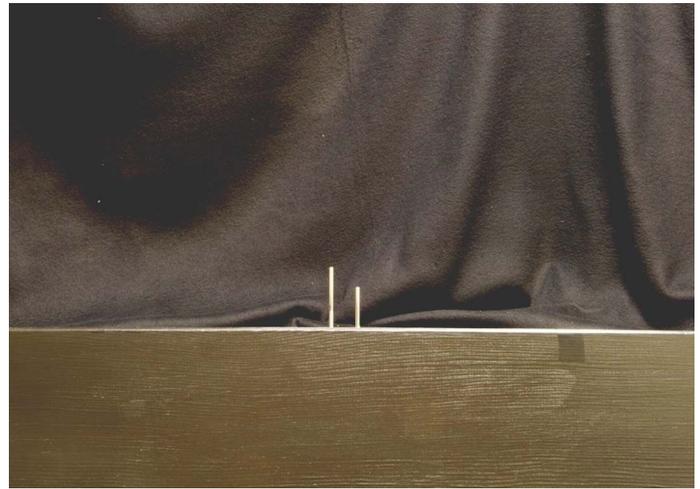


Figure 2. Austere setting for metal test rods from approximate viewpoint of observer. The variable test rod is on the left. Long duration photograph taken without flash to show lighting and shadows.

other objects in the room were visible in the periphery, since we made no attempt to conceal them. In the first experiment using this setting, the shutters were open for 1 s. In all subsequent experiments, the shutters were open until the observer made a judgment. We compared thresholds for a 1-s duration with thresholds for an unlimited duration and found that they were not significantly different.

## Enriched setting

For the "enriched setting," we added some additional objects surrounding the metal rods. Wrapping paper was attached to flat panels mounted adjacent to the test rods, and the panels were tilted slightly so that the observer could easily see the regular texture beside the rods. Grocery items were placed 0.4–0.6 m behind the rods to provide occlusion and relative depth cues (see Figures 1 and 3).

We asked observers to judge whether shadows and shape from shading were apparent when viewing the enriched setting from the position of the shutters. They all noted that the rods looked unevenly lit and rounded, with reflected coloring from adjacent surfaces (the black baseboard or the textured panels). They also noted highlights on the glass jars and faint shadows from the white rods onto the textured surfaces, as well as shadows on the other grocery objects. When looking through the shutters, they said they were able to see much of the rear of the room, including some furniture and cabinet doors, as well as the computer mouse on the table in front of them.



Figure 1. Experimental setting. The micro-stage lies between the two tilted panels in this picture of the enriched setting. Photograph taken *without flash* using an SLR camera mounted on a tripod, so that lighting and shadows would be visible.



Figure 3. Enriched setting from observer's viewpoint. Variable test rod on left. Long duration photograph taken without flash. The varying photographic durations alter colors somewhat from picture to picture.

In a separate control experiment on monocular motion parallax, we removed the goggles and mounted a large shutter in front of the rods themselves, which concealed the micro-stage movements during the intertrial interval. For this experiment only, the observer wore a patch over the left eye. Without the goggles in place, observers could move their heads over a larger distance, improving motion parallax information. We instructed each observer to move her head laterally from one side of the head holder to the other (18 cm); our measurements showed that this lateral head movement translated the center of the right eye 8–10 cm.

## Stereoscope measurements

We also measured disparity thresholds in a mirror stereoscope, composed of two pairs of mirrors arranged so that each eye could see only the image on one computer screen. Observers viewed the screens from a head holder that minimized head movements. Before any set of the threshold measurements was made, the observer adjusted mirrors closest to her eyes so that the Nonius lines were aligned with a minimum effort. The viewing distance was 1.22 m for the stereoscope measurements.

The stereoscopic display consisted of two dark lines that matched the dimensions (height, width, and separation in arcmin) of the two metal rods. The luminance of the lines was  $\sim 1$  cd/m<sup>2</sup> and that of the background was 30 cd/m<sup>2</sup>. Stimulus duration was unlimited for all measurements in the stereoscope; trials were terminated when the observer gave a response.

For the training sets in the stereoscope, we used a sparse random dot pattern composed of bright (50 cd/m<sup>2</sup>) points;

density was 189 points per square degree. The pattern consisted of a central circular test region (1.3 deg in diameter), surrounded by a square annulus, 2.3 deg on a side. Observers judged the depth interval that separated the central disk from the square annulus, using the same procedure described above for the metal rods. Observers were given 3 days of practice on the random dots; each data point shown in the training graph is based on 100–200 trials.

The displays were programmed on a Macintosh computer and presented on two Sony Trinitron monitors, Model 110GS. We presented the two half-images in the central 3 deg of the monitors where screen curvature was minimal. We ran the two monitors at 75 Hz, using a 1024 × 768 resolution level. Each pixel subtended 0.71 arcmin at the 1.22-m viewing distance. We used dithering to produce sub-pixel shifts in disparity. A Pritchard photometer was used to measure luminance as a function of monitor gray levels, and these values in turn were used in the dithering calculations.

## Observers

Both authors were observers for this study. The other four observers were naive volunteers who had not previously participated in any psychophysical study. All observers had normal or corrected-to-normal visual acuity.

Disparity thresholds for real objects, when expressed in arcmin, depend on the observer's interpupillary distance. In Figures 4 and 6, we have expressed thresholds in centimeters but have added second axes to show disparity thresholds in arcmin. We did not correct these second axes for interpupillary distances because the proliferation of scales would have been confusing. For these three figures, we assumed an interpupillary distance of 6 cm (the average interpupillary distance of our observers) to convert centimeters into arcmin. In Figures 8 and 9, we corrected disparity thresholds for the interpupillary distance of each observer. Interpupillary distances of our observers in centimeters were: S1 = 7.15; S2 = 5.75; S3 = 6; S4 = 5.8; S5 = 5.75; S6 = 6.

