Representation and measurement of stereoscopic volumes

Ross Goutcher
Faculty of Natural Sciences (Psychology), University of Stirling, Stirling, UK

L. M. Wilcox
Department of Psychology, Centre for Vision Research, York University, Toronto, ON, Canada

Binocular disparity information provides the human visual system with a basis for the compelling perception of both three-dimensional (3-D) object shape, and of the 3-D space between objects. However, while an extensive body of research exists into the perception of disparity-defined surface shape, relatively little research has been conducted on the associated perception of disparity-defined volume. In this paper, we report three experiments that examine this aspect of binocular vision. Participants were asked to make judgments about the 3-D spread, location-in-depth, and 3-D shape of stereoscopic volumes. Volumes were comprised of random dots with disparities drawn from a uniform distribution, a Gaussian distribution, or a combination of both. These results were compared to two models: One of these made judgments about stereoscopic volumes using information about the distributions of disparities in each stimulus, while the other was limited to only maximum and minimum disparity information. Psychophysical results were best accounted for by the maximum-minimum decision rule model. This suggests that, although binocular vision affords a compelling phenomenal sense of 3-D volume, when required to make judgments about such volumes, the visual system's default strategies make only limited use of available binocular disparity signals.

Introduction

Stereoscopic vision, the compelling sense of three-dimensional (3-D) structure derived from disparate binocular images, has been the subject of extensive study since its discovery by Wheatstone (1838). These have included investigations of the upper (Badcock & Schor, 1985; Foley, Applebaum, & Richards, 1975; Ogle, 1952, 1953; Westheimer & Tanzman, 1956; Wilcox & Hess, 1995) and lower (McKee, 1983; Tyler, 1973) limits of stereopsis, and its role in defining surface shape and slant (e.g., Gillam, 1968; Johnston, 1991; Lunn & Morgan, 1997; Rogers & Cagenello, 1989; Stevens & Brooks, 1988). The introduction of the random dot stereogram (RDS) by Julesz (1964, 1971) provided a useful tool for the study of stereoscopic processing without additional cues or conflicts from monocularly visible form.

Using RDS patterns researchers have shown that, using stereopsis alone, observers can perceive multiple pseudotransparent planes in depth (Akerstrom & Todd, 1988; McKee & Verghese, 2002; Tsirlin, Allison, & Wilcox, 2008, 2012; Wallace & Mamassian, 2004; Weinshall, 1989, 1991, 1993). However, if there is insufficient separation of two planes in depth, they are perceived as a disparity-defined volume, a phenomenon that Parker and Yang (1989) referred to as *disparity averaging*. This sense of disparity-defined volume was one of the key aspects of stereopsis noted by Susan Barry in her autobiographical account of the apparent restoration of this component of her vision. Barry noted the "palpable volume of empty space" that could only be perceived stereoscopically, and stated that she "could see, not just infer, the volume of space" (Barry, 2009, pp. 94–95). Of interest here is the fact that, in the experiments of Parker and Yang (1989) and others (Stevenson, Cormack, & Schor, 1989; Tyler, 1991), below a certain disparity threshold, observers are unable to distinguish between distinct surfaces and a volume of dots in depth. This demonstrates that the perception of surface structure is not a prerequisite for the perception of disparity-defined space. However, at present, there has been little empirical attention paid to how volumetric depth percepts are obtained from stimuli that do not contain surface structure.

One notable recent exception is the work of Harris (2014). This study examined judgments of depth extent in stimuli depicting either two surface stereo transparency, or a volume of points in depth. Her results showed that the perceived depth of stereoscopic volumes was significantly smaller than for a pair of transparent planes. These results were consistent with

Citation: Goutcher, R., & Wilcox, L. M. (2016). Representation and measurement of stereoscopic volumes. *Journal of Vision*, *16*(11):16, 1–17, doi:10.1167/16.11.16.

doi: 10.1167/16.11.16 Received December 10, 2015; published September 20, 2016 ISSN 1534-7362

the performance of a cross-correlation model of disparity measurement. Stereoscopic volumes thus appear to be subject to the same processes of disparity averaging (Parker & Yang, 1989) seen in stereo transparency, albeit to a different degree. In a similar vein, van Ee and Anderson (2001) found that the perceived depth-to-width ratio of disparity-defined volumes was affected by the distribution of element orientations, as well as by their speed and direction of motion. These results suggest that factors that reduce correspondence-matching ambiguity lead to an increase in perceived depth.

Thanks to the work of Harris (2014) and van Ee and Anderson (2001), we have some understanding of the quantitative perception of stereoscopic volume, compared to segmented surfaces in depth. Three-dimensional patterns defined by elements distributed randomly in depth are not simply perceived as isolated points in space. As noted above, they cohere to create the percept of a volume. It seems reasonable to assume that the form of a given volume will depend on the distribution of the points within it. However, to date we do not know if or how the visual system represents such 3-D volumes. Here, we examine the visual system's ability to compare 3-D volumes and make use of information about the distribution of disparities within them. In three experiments, we assess the visual system's capacity to discriminate disparity-defined volumes, and examine the effects of varying the distribution of disparities within volume stimuli. Our results suggest that the visual system makes only limited use of available binocular disparity information within 3-D volumes so the perceived form of the volume is largely determined by its most extreme disparity values.

Experiment 1: 3-D spread discrimination

In Experiment 1, we examined observers' ability to discriminate differences in the spread-in-depth of RDS volumes, with uniform or Gaussian distributions, providing an initial marker for our capacity to make judgments about these stimuli. We chose these distributions as they have previously been used to study the representation of two-dimensional (2-D) distributions (Hirsch & Mjolsness, 1992; Juni, Singh, & Maloney, 2010; Morgan & Glennerster, 1991). In addition, the mathematical definition of spread is quite different for these two distributions. Estimates of variance V depend upon the entire sample of points for Gaussian distributions and on only maximum and minimum values for uniform distributions (see Equations 1 and 2).

$$V_{\text{Gaussian}} = \frac{\sum (d - \bar{d})^2}{N - 1} \quad (1)$$

where d is the disparity of a given dot, $\bar{d}$ is the mean disparity, and N is the total number of samples.

$$V_{\text{uniform}} = \frac{1}{12}(b - a)^2 \quad (2)$$

where b and a are, respectively, the maximum and minimum displayed disparities. Results were compared to model observers, which applied different decision rules to the distribution of dot disparities in each volume. These decision rules reflect how variability is calculated for each volume, through the standard deviation $\sqrt{V_{\text{Gaussian}}}$ for Gaussian distributions and the difference between maximum and minimum values $b - a$ for uniform distributions.

Methods

Participants

Five participants completed Experiment 1. All were staff or students of York University and had normal or corrected-to-normal vision. Stereoacuity was measured using a RanDot stereotest, and was less than 40 s of arc for all participants. Participants were naive as to the purposes of the experiment and provided written consent for their participation. All experimental procedures were approved by the University ethics board, and meet the requirements of the British Psychological Society guidelines, and the Declaration of Helsinki.

Stimulus and apparatus

The experiment was conducted using a G5 Power Macintosh computer (Apple, Inc., Cupertino, CA) with two ViewSonic G225f CRT displays (ViewSonic, Los Angeles, CA). Monitor resolution was 1280 × 960 pixels, with a refresh rate of 100 Hz, resulting in a single pixel subtending an angle of 1.77 arcmin at the 60-cm viewing distance. Stereoscopic presentation was achieved using a mirror stereoscope, correctly calibrated for the viewing distance to ensure there was no conflict between vergence angle and the accommodative state of the eye. Participants used a chin rest during testing to stabilize head position. All stimuli were created and presented using Matlab (Mathworks, Inc., Natick, MA), together with the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997).

Stimuli for this experiment were random-dot volumes in depth, with binocular disparities drawn from one of two distributions, either uniform or Gaussian (further details in Design and procedure section,

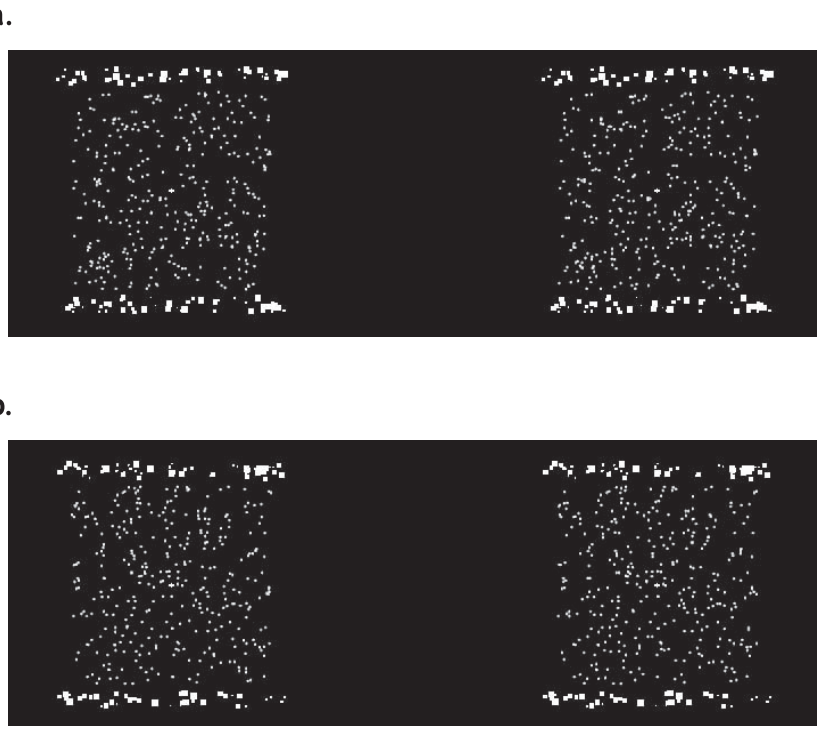

Figure 1. Examples of the disparity volume stimuli used in Experiment 1. (a) Stereo pair for a Gaussian-distributed random-dot volume. (b) Stereo pair for a uniformly distributed random-dot volume.

below). While distribution spread was varied parametrically, location-in-depth was chosen at random from a uniform distribution of range ±3.54 arcmin. This randomization ensured that participants could not simply use extreme disparity values to discriminate differences in spread. For each volume, dot x- and y-coordinates were drawn at random from a uniform distribution, covering an area of $7.67° \times 7.67°$. Each dot was 8.85 arcmin in diameter, with an overall dot density of 5.9 dots per degree^2. RDS elements were high contrast white circles, presented against a black background. A small white fixation cross was presented at the center of the display, together with zero disparity reference planes, comprised of randomly placed white squares, placed above and below the stimulus (see Figure 1). These were also present during stimulus presentation.

Design and procedure

Participants were presented with pairs of uniformly or Gaussian-distributed disparity volumes using a method of constant stimuli, two-interval forced-choice procedure. Disparity volumes were at one of four standard 3-D spreads of 3.54, 7.08, 10.62, and 14.16 arcmin and were presented for 600 ms. The 3-D spread values define the maximum disparity magnitude for uniformly distributed volumes and the standard deviation of Gaussian-distributed volumes. Standard volumes were presented at random in one of the two intervals, with the other interval containing a volume of dots drawn from the same distribution shape, at one of nine different spread ratios. Spread ratios were defined on a $\log_2$ scale, relative to the standard, with positive values indicating a larger spread than the standard, and negative values indicating a smaller spread than the standard. Ratio values varied between ±2.5, equivalent to spreads of between 0.177 and 5.66 times the standard. Participants were asked to judge which of the two intervals contained the greater 3-D spread. Responses were made using left and right buttons on a Logitech gamepad. Each participant completed a total of 50 repeated trials of each stimulus, over a total of 10 blocks. Presentation of standard spreads was randomized within blocks, as was interval order and spread ratio. While only a single distribution shape was measured in each block, block orders were randomized.

Task performance was measured by fitting an inverted cumulative Gaussian distribution to the proportion of "standard greater" responses P for each standard spread, and distribution shape. The fitted function could also be scaled and shifted, to account for asymmetries in participants' responses at disparity ratios greater than and less than the standard (see below), resulting in a four-parameter fit, as described in Equation 3.

$$P = \frac{1 - \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]}{\alpha} + \beta \quad (3)$$

where μ is the function mean, σ is the slope, and α and β are, respectively, scaling and shifting factors. Fitted σ values were used to compare precision of estimates across conditions.

Results and discussion

Psychophysical performance

Figure 2 shows example fits for a single participant (Figure 2a, b) for both Gaussian and uniform distributions, together with fitted σ values for each participant (Figure 2c, d), across all conditions. From these fits it is notable that, for disparity ratios smaller than the standard, there is little consistent effect of either standard spread, or distribution shape, on performance. The 3-D spread discrimination is a relatively constant function of the ratio of variable-to-standard spread, with 75% threshold performance requiring a variable spread around -0.56 $\log_2$ units or -0.43 $\log_2$ units from the standard spread for Gaussian and uniform distributions, respectively. Statistical analysis using a two-way repeated measures analysis of variance (ANOVA) shows no significant effect of distribution type, $F(1, 4) = 1.568$, $p = 0.279$, or standard spread, $F(3, 12) = 0.536$, $p = 0.667$, on fitted σ values, with no significant interaction, $F(3, 12) = 2.623$, $p = 0.099$.

For disparity ratios greater than the standard ($\log_2$ ratios > 0 in Figure 2a, b), there is a marked tendency for responses to rebound at larger standard spreads: When variable spread is larger than the standard, participants either maintain near chance performance, or show an increasing tendency to report the standard as having a greater spread in depth. Note that this rebound did not occur for all observers, and tended to occur only at larger standard spreads. This suggests a role for individual differences in the range of perceptible disparities, with the 14.16 arcmin standard very much at the upper end of this range. Rebounding at larger spreads is thus consistent with a reduction or disruption in perceived depth for distributions with large 3-D spreads, either due to false matching or to reaching the upper limit for disparity at that point (i.e., dots with large disparities appear unmatched and at indeterminate depth). Reductions in perceived disparity may also represent the action of residual flatness cues (van Ee, Adams, & Mamassian, 2003), due to the increased occurrence of false matches at such large disparities.

Model observer analysis

To better understand performance in this task, we examined the information use consistent with participants' responses. To do so, we created two model observers with the assumptions that (a) the visual system is able to solve the correspondence problem for 3-D volume stimuli, and (b) recover disparity measurements with reasonable accuracy, such as may be achieved using cross-correlation–based disparity estimation processes (see, for example, Allenmark & Read, 2011; Banks, Gepshtein, & Landy, 2004; Filippini & Banks, 2009; Goutcher & Hibbard, 2014; Harris, 2014). Given these assumptions, each model had knowledge of the disparity of each stimulus dot, subject to limitations in sampling efficiency (the randomly selected subset of dot disparities informing the decision stage) and internal noise. This approach follows earlier modeling strategies adopted by Harris and Parker (1992, 1994) for disparity discrimination tasks. Internal noise was defined as a set of random, uniformly distributed additive shifts in disparity, and was varied by altering the range of the uniform distribution from which random disparity shifts were drawn.

Our two models differed in the decision rules they applied. Here, and in each experiment in this article, one model applied a decision rule that made full use of the available disparity information, while the other relied only on the minimum and maximum disparity values. We refer to these models as the full distribution and min-max models, respectively. For the discrimination of 3-D spread, the full distribution model calculated the standard deviation of the sampled subset of dot disparities, and selected the interval that contained the greater value. The min-max model took the difference between the minimum and maximum absolute disparity values, and selected the larger interval. Monte Carlo simulations were run for each model, for both Gaussian and uniform distributions, at all standard spreads and disparity ratios shown to participants. The modeling results are shown in Figure 3, as best-fitting efficiencies and internal noise levels for thresholds averaged across all five participants.

As with human participants, thresholds for both models, in the form of fitted σ parameters, were constant across each level of standard spread. That is, discriminability was a function of the ratio of variable to standard 3-D spread. While both models therefore provided qualitatively good fits to human threshold data, neither model predicted the observed rebound in "standard greater" responses for large disparity ratios, at larger standard spreads. Model fits show that average thresholds for human participants were equivalent to sampling efficiencies of between 1.6% and 2.2%, and to internal disparity noise levels of between ± 1.68 and ± 1.97 arcmin for full distribution and min-max models, respectively. The effects of varying internal noise were small, however, compared to the effects of varying sampling efficiency. While low, our estimates of sampling efficiency are within the ranges