## Results

### Austere setting

The binocular and monocular thresholds for the austere setting are shown in Figure 4 for depth intervals ranging from 0 to 8 cm, or equivalently for 0–12 arcmin of disparity. The monocular thresholds are all far higher than the binocular thresholds. At the smallest intervals, the monocular thresholds are more than a log unit worse than

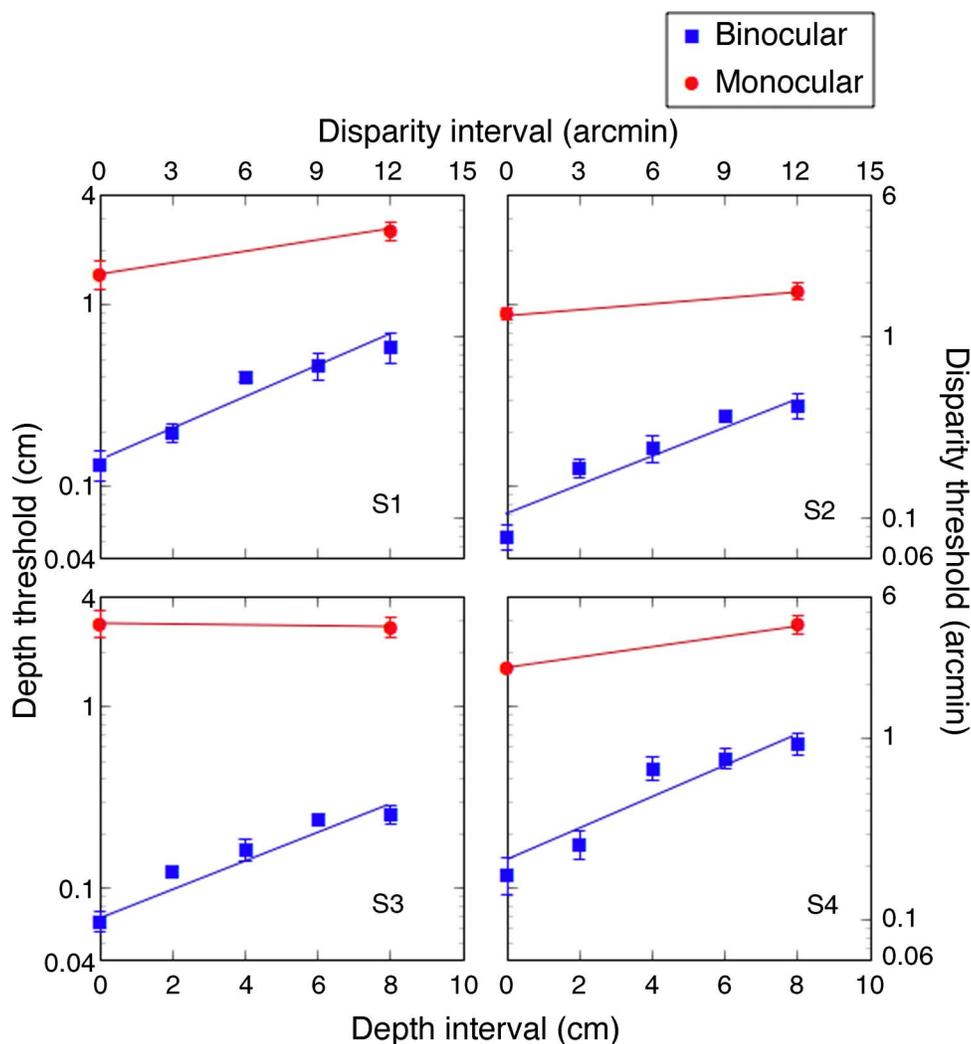


Figure 4. Binocular and monocular depth interval thresholds for four subjects for the metal rods shown in Figure 2. Left and lower axes in centimeters, right and upper axes in arcmin of disparity. Viewing distance 1.12 m; duration = 1 s. Straight lines were fitted by eye.

the binocular thresholds, confirming the results of Allison et al. (2009).

The binocular thresholds rise proportionately with the magnitude of the depth interval. The average Weber fraction for the four observers was 0.056, which is consistent with similar measurements, made using stereoscope displays presented for 1 s (McKee, Levi, & Bowne, 1990). Durgin et al. (1995) estimated Weber fractions for shape disparities greater than 3 arcmin as roughly 5–15%, again consistent with our measurements.

In contrast, the monocular thresholds show a very shallow dependence on depth interval measured over our test range of 0 to 8 cm. The average threshold for the zero interval, i.e., when the observers judged which rod was in front of the other, equaled 1.7 cm or 2.9 arcmin. In Figure 5, we have drawn a schematic of the two metal rods corresponding to this monocular threshold. Since  $z$ -axis information is not accessible monocularly, observers must rely on one of two sources of  $x$ -axis information: (1) the projected lateral separation between the rods as the depth

between them is changed (left side of figure), or (2) the difference in angular subtense of the variable rod with depth (right side of figure). Recall that the rods have the same physical width, so the difference in angular subtense is useful in this context, although it would not generally be useful for judging the depth of random objects. The observers could have used motion parallax, but the viewing aperture of the goggles was 2.5 cm and the stimulus duration was 1 s, making the motion cues weak. In a subsequent experiment, we examined the effect of motion parallax under more optimal circumstances.

A change of 1.7 cm (2.9 arcmin) along the  $z$ -axis produces a change in the projected  $x$ -axis separation of .02 deg (1.2 arcmin) or about 1.6% of the initial separation between the rods. The change of the angular subtense of the variable test rod, produced by a 1.7-cm depth increment, is roughly 0.2 arcmin or about 1.6% of the angular width of the rods. These estimated changes in lateral separation or angular subtense are close to the best thresholds for separation or width in the literature, which

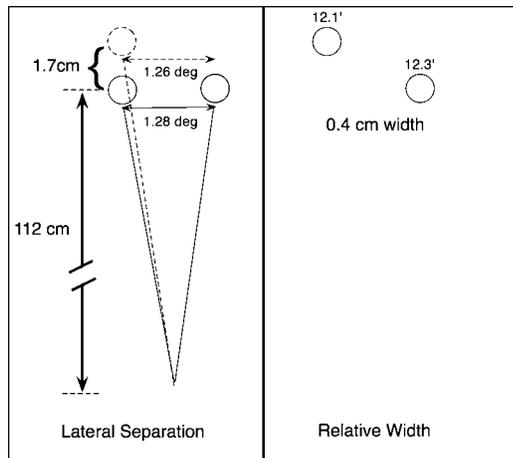


Figure 5. Schematic of the monocular information available at threshold for a depth interval of zero.

range from 1% to 3% (Burbeck, 1987; McKee, Welch, Taylor, & Bowne, 1990; Yap, Levi, & Klein 1987). The monocular depth threshold is limited by the same source of internal noise that limits estimates of lateral extent in other contexts. In short, the monocular threshold for zero interval is as good as it can be, based on the available spatial information.

Assume that the observer relies only on the projected  $x$ -axis separation between the rods (left side of Figure 5) and that the detectable change in that projected separation is always equal to the Weber fraction for width, namely  $\sim 1.5\%$ . From simple geometry, the  $z$ -axis change required to produce this change is also roughly 1.5% of the viewing distance to the rods. Since this distance does not change much at the largest depth interval (112-cm viewing distance + 8-cm interval = 120 cm), the monocular thresholds should be roughly constant, which they are. Actually, the thresholds for the larger depth interval (8 cm) are on average a little less precise than the zero interval thresholds, meaning that they should increase less than they do. Nevertheless, the Weber fraction ( $\sim 2.2\%$ ) for the larger interval still falls well within traditional estimates of sensitivity for lateral separation judgments.

## Enriched setting

The monocular information in natural scenes is generally far richer than that in the austere setting shown in Figure 2. In particular, objects often sit on textured surfaces and are surrounded by clutter. To simulate a more typical environment, we added patterned paper mounted on wooden surfaces positioned adjacent to the metal rods (see Figures 1 and 3); the papered surface was as close (0.75 deg) as possible while still permitting free movement of the micro-stage. Since the test rod was near the edge of the paper, we thought that the observers might be

able to read off the position of the test rod monocularly in “pattern” units. We also added grocery objects behind the two rods to provide familiar size and occlusion cues; as the test rod changed position in depth, different parts of the words on the package would be occluded. Observers were given unlimited time to make these judgments; trials were terminated when they pressed one of two mouse buttons.

We have plotted the binocular and monocular thresholds for the enriched setting in Figure 6. The monocular thresholds (red circles) and the binocular thresholds (open blue squares) may look remarkably similar to those in Figure 4, but, in fact, the difference between them has shrunk. This result will be particularly evident if you compare the two bottom graphs of Figure 4 with the two top graphs of Figure 6 (same subjects). Their monocular thresholds for the enriched setting are approximately half those for the austere setting. Thus, the enriched environment has improved the monocular thresholds substantially, but they are still about 10 times worse than the comparable binocular thresholds.

Given the additional reference targets in the enriched setting, we anticipated that the binocular thresholds would also improve. We have plotted binocular measurements for both experimental settings in Figure 6 (the open and closed blue squares). For three of the four observers, the enriched setting had no systematic effect on binocular thresholds. However, observer S6 clearly took advantage of the additional information. Her thresholds show almost no dependence on the depth interval between the metal rods, because the marks on the red patterned paper and the white plastic rods holding the adjacent surfaces are much better disparity references than the reference rod itself.

She has chosen an optimal strategy for estimating the depth interval from disparity. Her performance highlights an important fact about natural scenes—the rich array of information benefits both monocular and binocular judgments of depth.

## Motion parallax

The 2.5-cm aperture of the electronic shutter goggles made it difficult for observers to use monocular motion parallax to judge the size of the depth interval separating the metal rods. To remedy this situation, we removed the goggles and the chin rest from the head holder and mounted a large shutter immediately in front the rods.

This shutter concealed the movements of the variable test rod between trials; otherwise, it remained below the black baseboard during a block of trials. While viewing the rods in the enriched setting, observers were asked to move their heads laterally from one side of the head holder to the other repeatedly before making a judgment—a distance of 18 cm. This head movement translated the viewing eye over a distance of 8–10 cm, a distance somewhat larger than the interpupillary separation.