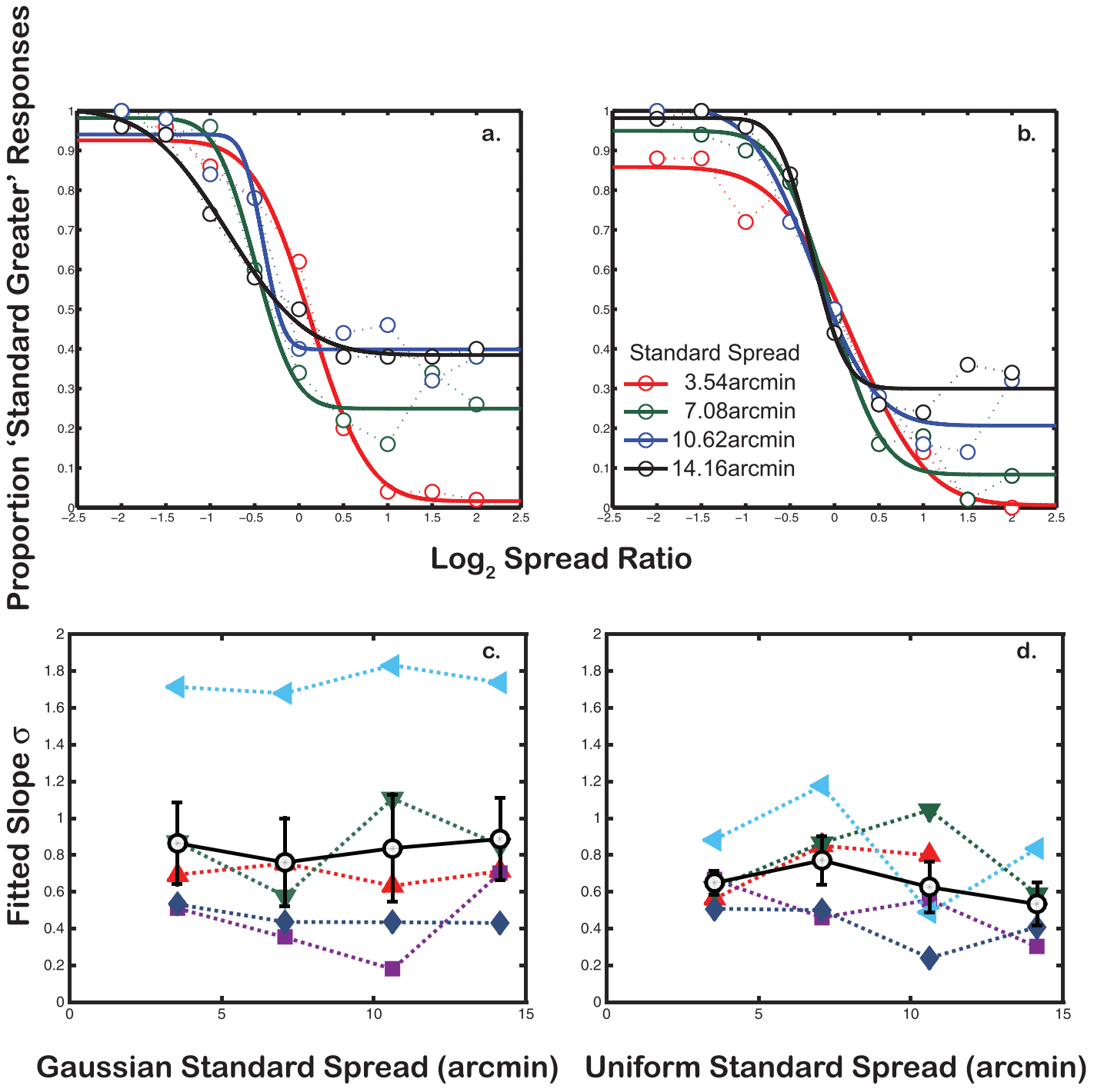

Figure 2. Results for Experiment 1. Example psychometric functions are shown for a single participant in the top row for (a) Gaussian-distributed volumes and (b) uniformly distributed volumes. The *x*-axis plots the spread ratio, (variable/standard spread); the *y*-axis plots the proportion of "standard spread greater" responses. Differing symbols show each standard spread. The lower graphs show the slopes (σ) for (c) Gaussian and (d) uniformly distributed volumes. Colored symbols and dashed lines indicate results for individual participants, while the larger black circles show the mean thresholds across all participants (error bars represent standard error of the mean). On average, there is no effect of standard spread on sensitivity to variable/standard spread ratios in spread discrimination.

established in previous studies where, depending on the task and stimulus configuration, measured efficiency has varied between around 1% and 20% (Harris & Parker, 1992; Wallace & Mamassian, 2004). Best-fitting sampling efficiencies were similar for both the full distribution and min-max models, for both uniform and Gaussian-distributed volumes. In Experiments 2 and 3, we examine these models in more detail.

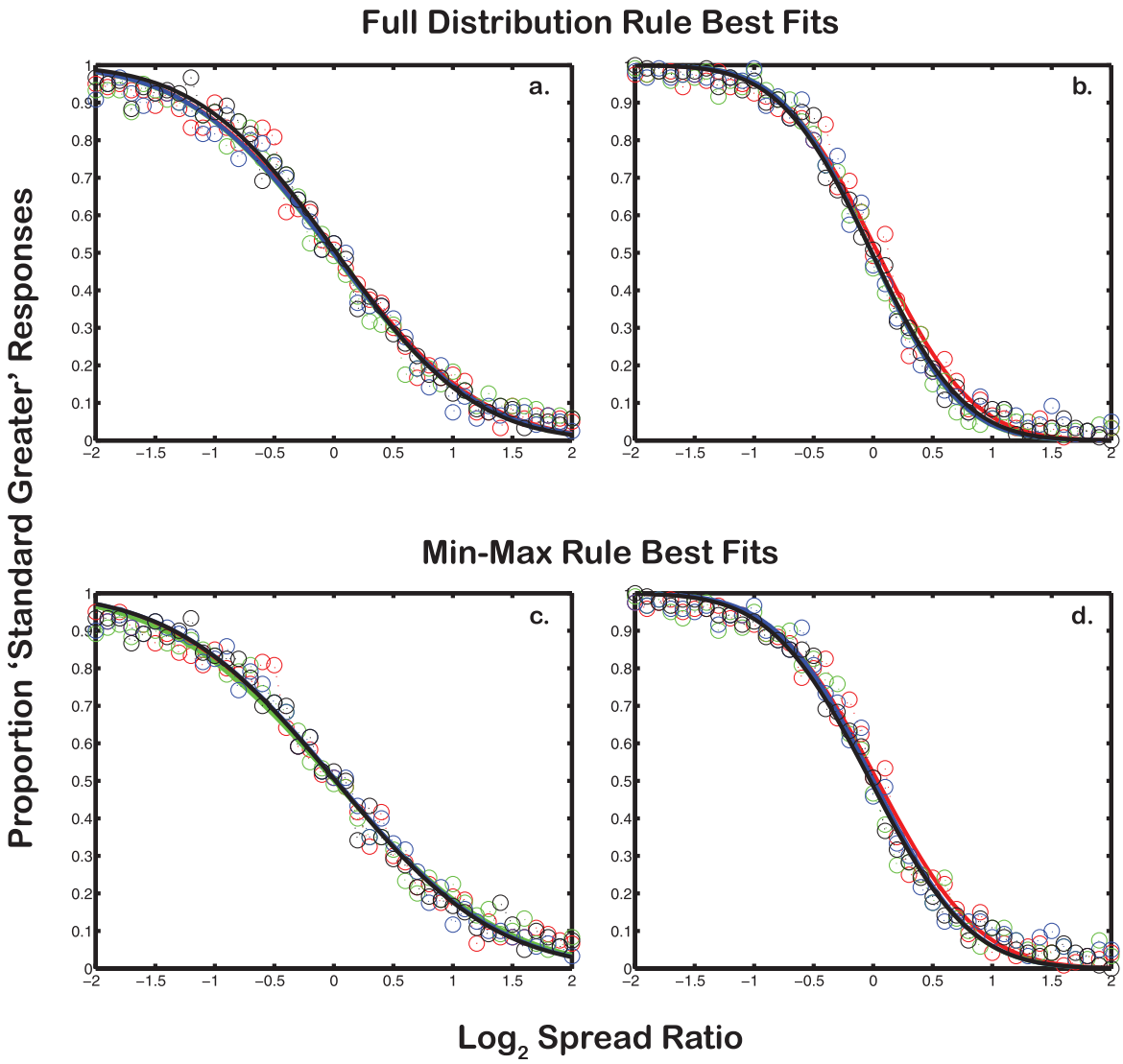

Figure 3. Model performance on the spread discrimination task used in Experiment 1. Graphs show best-fitting psychometric functions for the full distribution model for (a) Gaussian and (b) uniformly distributed volumes, and (c–d) for the same distributions under the min-max model. As in Figure 2, the *x*-axis plots the spread ratio and the *y*-axis plots the proportion of "standard spread greater" responses. Under both models, as with human observers, there is no effect of standard spread on disparity ratio thresholds. Performance in the spread discrimination task cannot, therefore, be used to discriminate between these models.