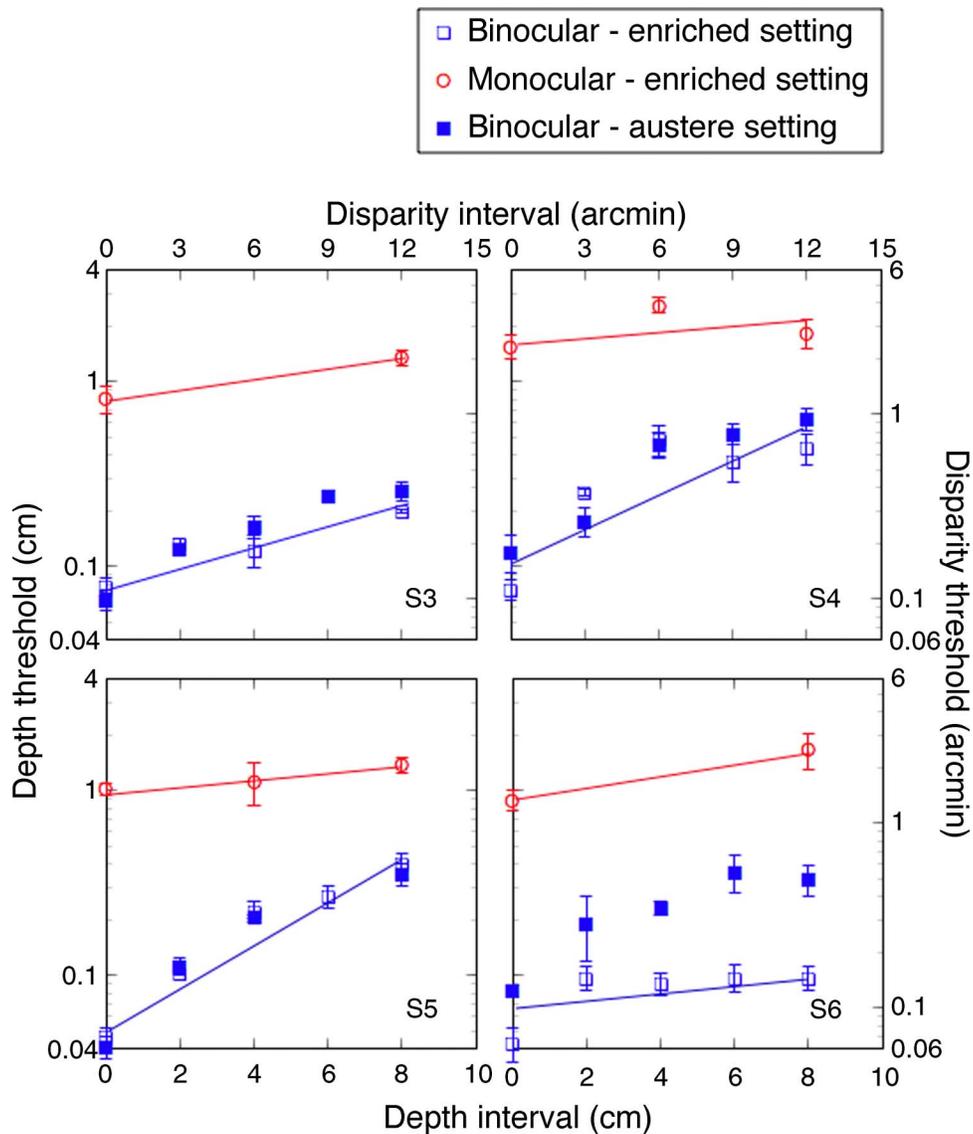


Figure 6. Monocular depth interval thresholds (open circles) for four subjects for the enriched setting (shown in Figure 3); binocular depth interval thresholds for enriched setting (open boxes) and austere setting (filled boxes). Left and lower axes in centimeters, right and upper axes in arcmin of disparity. Viewing distance = 1.12 m. Straight lines were fitted by eye.

As shown in Figure 7, motion parallax improved monocular thresholds significantly for only one (S5) of our three observers. Even for this observer, the monocular thresholds with motion parallax were significantly worse than the binocular thresholds. Several previous studies have found that monocular judgments of depth or shape based on motion parallax are not nearly as accurate or precise as binocular judgments, in agreement with our findings (Durgin et al., 1995; Frisby et al., 1996; LeClair & Durgin, 2008; Wheeler, 1982).

### Stereoscope measurements

The overwhelming message from the data taken in our natural settings is simple—binocular judgments are far

more precise than monocular judgments, no matter what cues are available. It appears that, if all monocular cues were removed, observers could base their judgments of depth interval, and implicitly three-dimensional shape, solely on disparity information. To put this idea to the test, we measured thresholds for computer-generated targets presented in a mirror stereoscope.

The test and reference “bars” in the stereoscope were two dark lines that matched the vertical and horizontal dimensions of the metal rods used in the austere setting. In Figure 8, we have plotted thresholds for the real and stereoscope displays. Surprisingly, three of the four observers were far less sensitive to disparity increments measured in the stereoscope than in the real setting.

Indeed, one observer could not see any depth difference between the dark lines, despite further experimental

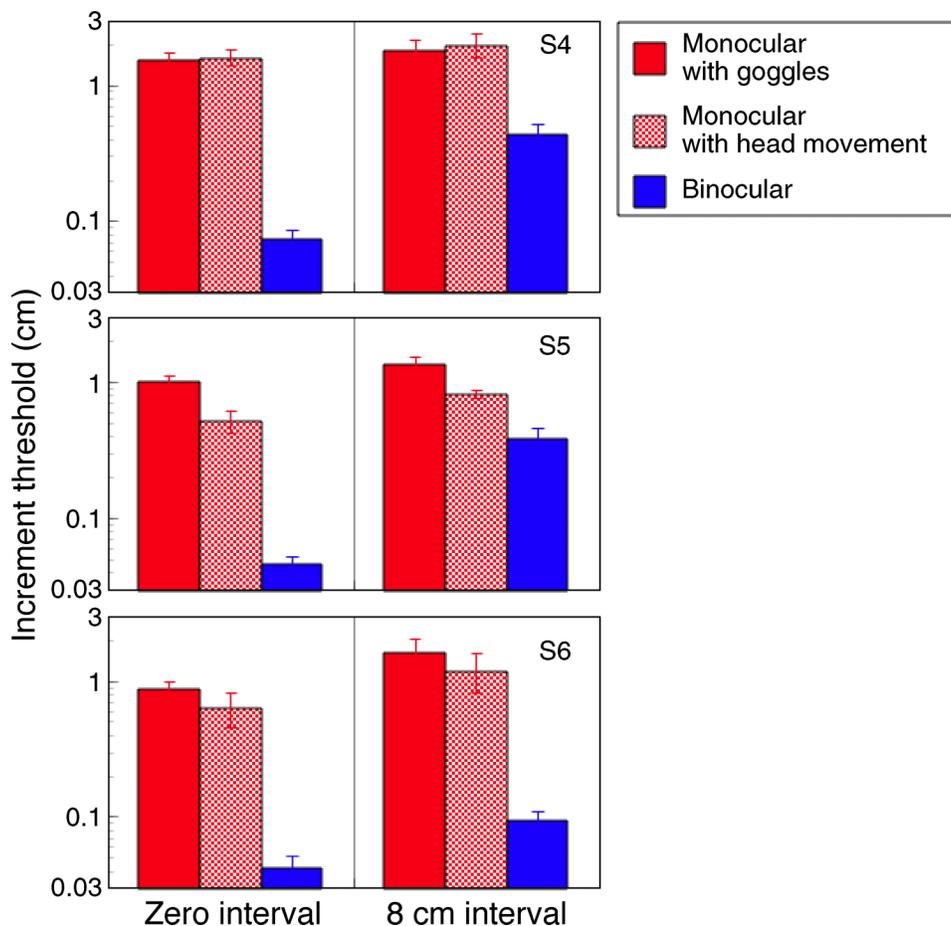


Figure 7. Monocular thresholds in centimeters for viewing enriched setting through a 2.5-cm aperture (red solid bars) or viewing enriched setting while moving head 18 cm laterally (hatched red bar); binocular thresholds (blue bars) for viewing enriched setting through 2.5-cm aperture. Viewing distance = 1.12 m; unlimited viewing time.

manipulations of step size, lateral separation, and the addition of perspective changes, i.e., changes in length and width consistent with bar depth. Only the senior author (S5) has identical thresholds for the two settings. She has had perhaps half a million trials for various stereoscopic displays, so her performance may simply reflect extensive practice. The second author (S3) has also had practice with stereoscopes but has always experienced difficulties with large depth intervals. For him, the interval sometimes flattens midway through a block of trials. The two naive observers had had no previous experience with stereoscope displays.

It is well known that stereoacuity improves with practice (Fendick & Westheimer, 1983; Gantz, Patel, Chung, & Harwerth, 2007; O'Toole & Kersten, 1992; Sowden, Davies, Rose, & Kaye, 1996). In fact, stereoacuity takes many more trials to reach asymptotic values than other spatial judgments (Kumar & Glaser, 1993). The puzzle here is that our naive observers needed almost no practice (at most 50 trials) for the stereoacuties measured with the real metal rods; their thresholds ranged from 6 to 14 arcsec in the austere setting, and 4–8 arcsec in the

enriched setting. The studies, cited above, showing that hundreds of trials were needed to produce fine stereoacuity were made with stereoscopic displays rather than with real objects. We suggest that in the studies showing significant improvement with practice observers were not learning to detect disparity *per se* but rather learning to detect disparity in a stereoscope.

What is it about a stereoscope that interferes with disparity sensitivity? Watt, Akeley, Ernst, and Banks (2005) found that the absence of correct focus cues affects perceived depth for surfaces presented in a stereoscope. The focus cues in our stereoscope would indicate that both the test and reference bars were in the same plane, in conflict with the disparity information. However, our viewing distance was 1.2 m and the maximum simulated depth interval was 8 cm, which should be well within the human depth of field of 0.33 diopters (Charman & Whitefoot, 1977).

There are other possible sources of cue conflict. Motion parallax from small head movements would indicate that both test and reference bars were in the same plane, contrary to the bar disparity. In addition, the two bars

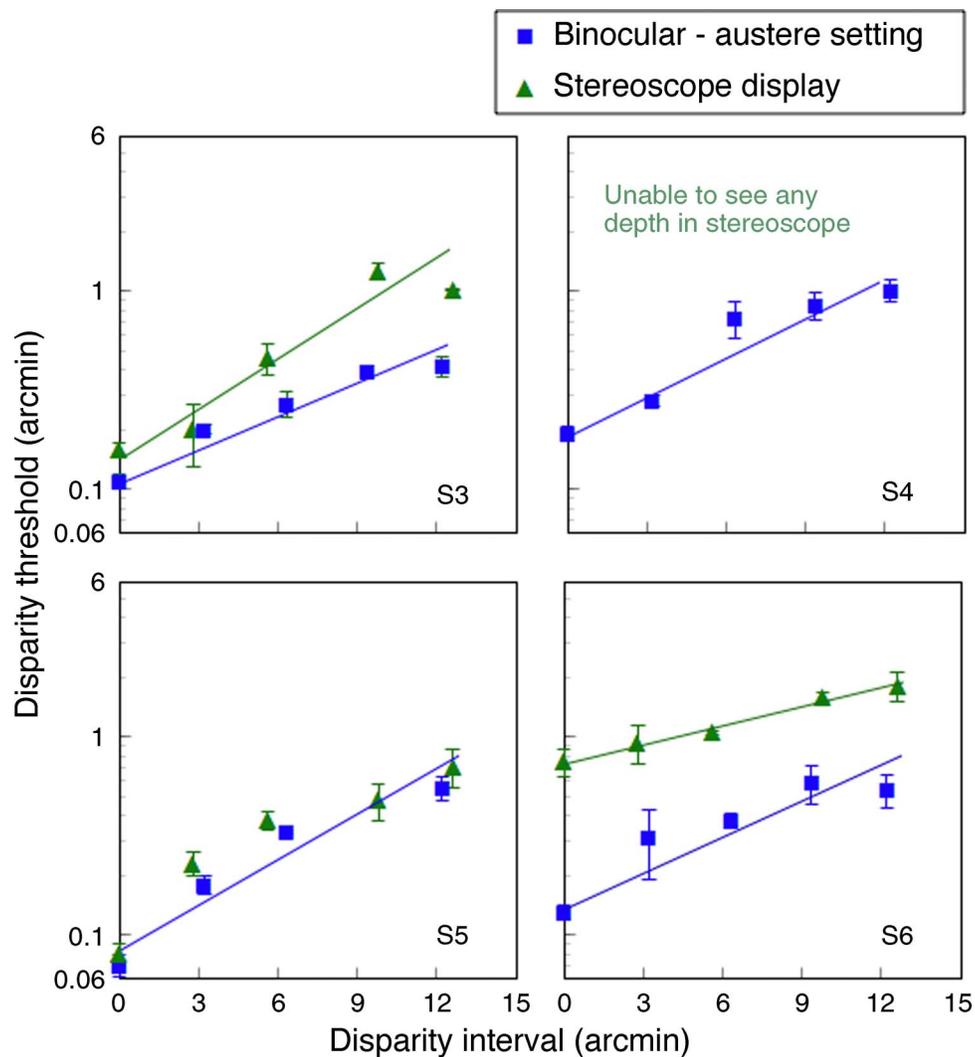


Figure 8. Disparity thresholds for depth interval judgments made with two real metal rods (blue squares) and for computer-generated bars presented in a mirror stereoscope (green triangles).

define an implicit slanted plane, and some observers have great difficulty detecting rotation about the vertical axis for simulated surfaces (Gulick & Lawson, 1976; Mitchison & McKee, 1990). Based on their calculations, Backus, Banks, Van Ee, and Crowell (1999) and Gårding, Porrill, Mayhew, and Frisby (1995) showed that eye position (vergence and version) could introduce horizontal disparities into frontoparallel surfaces. This ambiguity about what is causing the slant disparity, i.e., eye position or stimulus slant, may make it impossible to see depth between the simulated bars. Note, however, that the information supporting the percept of a slanted surface is very weak; the frontoparallel black bars were 12 arcmin wide and separated by a bright gray expanse of 1.3 deg. A single pair of lines or points is not sufficient to induce ambiguity for most observers (McKee, 1983; Fahle & Westheimer, 1988); stereoacuity thresholds for two lines are typically less than 10 arcsec.