Experiment 2: Location-in-Depth Discrimination

In Experiment 1 we measured observers' ability to discriminate differences in the second moment (spread) of stereoscopic volume stimuli. In Experiment 2 we looked instead at the first moment of stereoscopic volumes, and examined observers' ability to discriminate location-in-depth. While location judgments in 2-D have been shown to involve a range of measures, including centroid, midpoint, and peak locations (Badcock, Hess, & Dobbins, 1996; Hess, Dakin, & Badcock, 1994; Hess & Holliday, 1992; Hirsch & Mjolsness, 1992; Morgan & Glennerster, 1991), we do not currently know what stimulus information is used by participants to judge location-in-depth. To address this issue, participants were required to judge whether the center of a disparity-defined volume was in front of or behind the fixation plane. Responses were compared to the performance of models using variants of the full distribution and min-max rules applied in Experiment 1. To differentiate between these decision rules, stimuli were comprised of dots with disparities drawn from a combination of two random distributions. Fifty percent of dots were Gaussian distributed in depth, while the remaining 50% were drawn from a uniform distribution with a larger range of disparities. Locations-in-depth were independently varied for each distribution, and participants were

free to judge the location of the distribution in depth using any available decision criteria. If participants rely on minimum and maximum disparities, then judgments of location-in-depth will be largely unaffected by the locations of Gaussian-distributed dots. In contrast, Gaussian-distributed dots will have a greater effect on perceived location-in-depth if participants make full use of the full distribution of dot disparities.

Methods

Participants

Five participants completed Experiment 2, including author RG. All participants were staff or students of the University of Stirling and had normal or corrected-to-normal vision. Stereoacuity was measured using the RanDot2 test (VAC, Elk Grove Village, IL), and was less than 40 s of arc for all participants. Nonauthor participants were naive as to the purpose of the experiment and the structure of the stimuli. Participants provided written consent for their participation. All experimental procedures were approved by the University ethics committee, and were consistent with the guidelines of the British Psychological Society and the Declaration of Helsinki.

Stimulus and apparatus

The experiment was conducted using a MacPro computer and a 49- × 31-cm Apple HD Cinema display. Display resolution was 1920 × 1200 pixels, with a refresh rate of 60 Hz. As in Experiment 1, stimulus presentation and creation was controlled using Matlab (Mathworks, Inc.), together with the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). Display output was calibrated to a linear luminance scale using a SpyderPro2 calibration device (ColorVision, DataColor, Lawrenceville, NJ). Participants viewed the display through a mirror stereoscope, calibrated for a viewing distance of 76.4 cm. At this viewing distance a single pixel measured 1.1 arcmin in both x and y dimensions.

As in Experiment 1, stimuli were random-dot disparity volumes. Dot density within each volume was 15.2 dots per degree^2, with each volume measuring 4.77° × 4.77°. A single dot measured 5.5 arcmin in diameter. RDS elements were once again high contrast white circles, against a black background. A small fixation cross measuring 7.7 × 7.7 arcmin, with the same luminance as the stimulus dots was presented at the center of the display. As in Experiment 1, zero disparity reference planes were presented above and below each stimulus. Identical reference planes were used for each trial within an experimental block, and were visible both before and after each stimulus presentation. Each reference plane covered an area of 5.5° × 0.37°. Stimuli were presented for 600 ms, preceded by a 500-ms presentation of the fixation cross and zero disparity reference planes. The fixation cross and reference planes were also present during stimulus presentation. While the experimental set-up for Experiment 2 (and 3, below) is not identical to that used in Experiment 1, all efforts were taken to ensure that angular dimensions for stimuli, including dot size, stimulus size, and dot density, were consistent between experiments.

The disparity content of each volume was defined as a combination of uniformly distributed and Gaussian-distributed dots in depth. For any given stimulus, 50% of the dots had disparities selected at random from a uniform distribution with maximum and minimum values ±6.6 arcmin from its center. The remaining 50% of dots had disparities selected at random from a Gaussian distribution with a standard deviation of 2.2 arcmin. Note that all stimulus disparities were well within the perceptible range, as defined by the results of Experiment 1. We parametrically varied the mean location-in-depth of the Gaussian distribution and the center of the uniform distribution from which dot disparities were drawn. As in Experiment 1, a dot's disparity did not depend upon its x- or y-coordinates.

Design and procedure

Over eight blocks, each participant viewed 40 repeated trials of each combination of Gaussian mean and uniform center in a single interval, two-alternative forced choice task. Five Gaussian mean locations-in-depth and five uniform distribution centers were used, spread evenly across a range of ±2.2 arcmin, resulting in 25 combinations of Gaussian and uniform locations-in-depth. Two participants viewed a larger range of locations, spanning ± 6.6 arcmin, to provide a fuller range of responses across the range of the psychometric curve. Stimulus order was randomized within blocks, with each block containing five repeated trials of each combination of Gaussian and uniform location. On each trial, participants were asked to decide whether the middle of the cloud was in front of or behind fixation, with the proportion of "in front" responses recorded for each combination of Gaussian and uniform mean location-in-depth.

Results and discussion

Psychophysical performance

Psychometric functions were fitted to the proportion of "in front" responses for each Gaussian mean location, as a function of uniform distribution location. PSEs were extracted from each function, giving the location for the uniform distribution required for

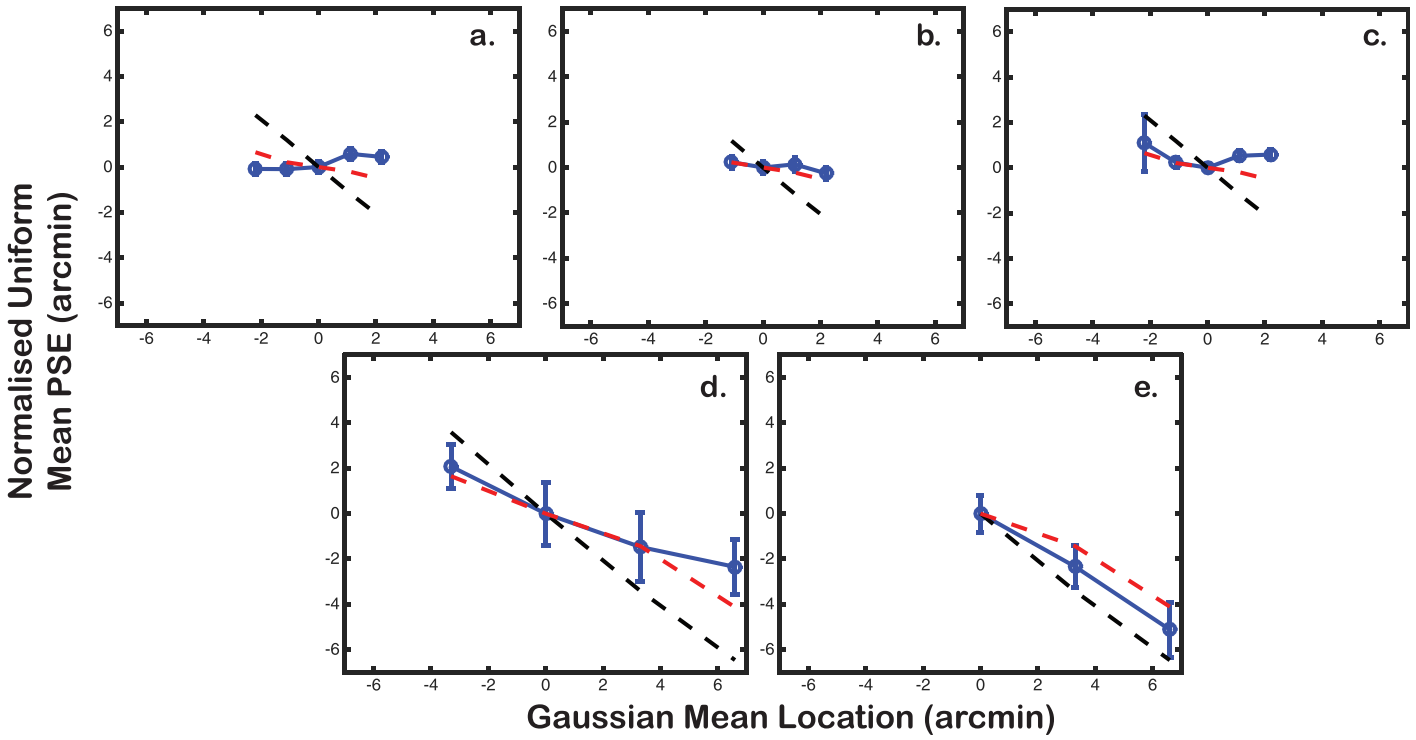

Figure 4. Psychophysical and modeling results from Experiment 2. Uniform distribution PSEs for each participant, as a function of Gaussian mean location (error bars show the standard error of the mean). Blue circles show measured normalized uniform distribution PSEs, with red solid lines showing the best-fitting predictions of the min-max model, and black dashed lines showing the best-fitting predictions of the mean model. Human performance is better accounted for by the min-max decision rule.

participants to respond that the cloud was "in front" on 50% of trials. PSEs were then normalized by dividing by the PSE obtained, for each participant, when the Gaussian mean location was at fixation. This process of normalization was necessary to account for large biases for crossed disparities, exhibited by all observers (i.e., participants tended to favor the "in front" response). Bias for "in front" responses also led to an inability to fit a total of four functions to crossed disparity displacements, from three different observers. Normalized PSEs are plotted as a function of Gaussian mean location, for each participant, in Figure 4. For most observers, PSEs were largely unaffected by changes in the mean location of Gaussian-distributed dots (a maximum shift of 1 arcmin for a 2.2-arcmin shift in mean location). Larger effects were found, however, for one of the participants who viewed a larger range of locations-in-depth (a 4.7-arcmin shift for the larger 6.6-arcmin range). The modest effect of mean Gaussian location is surprising, given that changing the location-in-depth of Gaussian-distributed dots alters the disparity of half the dots in the display. Such results suggest that the visual system does not make full use of the available stimulus information. Below, we apply adapted versions of the models used in Experiment 1 to judgments of location-in-depth, and compare these to our psychophysical data.

Modeling location discrimination in depth

The limited effect of varying the mean location-in-depth of Gaussian-distributed dots suggests that most participants' judgments of location are determined by a representation of the stimulus that contains information about only the extremes of the disparity volume. The visual system appears to have little knowledge of the distribution of disparities within the volume. This can be demonstrated more fully by considering variants of the full distribution and max-min models, applied in Experiment 1. The two models from Experiment 1 were modified to provide responses appropriate to the location-in-depth discrimination task used in Experiment 2. For the full distribution model, this meant finding the mean of the subset of sampled dot disparities. For the min-max model, volume location was determined by taking the difference between the maximum and minimum sampled disparities. For each model the sign of the calculated value (mean or min-max difference) is taken as the two-alternative forced choice response for the location of the middle of the distribution.

Using Monte Carlo simulations, best fits for each model were once again obtained by varying sampling efficiency and internal additive noise. As with model results for Experiment 1, within the range examined here, additive noise had little impact on model

performance, although best-fitting internal noise was around ±1.1 arcmin. Best-fitting values for sampling efficiency were ~5%. This was higher than observed in Experiment 1 but once again within the range found in other relevant studies. Importantly, the min-max model offers a significantly better fit to participants' data than the full distribution model.

Figure 4 shows lines of best-fit for each participant and each model. For four of the five participants, the mean model predicts a greater than observed shift in normalized PSE, with bootstrapped 95% confidence intervals (CIs) for the difference in mean squared error terms between models showing significantly smaller errors for the min-max model than for the full distribution model. The 95% CIs for the difference in mean squared errors for these four participants were as follows: 1.86 to 3.62; 0.68 to 2.26; 1.11 to 3.02; and 2.18 to 6.54. For the single participant where the min-max model did not show a significant advantage, 95% CIs for the difference in mean squared error ranged from –0.94 to 2.32. Note that, while this model can account for the observed pattern of results in terms of normalized PSEs, neither model is able to account for the observed bias towards "in front" responses. Such biases could stem from a range of factors, including issues related to viewing geometry, eye movements (e.g., Jaschinski, Švede, & Jainta, 2008; Zaroff, Knutelska, & Frumkes, 2003), or cognitive factors. Given that this bias should be constant across conditions and is not central to the question under study, we have not pursued it in more detail here.

Data were also analyzed by comparing R^2_{adj} values for linear fits of measured to predicted PSEs. Monte Carlo simulations of best-fitting model parameters were used to generate multiple PSE predictions, with 95% CIs obtained through bootstrap resampling. The average R^2_{adj} value across all participants was 0.78 for the min-max model, with bootstrapped 95% CIs ranging from 0.75 to 0.79. For the full distribution model the mean R^2_{adj} value was 0.63, with 95% CIs ranging from 0.60 to 0.64. Bootstrapped 95% CIs for the difference in model R^2_{adj} values ranged from 0.12 to 0.19. The min-max model therefore offered a significantly better fit to the data than the full distribution model.

That estimates of location-in-depth are better fit by a min-max decision rule than by the distribution mean is in sharp contrast to findings for the perception of the location of 2-D distributions. For 2-D distributions, perceived location follows a robust mean rule (Juni et al., 2010; Morgan & Glennerster, 1991). Deviation from this mean decision rule when making judgments in depth suggests that the visual system processes distributions in depth very differently from 2-D distributions. We consider possible explanations for this difference in the Importance of stereoscopic volume judgments section, below.

Experiment 3: Odd-one-out discrimination

In Experiments 1 and 2, participants made judgments about the spread and location-in-depth of stereoscopic volumes (respectively). The results of Experiment 2 suggested that participants made use of only a limited amount of the information available in these stimuli. Judgments of location-in-depth appear to be defined by the extreme values of the volume, and not by any measure that takes into account the disparities of all stimulus dots. If such limited use of the information in disparity volumes is the norm, this is likely to limit the extent to which observers are able to make more complex judgments about volume structure, such as discriminating the shape of depth distributions. This issue is examined in Experiment 3, where participants were asked to determine the odd-one-out of three stimuli, where the target volume differed from comparators in both spread and shape.

Methods

Participants

Experiment 3 was completed by four participants, including author RG. Two of these participants, including author RG, also completed Experiment 2. All participants were staff or students of the University of Stirling and all had normal or corrected-to-normal vision. Stereoacuity was measured using the RanDot2 stereotest, and was less than 40 s of arc for all participants. Nonauthor participants were naive as to the purpose of the experiment and the structure of the stimuli. All experimental procedures were approved by the University ethics committee, in accordance with British Psychological Society guidelines and the Declaration of Helsinki.

Stimulus and apparatus

Stimuli were created and displayed using the apparatus described in Experiment 2. Stimulus parameters (e.g., angular size, dot density) were the same as outlined in Experiment 2, with the exception of the differences noted below.

Two of the three volumes presented on each trial contained disparities drawn at random from identical Gaussian distributions. These distributions were always of mean 0, with standard deviations of either 1.1 or 4.4 arcmin, depending on the experimental block. We refer to these volumes as the comparison intervals. The third volume, the target interval, contained a proportion of dots with disparities drawn at random from a Gaussian distribution identical to that used for the comparison

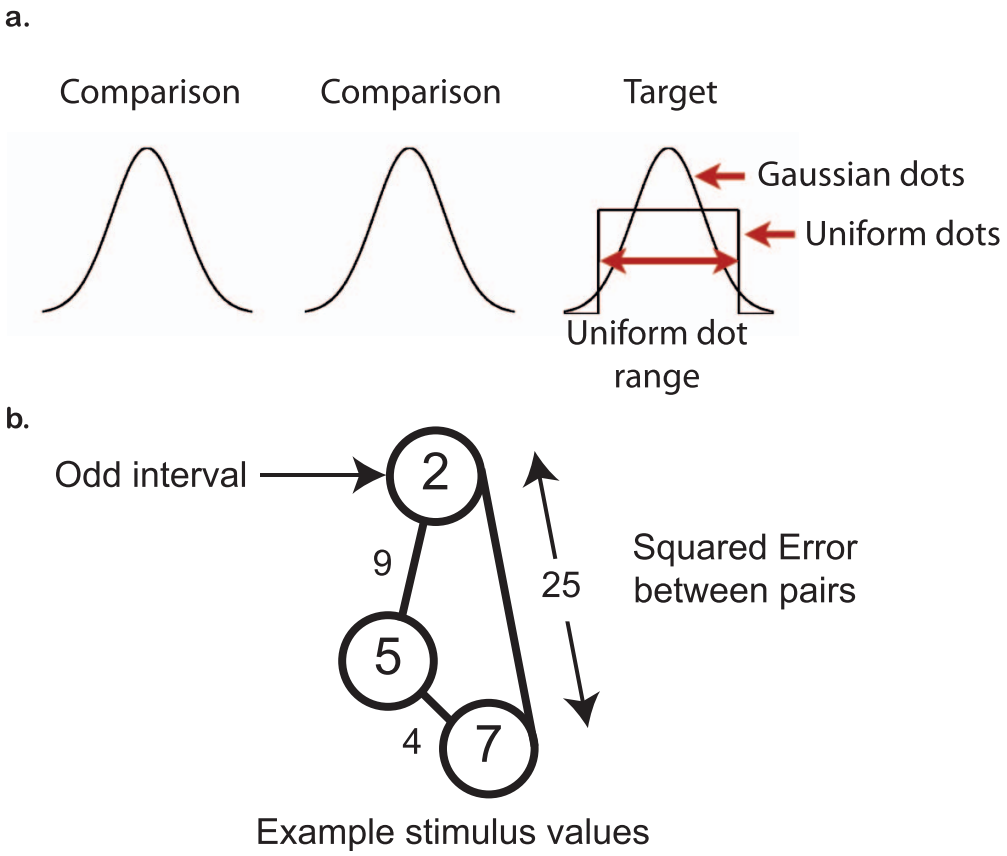

Figure 5. (a) An illustration of an example trial in Experiment 3. Participants are presented with three intervals (in random order), where the target interval varies in the proportion of dots drawn from a uniform distribution and/or the range of the uniform distribution from which dot disparities are drawn. Comparison intervals contain disparities drawn from identical Gaussian distributions. (b) Illustration of the least squared error decision rule employed by both min-max and full distribution models in the 3IFC oddity detection task. Let us assume that three random stimuli may be summarised by random integers. The difference between pairs of these stimuli can be described by their squared errors—the smaller the squared error term, the more similar the stimulus pair. For any stimulus triplet, the odd interval may therefore be defined as the stimulus that is not part of the pairing with the least squared error (i.e., the most similar stimulus pairing). This may be seen in the illustration, where the squared error terms determine the line length between stimulus pairs. By necessity the odd interval is farther from each of the other stimulus intervals, than they are from each other.

intervals. The remaining dots had disparities drawn at random from a uniform distribution of variable range. We parametrically varied both the proportion of uniformly distributed dots and the range of the uniform distribution from which dot disparities were drawn. Four proportions (0.25 to 1) of uniformly distributed dots were used, with seven ranges between ±1.1 and ±7.7 arcmin (see Figure 5a).