Another issue is that real objects, like our metal rods, have “solidity”; their front and back surfaces have different disparities. The stereoscope “bars” have no solidity but appear as infinitely thin films, floating in space. Moreover, one of these ghostly bars is apparently floating in front of the monitor screens. Naive observers might find that the stereoscope bars are contrary to their expectations about objects in depth. Not all natural objects have solidity—specks of dust hovering in midair do not produce detectable differences in disparity between front and back surfaces.

We thought that if we modified the stereoscope display to minimize all these potential conflicts, our observers would be able to respond more easily to the test disparities. To simulate “dust clouds,” we used sparse random dot displays, reasoning that small bright points should produce less cue conflicts than the virtual bars, because random dots minimize texture and/or perspective

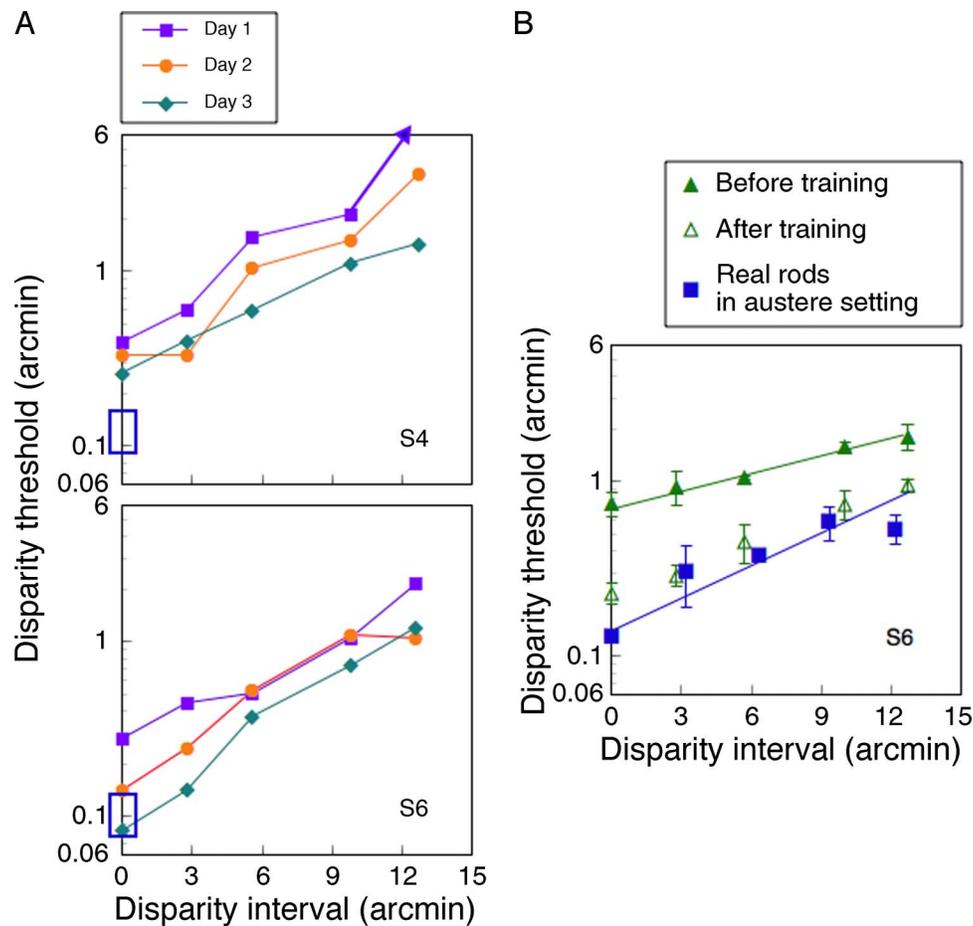


Figure 9. (A) Left graphs show three practice sessions for disparity increment thresholds measured with random dot displays presented in the stereoscope. The blue boxes on the ordinate show the stereoacuity thresholds for the metal rods in the austere setting. (B) The right graph shows transfer of training from the random dot displays to the “bar” targets presented in the stereoscope for one observer; the other observer was unable to judge depth between the bars even after training with random dot displays.

cues (Zabulis & Backus, 2004). The random dot display consisted of a circular central region that changed incrementally from trial to trial, surrounded by a fronto-parallel annular region that served as a reference surface. This arrangement made it impossible to interpret the display as a slanted surface. To teach our observers to detect fine disparities in the stereoscope, we trained them with these random dot displays.

### Learning to discriminate disparity in the stereoscope

Each naive observer was given 3 days of practice (1500–2000 trials) with the random dot displays. They made incremental judgments for depth intervals covering the same range as the bar targets. As shown by the graphs on the left of Figure 9, they were able to respond to the incremental changes in the central test region. Observer S4, who had previously been unable to see any depth difference

between the virtual bar targets, had less difficulty seeing depth differences in these sparse random dot displays, although her thresholds for large intervals were initially quite poor (purple squares in Figure 9A).

Both observers showed improved sensitivity; the blue boxes on the ordinate show their stereoacuity thresholds for the real metal rods in the austere setting. For observer S6, the stereoacuity threshold for the random dot display is, in fact, slightly better than her threshold for the metal rods. Stereoacuity for the other observer (S4) was still about a factor of two worse than her rod threshold.