Design and procedure

Participants in this study were sequentially shown three RDS volumes per trial in a three-interval forced-choice (3IFC) paradigm. On each trial they were asked to determine which of the three intervals contained the "odd-one-out," with no further instruction given (see Figure 5a). As such, each participant was free to respond to any combination of possible stimulus features when selecting the odd interval. With instructions of this kind, the odd-one-out task offers a criterion-free, objective means of assessing observers' perception of difference in the three intervals, with any detectable difference able to provide a basis for observers' responses (cf. Frijters, 1981; Hillis, Ernst, Banks, & Landy, 2002).[1] Participants indicated their choice by pressing one of three keys on a keyboard.

Each disparity volume was presented for 600 ms and was preceded by a 500-ms fixation cross (also present during stimulus presentation), with the order of comparison and target intervals randomized across trials. As stated above, 2 standard deviations, of 1.1 or 4.4 arcmin, were used for the Gaussian-distributed comparison intervals. Participants completed six experimental blocks, resulting in 30 repeated trials of 28 combinations of uniform distribution range and proportion uniformly distributed dots for each com-

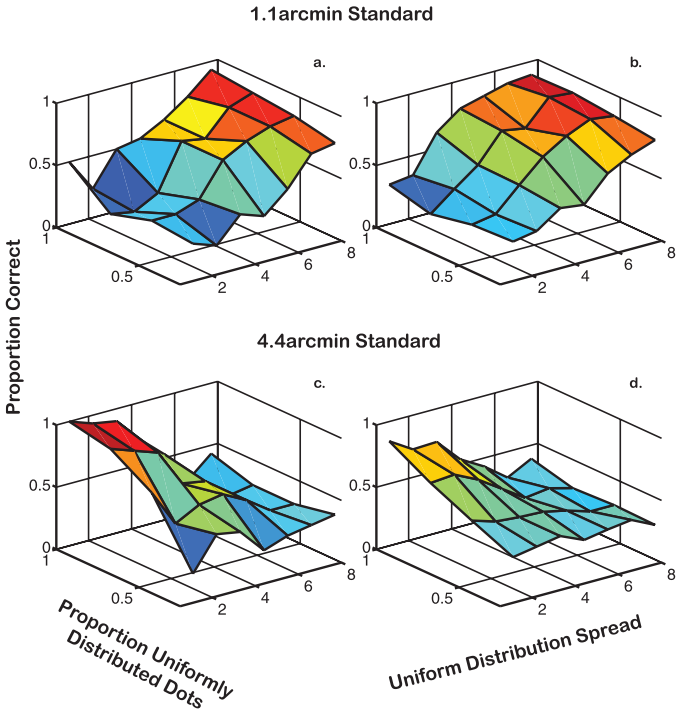

Figure 6. Psychophysical results from the odd-one-out detection task completed in Experiment 3. Plots of experimental results for (a) an example participant and (b) for the average of all four participants, in the 1.1 arcmin comparison interval condition. Plots show proportion correct scores for the odd-one-out task (*z*-axis) for each combination of uniform distribution range and proportion of uniformly distributed dots. Equivalent plots are also shown for (c) the same example participant and (d) the average of all four participants in the 4.4 arcmin comparison interval condition. Performance is determined largely by the uniform distribution range, suggesting that participants are insensitive to distribution shape.

parison interval standard deviation, for a total of 1,680 trials for each participant.

Results and discussion

Psychophysical performance

Performance in the odd-one-out detection task is shown in Figure 6a and c, for an example participant, and Figure 6b and d for the average of all four participants. Figure 6a through d show participants' performance when the comparison interval contained elements drawn from Gaussian distributions with standard deviations of 1.1 arcmin and 4.4 arcmin, respectively. Note that chance performance is 0.3 for this task. Inspection of the averaged results shows that when the standard deviation of the Gaussian distribution used to generate the standard stimulus is relatively small (1.1 arcmin), proportion correct increases as the spread of the uniform distribution increases from 2 to 8 arcmin. While there is a tendency for performance to improve as a function of the proportion of uniformly distributed elements, this trend is weak and variable compared with the impact of distribution range. The same pattern of results is obtained when the standard deviation of the Gaussian distribution used to generate the standard is large (4.4 arcmin). In this case (Figure 6d) accuracy improves when the target has a substantially *lower* range of disparities. In the latter case, overall performance is quite poor, and there is a wide range of distributions that are indistinguishable, including some cases in which the target consists entirely of uniformly distributed elements. The differences between the 1.1 and 4.4 arcmin conditions likely reflects degraded depth percepts in the large disparity range for these RDS stimuli.

Modeling odd-one-out detection

Participants' reliance on differences in range during the odd-one-out detection task suggests a general lack of sensitivity to the information contained in disparity volume stimuli, or at least suggests a lack of sensitivity to information about distribution shape. Such performance may also arise, however, if differences in the shapes of the distributions offer a comparatively poor cue to odd-one-out detection. To investigate this possibility, we once again examined the responses of variants of the full distribution and min-max models used in Experiments 1 and 2.

The variant of the full distribution model applied here ordered the disparities in each volume. Ordered dot disparities were then used to calculate the cumulative distribution function for each volume across all possible disparity values. The min-max model simply identified the maximum and minimum disparity values for each volume. As in Experiments 1 and 2, model performance was impaired through limiting sampling efficiency, and through the addition of uniformly distributed random disparity noise. For the full distribution model, comparisons between intervals were made by calculating the sum of squared differences between stimulus pairs (i.e., the sum of squared differences between cumulative distributions of disparities). For the min-max model, stimulus pairs were compared by finding the sum of squared differences in maximum and minimum disparities. For both models, the odd interval was selected by finding the stimulus pair where the sum of squared differences was minimized (i.e., the pair with the greatest similarity), and choosing the stimulus not included in that pair (see Figure 5b).

Monte Carlo simulations were used to find best-fitting sampling efficiencies and additive noise values for both the full distribution and min-max models, which allowed us to determine the extent to which each