Would training on the random dots transfer to the virtual bars? We repeated our increment threshold measurements with the bright bars in the stereoscope. Once again, observer S4 was unable to see any depth difference between the bars. Apparently, despite her capacity to respond to the disparities of the random dots, the various cue conflicts continued to interfere with her ability to respond to the disparity of the bars. On the other hand, observer S6 showed nearly perfect transfer of

training; her thresholds for the virtual bars after training are close to her thresholds for the real rods (see open triangles and filled squares in [Figure 9B](#)).

## Discussion

In natural settings, monocular information about depth is very imprecise. Our results show that, even for objects in rich local surroundings, monocular depth thresholds are as much as a log unit higher than binocular depth thresholds. We argue that, for static viewing, this imprecision follows from the viewing geometry; monocular information about relative depth along the  $z$ -axis depends on the *projected* distance separating features along the  $x$ -axis. To produce a detectable change in depth monocularly, the associated change in the  $x$ -axis projection has to reach threshold levels (see [Figure 5](#)). Thresholds for lateral separation are 1–2%. To produce a 1–2% increment in the  $x$ -axis projection, the change in the viewing distance to the test object has to be roughly 1–2%. Our monocular depth thresholds in the austere setting correspond to the 1–2% change in the  $z$ -axis distance needed to produce the 1–2% change in the  $x$ -axis projection.

In our enriched setting, the textured paper provided many marks that served as additional reference points. Thresholds in the enriched setting were therefore somewhat lower than thresholds for the austere setting. Would monocular thresholds be even lower if the rods were superimposed directly on the textured surface, or better yet, superimposed on a ruler with demarcations specifying numbered units? If an observer were estimating the position of a test rod positioned on a ruler lying on the  $z$ -axis, then determining where the rod fell on the ruler, e.g., where exactly the rod was sitting between the 5- and 6-cm marks, would still be imprecise for the same geometrical reasons described above. Of course, the optimum strategy for the monocular observer is simple. Walk to one side of the display, so that the  $z$ -axis is directly converted into an  $x$ -axis. Then, reading the position from a ruler is limited by the exquisite human sensitivity for lateral separation. In fact, in any natural setting, the optimum strategy for utilizing monocular cues is to view the depth relationships off the line of sight, so that the  $z$ -axis is converted into an  $x$ -axis judgment. This strategy obviously will not work if the objects are very far away; it also takes time. Fine stereopsis provides a rapid, precise assessment of depth differences along the line of sight without any need to change position.

In the [Introduction](#) section, we noted that thresholds are usually limited by internal sources of noise. From the poor thresholds, one might guess that all monocular processing is inherently noisier than binocular processing. This conclusion is incorrect. The Weber fraction for width is

about 1–2%, much better than the Weber fraction for disparity, which is 5–6%. If the monocular noise is so low, then why are the monocular thresholds so bad? Keep in mind that we are not measuring monocular thresholds for incremental changes along the  $x$ -axis, i.e., width. Instead, we are measuring the ability to discriminate changes along the  $z$ -axis from the information in the monocular image. Changes along the  $z$ -axis necessarily produce angular changes in  $x$ -axis dimensions—in the projected lateral separation between the rods or in the angular subtense of the rods—but as noted above, these changes are remarkably small. In short, monocular thresholds for real objects are largely limited by the lack of physical information, rather than by internal noise.

## Depth in the stereoscope

In a prescient comment, Buckley and Frisby (1993) warned against drawing conclusions about depth cue combination from computer-generated displays presented in a stereoscope (see also Frisby, Buckley, & Horsman, 1995). In the current study, we were interested only in disparity sensitivity for a pair of computer-generated lines, not perceived surface slant or cue combination. Nevertheless, three of the four observers showed higher thresholds for the stereoscope bars than for the real metal rods. This finding is only surprising in the context of our main results. For the real rods, the stereoacuity thresholds for all three observers were less than 10 arcsec, without significant practice. Apparently, the absence of monocular cues consistent with the disparity of the virtual bars interferes with fine stereoacuity. Yet our main results also show that the monocular cues provide unreliable information about depth. An ideal observer would ignore the monocular information in favor of the far more precise disparity information.

Real observers, however, are affected by cue conflict. Girshick and Banks (2009) found that disparity thresholds increased when there was a large conflict between the depth specified by texture and depth specified by disparity. This explanation works for our larger depth intervals, because the conflict between the standing disparity (12 arcmin) that defines the interval and the monocular cues (depth = zero) are in significant conflict. One would predict a pattern of an increasing discrepancy between thresholds for the real rods and those for the stereoscope as the standing disparity increases, thereby increasing the cue conflict between the disparity and the monocular cues. This pattern is actually observed for the second author (observer S3 in [Figure 8](#)). It is hard to see how this explanation works for stereoacuity. When the standing disparity is zero and the threshold increments are small, the conflict is trivial. Nevertheless, our two naive observers had poor stereoacuity for the virtual bars and, initially, for stereoacuity measured with the random dot display.

From our extensive experience measuring the properties of stereopsis in a stereoscope, we can assert that there are few observers like S4, who are unable to see any depth conveyed by virtual bar targets. Most naive observers do need practice in the stereoscope to produce fine stereoacuity thresholds. Our results here suggest that, for real objects, disparity judgments require no more practice than judgments about other dimensions.

## Practical consequences

Three to five percent of the population has no stereopsis because of strabismus during early development. The greatest concern of most pediatric ophthalmologists is the loss of visual acuity in the deviating eye (amblyopia). Our results show that the loss of stereopsis is also an important concern, even for those strabismics who do not suffer from amblyopia. It greatly increases their uncertainty about the location of features along the line of sight, and based on our calculations, it seems unlikely that monocular depth information can compensate for its absence. The optimum solution for these individuals is to move around the objects, converting the  $z$ -axis information into  $x$ -axis information. However, these movements take time and are only useful for objects that are fairly close. Of course, there are surgeons and professional athletes who manage superbly without stereopsis. It would be interesting to know what information these extraordinary individuals use to compensate for the loss of the disparity information. For less gifted individuals, the loss of stereopsis certainly hampers visual processing of object location and shape. Happily, extensive training and some types of treatment for amblyopia not only improve the acuity of the amblyopic eye but also lead to the recovery of stereopsis (Levi & Li, 2009; Li, Provost, & Levi, 2007).