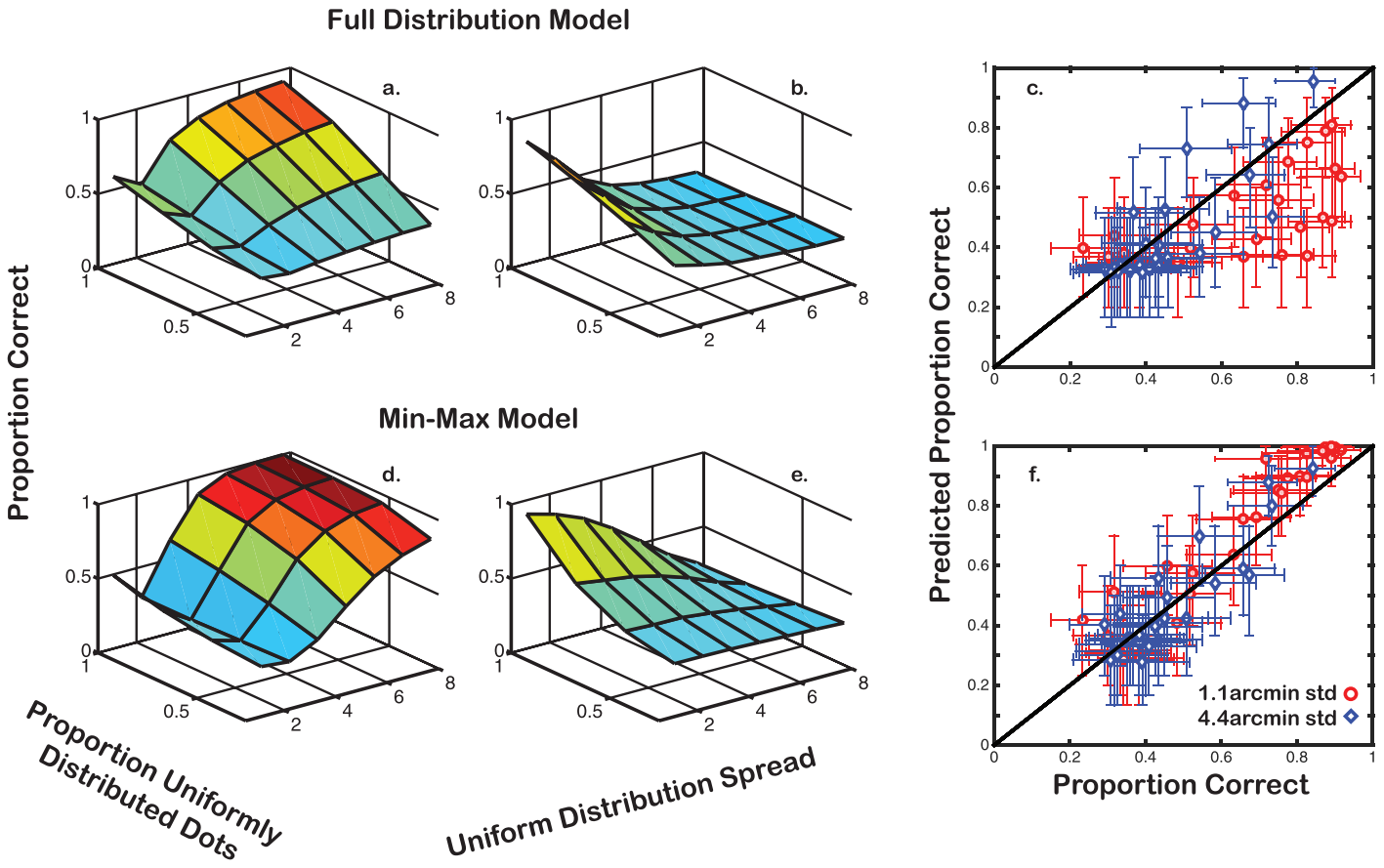

Figure 7. Modeling results for Experiment 3. Plots show best-fitting performance for the full distribution model in the (a) 1.1 arcmin and (b) 4.4 arcmin comparison interval condition, together with (c) a plot of predicted versus measured proportion correct scores for the average of all four participants. Red circles show results for the 1.1 arcmin condition, with blue diamonds showing the 4.4 arcmin condition (error bars show bootstrapped 95% CIs on both measured and predicted proportion correct scores). Equivalent plots are shown in (d–f) for the min-max model. Human psychophysical performance is better accounted for by the min-max decision rule.

decision rule accounted for participants' psychophysical performance. Figure 7 shows the responses of the best-fitting models for each comparison interval standard deviation and each decision rule.

The results of the best-fitting full distribution model are shown in Figure 7a through c, with the best-fitting min-max model shown in Figure 7d through f. Figure 7c and f plots the relationship between psychophysical performance and the results of the best-fitting models. Best-fitting sampling efficiencies were 5.3% for the min-max model and 5.5% for the full distribution model, with a fitted additive noise level of 1.1 arcmin for both models. R^2 values for the correlation between predicted and observed responses were 0.8 for the full distribution model and 0.94 for the min-max model, with bootstrapped 95% CIs for the difference in R^2 values ranging from 0.09 to 0.19. These results indicate that the min-max model offers a significantly better fit to the data than the full distribution model. In addition, while both models do well in matching psychophysical performance, the min-max model is noticeably better at predicting occasions where the proportion of correctly discriminated intervals is high. This is particularly evident for the 1.1 arcmin comparison intervals. For the full distribution model, however, much of its ability to match performance resides at the lower end of the scale, where discriminability is at or near chance level.

These modeling results are consistent with the proposal that performance in the odd-one-out detection task is largely driven by the use of stimulus information related to the range of disparities within each volume, and not by information related to the shape of the distribution of such disparities. While our results do not preclude sensitivity to the shape of the distribution of disparities in general, they suggest that, as in Experiment 2, participants do not readily make use of such information, preferring, instead, to rely on simpler measures of location and extent as the basis for their responses in our psychophysical tasks.

If use of a min-max decision rule does reflect the extent of the visual system's representation of stereoscopic volumes, then one must ask why the visual system does not make use of a more complete description of such stimuli. One possibility is that the

min-max rule offers the visual system a faster, "good enough" response to most behaviorally relevant questions, than would be possible if the full distribution of disparities were analyzed. Another possibility is raised by looking at how the responses of the min-max and full distribution models vary with changes in sampling efficiency and additive noise. Since the full distribution model uses more of the available stimulus information, its performance exceeds the min-max model at low levels of internal noise and at high efficiencies. This is not the case, however, when these factors are impaired. As additive noise increases and efficiencies decline, the performance of the full distribution model falls off, while the min-max model remains relatively stable. This suggests that the min-max model may offer a more robust means of describing the volume, or at least limited aspects of the volume, given limitations in processing efficiency.

General discussion

The experiments reported in this paper examined the perception of disparity-defined volume, an aspect of stereoscopic vision that has been comparatively neglected, despite its importance to the phenomenology of stereopsis. Our results are consistent with an account of stereoscopic volume perception that relies on the limited use of disparities within the volume, contrary to findings for 2-D distributions (e.g., Juni et al., 2010; Morgan & Glennerster, 1991). Such limitations may have important consequences for judgment of space-in-depth. We discuss these possibilities below, and address possible factors underlying observed performance limitations.

Importance of stereoscopic volume judgments

The suggestion that the visual system uses only a limited representation of the information within disparity volumes raises questions about the extent to which such representations are suited for real world judgments of 3-D space. For example, although stereoscopic vision provides a qualitatively distinct perception of the space between tree branches, the ability to navigate through changing branch densities may be somewhat restricted. Instead, volume judgments may be more suited to perceiving the depth extent of a cluttered space. This has implications both for how we consider the higher functioning of postdisparity-measurement stereo processing (see Parker, 2007), and for the production and use of stereoscopic images depicting depth volumes. There may be, for example, unforeseen problems in the use of stereoscopic presentation for data visualization (e.g., Landsberg, Moran-Jones, & Smith, 2006; Vogt & Wagner, 2012).

Apparent limits in judgments of volume location-in-depth also compare unfavourably to similar judgments in 2-D. In such stimuli location judgments appear consistent with either the 2-D mean of element positions (Hirsch & Mjolsness, 1992; Juni et al., 2010; Morgan & Glennerster, 1991) or the mode (Hess & Holliday, 1992). While some studies of 2-D localization have found use of equivalent midpoint (i.e., max-min) calculations, this has been in addition to the calculation of mean location, not at its expense (Badcock et al., 1996; Hess et al., 1994). One must, therefore, consider potential reasons for this apparent discrepancy between 2-D and 3-D localization processes, and for the observed limited use of depth distribution information. Below, we consider whether disparity measurement processes may play a role in limiting the representation of stereoscopic volumes. Here, however, we consider whether such limited representations may act as an appropriate initial description of distributions in depth.

Our results show that observers default to a very basic description of depth distributions. These results do not show, however, that such descriptions are either completely unavailable for use, or that they cannot become available over time. Such changes in information use have been noted, for example, within the literature on binocular slant perception. With stereoscopic slant stimuli, estimates of slant increase, and become more accurate, with increasing presentation time up to 10 s, well above the 600 ms used here (van Ee & Erkelens, 1996). Slant estimates have also been found to improve with the presence of stereoscopic boundaries (Gillam, Flagg, & Finlay, 1984). It may be the case that, in a similar fashion, representations of the internal structure of disparity volumes gain complexity over time, such that judgments of density changes or center of mass are possible with prolonged presentation. Prolonged stimulus presentation could also allow for the increased use of other cues, such as sequential stereopsis (Enright, 1996), brightness gradients (Samonds, Potetz, & Lee, 2012), or (in real-world stimuli) blur gradients (Watt, Akeley, Ernst, & Banks, 2005). Note that, while the absence of these cues may increase prior biases for flatness (e.g., van Ee et al., 2003), such biases should not, in and of themselves, limit the visual system's ability to characterize the distribution of disparity volumes. It remains to be seen if volume localization exhibits similar dependencies on extending viewing time and internal structure.