## Acknowledgments

This research was supported by National Eye Institute Grants R01-EY018875 and R01-EY06644 and by The Smith-Kettlewell Eye Research Institute. We thank Laurie Wilcox, James Elder, Preeti Verghese, Andrew Glennerster, Justin Ales, and Christopher McKee for valuable discussions about these results.

Commercial relationships: none.

Corresponding author: Suzanne P. McKee.

Email: [suzanne@ski.org](mailto:suzanne@ski.org).

Address: 2318 Fillmore St., San Francisco, CA 94115, USA.

## References

- Allison, R. S., Gillam, B. J., & Vecellio, E. (2009). Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision*, *9*(1):10, 1–14. <http://www.journalofvision.org/content/9/1/10>, doi:10.1167/9.1.10. [PubMed] [Article]
- 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]
- 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]
- Burbeck, C. A. (1987). Position and spatial frequency in large-scale localization judgments. *Vision Research*, *27*, 417–428. [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.
- Durgin, F. H., Profitt, D. R., Olson, T. J., & Reinke, K. S. (1995). Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 679–699. [PubMed]
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433. [PubMed]
- Fahle, M., & Westheimer, G. (1988). Local and global factors in disparity detection of rows of points. *Vision Research*, *28*, 171–178. [PubMed]
- Fendick, M., & Westheimer, G. (1983). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision Research*, *23*, 145–150. [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]
- Gantz, L., Patel, S. S., Chung, S. T. L., & Harwerth, R. S. (2007). Mechanisms of perceptual learning of depth discrimination in random dot stereograms. *Vision Research*, *47*, 2170–2178. [PubMed]

- Gårding, J., Porrill, J., Mayhew, J. E., & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Research*, *35*, 703–722. [PubMed]
- Gibson, J. J. (1950). *The perception of the visual world* (p. 108). Cambridge, MA: The Riverside Press.
- Girshick, A. R., & Banks, M. S. (2009). Probabilistic combination of slant information: Weighted averaging and robustness as optimal percepts. *Journal of Vision*, *9*(9):8, 1–20, <http://www.journalofvision.org/content/9/9/8>, doi:10.1167/9.9.8. [PubMed] [Article]
- Gulick, W. L., & Lawson, R. B. (1976). *Human stereopsis*. New York: Oxford University Press.
- 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):1, 967–992, <http://www.journalofvision.org/content/4/12/1>, doi:10.1167/4.12.1. [PubMed] [Article]
- 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]
- Kumar, T., & Glaser, D. A. (1993). Initial performance, learning and observer variability for hyperacuity tasks. *Vision Research*, *33*, 2287–2300. [PubMed]
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*, 389–412. [PubMed]
- LeClair, A., & Durgin, F. H. (2008). Depth interval perception: Comparing binocular stereopsis with motion parallax in “action space” [Abstract]. *Journal of Vision*, *8*(6):857, 857a, <http://www.journalofvision.org/content/8/6/857>, doi: 10.1167/8.6.857.
- Levi, D. M., & Li, R. W. (2009). Perceptual learning as a potential treatment for amblyopia: A mini-review. *Vision Research*, *49*, 2535–2549. [PubMed]
- Li, R. W., Provost, A., & Levi, D. M. (2007). Extended perceptual learning results in substantial recovery of positional acuity and visual acuity in juvenile amblyopia. *Investigative Ophthalmology and Visual Science*, *48*, 5046–5051. [PubMed]
- Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1996). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 906–921. [PubMed]
- Loomis, J. M., Philbeck, J. W., & Zahorik, P. (2002). Dissociation between location and shape in visual space. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 1202–1212. [PubMed]
- McKee, S. P. (1983). The spatial requirements for fine stereoacuity. *Vision Research*, *23*, 191–198. [PubMed]
- McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research*, *30*, 1763–1779. [PubMed]
- McKee, S. P., Welch, L., Taylor, D. G., & Bowne, S. F. (1990). Finding the common bond: Stereoacuity and the other hyperacuties. *Vision Research*, *30*, 879–891. [PubMed]
- Mitchison, G. J., & McKee, S. P. (1990). Mechanisms underlying the anisotropy of stereoscopic tilt perception. *Vision Research*, *30*, 879–891. [PubMed]
- Morgan, M. J., Watamaniuk, S. N., & McKee, S. P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research*, *40*, 2341–2349. [PubMed]
- O’Toole, A., & Kersten, D. (1992). Learning to see random dot stereograms. *Perception*, *21*, 227–243. [PubMed]
- Sowden, P., Davies, I., Rose, D., & Kaye, M. (1996). Perceptual learning of stereoacuity. *Perception*, *25*, 1043–1052. [PubMed]
- Svarverud, E., Gilson, S. J., & Glennerster A. (2010). Cue combination for 3D location judgements. *Journal of Vision*, *10*(10):7, 1–13, <http://www.journalofvision.org/content/10/1/5>, doi:10.1167/10.1.5. [PubMed] [Article]
- Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, *5*(10):7, 834–862, <http://www.journalofvision.org/content/5/10/7>, doi:10.1167/5.10.7. [PubMed] [Article]
- Wheeler, D. G. (1982). The role of motion parallax in the perception of distance, depth and size. (Dissertation, Rutgers-Newark, The State University of New Jersey) [URL]
- Yap, Y. L., Levi, D. M., & Klein, S. A. (1987). Peripheral hyperacuity: 3-Dot bisection scales to a single factor from 0 to 10 deg. *Journal of the Optical Society of America A*, *4*, 1554–1561. [PubMed]
- Zabulis, X., & Backus, B. T. (2004). Starry nights: A texture devoid of depth cues. *Journal of the Optical Society A*, *21*, 2049–2060. [PubMed]