Even if more complex representations of depth distributions are found to emerge under prolonged viewing conditions, the question still remains as to why the visual system appears to default to the simple min-max estimate. It could be that measurement of maxima

and minima provides a rapid, first measure of the scene, which is then used for subsequent detailed analysis. However, for judgments of location-in-depth in particular, there seems no obvious reason why a max-min decision rule would be preferred on the basis of processing speed or simplicity, since coarse scale disparity measurement processes would, by default, provide a measure of mean disparity (see further discussion of this issue in the following section). An alternative possibility is that the max-min decision rule has ecological validity for typical depth judgment tasks in natural scenes. Such validity would depend upon the relationship between meaningful depth judgment tasks and external characteristics of disparity volumes, such as extent, rather than internal characteristics like density. It would also depend upon uniform distributions in depth approximating the statistics of natural environments. While the former possibility is extremely difficult to define in any operational manner, support for the second proposal comes from work by Hibbard (2007), who found that uniform random depth distributions generate disparity statistics comparable to those found in natural scenes (Hibbard, 2008; Liu, Bovik, & Cormack, 2008; Yang & Purves, 2003). Such results suggest that a max-min rule for the judgment of disparity volumes may prove effective as a first approximation within natural scenes.

Limitations in disparity measurement

One pertinent question to address when considering performance limitations in disparity volumes is to ask the extent to which such limitations are attributable to processing involving disparity measurement. Limitations in disparity measurement processes have been used to account for performance in a number of stereoscopic tasks, with supporting evidence derived from models measuring disparity through a process of cross-correlation (Allenmark & Read, 2010, 2011; Banks et al., 2004; Filippini & Banks, 2009). Harris (2014) has also used such a model to account for reduced perceptions of depth in stereoscopic volumes.

Although we have not used a cross-correlation model of disparity measurement, our modeling did include factors affecting disparity measurement through manipulations of both additive disparity noise and sampling efficiency. While these factors may not directly replicate errors in measurement arising from disparity estimation processes, they do show how general reductions in the quality of such measurements affect performance. As such, while these manipulations suggest a role for disparity measurement in providing an overall limit on task performance, they cannot account for the patterns of results obtained in Experiments 2 and 3. To account for both sets of results the decision rule must make only limited use of available disparity information.

Modeling disparity estimation noise through, for example, cross-correlation modeling is unlikely to change this decision rule requirement. There are two main reasons for this, both related to the fact that cross-correlation processes will, by necessity, result in a regression to the mean disparity across the local correlation window. First, in Experiment 2, local cross-correlation processes should, by virtue of disparity averaging, bias disparity measurements towards the mean disparity, rather than the max-min decision. As an example of this, if the correlation window used for the task in Experiment 2 were the size of the stimulus, the resulting disparity measurement would take into account the disparities of all dots. Consequently, its output would be equivalent to the results of the mean decision rule applied in the full distribution model. We have shown that this model fails to account for our findings.

The second reason to doubt a cross-correlation based account of our findings is that, in Experiment 3, there is no a priori reason to suppose that the disparity averaging that occurs with cross-correlation would obscure the shape of the distribution any more than manipulations of additive noise and sampling efficiency. Averages of Gaussian-distributed disparities, through local windowed cross-correlation, should produce sets of disparity measurements that are themselves Gaussian distributed, with similar mapping occurring for uniformly distributed disparities. Performance in the odd-one-out discrimination task would therefore seem to be driven primarily by a limited use of the available stimulus information at the decision rule, rather than disparity measurement, stage.

Conclusions

The experimental results and analysis contained in this article suggest that, despite the fact that a compelling qualitative sense of depth is reported when viewing disparity-defined volumes, the visual system makes only limited use of the information available in such stimuli. Our results suggest that performance in volume-related psychophysical tasks is driven primarily by information about the range of disparity volumes, taken as a simple measurement related to maximum and minimum disparity values. The visual system appears to default to this simple measure of our volume stimuli, rather than make use of a fuller representation of the distribution of disparities. Further experimentation is require to determine whether these limitations generalize to other stimulus configurations and tasks, and whether they reflect an encoding strategy ineffi-

ciency or a lack of capacity to represent the distribution of elements in such disparity-defined volumes.

Keywords: binocular vision, stereopsis, odd-one-out task, 3-D volumes

Acknowledgments

This research was funded by a Carnegie Trust Travel Grant, Biotechnology and Biological Sciences Research Council Grant No BB/G004803/1 (RG), and by a National Sciences and Engineering Research Council Individual Discovery grant (LW).

Commercial relationships: None.
Corresponding author: Ross Goutcher.
Email: ross.goutcher@stir.ac.uk.
Address: Faculty of Natural Sciences (Psychology), University of Stirling, Stirling, UK.

Footnote

[1] Note that, given the stochastic nature of the stimuli used, the 3IFC method offers an objective means of judging oddity, where a 2IFC same-different measure would not. In a same-different measure the choice of different is always legitimate, given different sets of disparity samples (even if drawn from the same distribution).

References

Akerstrom, R. A., & Todd, J. T. (1988). The perception of stereoscopic transparency. *Perception & Psychophysics*, *44*, 421–432.

Allenmark, F., & Read, J. C. A. (2010). Detectability of sine- versus square-wave disparity gratings: A challenge for current models of depth perception. *Journal of Vision*, *10*(8):17, 1–16, doi:10.1167/10.8.17. [PubMed] [Article]

Allenmark, F., & Read, J. C. A. (2011). Spatial stereoresolution for depth corrugations may be set in primary visual cortex. *PLoS Computational Biology*, *7*(8), e1002142, doi:10.1371/journal.pcbi.1002142.

Badcock, D. R., Hess, R. F., & Dobbins, K. (1996). Localization of element clusters: Multiple cues. *Vision Research*, *36*, 1467–1472.

Badcock, D. R., & Schor, C. M. (1985). Depth-increment detection function for individual spatial channels. *Journal of the Optical Society of America, A*, *2*, 1211–1215.

Banks, M. S., Gepshtein, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? *The Journal of Neuroscience*, *24*, 2077–2089.

Barry, S. R. (2009). *Fixing my gaze: A scientist's journey into seeing in three dimensions*. New York: Basic Books.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Enright, J. T. (1996). Sequential Stereopsis: a Simple Demonstration. *Vision Research*, *36*(2), 307–312.

Filippini, H. R., & Banks, M.S. (2009). Limits of stereopsis explained by local cross-correlation. *Journal of Vision*, *9*(1):8, 1–18, doi:10.1167/9.1.8. [PubMed] [Article]

Foley, J. M., Applebaum, T.H., & Richards, W.A. (1975). Stereopsis with large disparities: discrimination and depth magnitude. *Vision Research*, *15*, 417–421.

Frijters, J. E. R. (1981). An olfactory investigation of the compatibility of oddity instructions with the design of a 3-AFC signal detection task. *Acta Psychologia*, *49*, 1–16.

Gillam, B. (1968). Perception of slant when perspective and stereopsis conflict: experiments with aniseikonic lenses. *Journal of Experimental Psychology*, *78*, 299–305.

Gillam, B., Flagg, T., & Finlay, D. (1984). Evidence for disparity change as the primary stimulus for stereoscopic processing. *Perception & Psychophysics*, *36*(6), 559–564.

Goutcher, R., & Hibbard, P.B. (2014). Mechanisms for similarity matching in disparity measurement, *Frontiers in Psychology*, doi:10.3389/fpsyg.2013.01014

Harris, J. M. (2014). Volume perception: Disparity extraction and depth representation in complex three-dimensional environments. *Journal of Vision*, *14*(12):11, 1–16, doi:10.1167/14.12.11. [PubMed] [Article]

Harris, J. M., & Parker, A.J. (1992). Efficiency of stereopsis in random-dot stereograms. *Journal of the Optical Society of America, A*, *9*, 14–24.

Harris, J. M., & Parker, A.J. (1994). Constraints on human stereo dot matching. *Vision Research*, *34*, 2761–2772.

Hess, R. F., Dakin, S.R.G., & Badcock, D.R. (1994). Localization of element clusters by the human visual system. *Vision Research*, *34*, 2439–2451.

Hess, R. F., & Holliday, I.E. (1992). The coding of spatial position by the human visual system: Effects of spatial scale and contrast. *Vision Research*, *32*, 1085–1097.

Hibbard, P. B. (2007). A statistical model of binocular disparity. *Visual Cognition*, *15*, 149–165.

Hibbard, P. B. (2008). Binocular energy responses to natural images. *Vision Research*, *48*(12), 1427–1439.

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

Hirsch, J., & Mjolsness, E. (1992). A centre-of-mass computation describes the precision of random dot displacement discrimination. *Vision Research*, *32*, 335–346.

Jaschinski, W., Švede, A., & Jainta, S. (2008). Relation between fixation disparity and the asymmetry between convergent and divergent disparity step responses. *Vision Research*, *48*, 253–263.

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

Julesz, B. (1964). Binocular depth perception without familiarity cues. *Science*, *145*, 356–361.

Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: University of Chicago Press.

Juni, M. Z., Singh, M., & Maloney, L. T. (2010). Robust visual estimation as source separation. *Journal of Vision*, *10*(14):2, 1–20, doi:10.1167/10.14.2. [PubMed] [Article]

Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007). What's new in PsychToolbox-3? *Perception*, ECVP Abstract Supplement 14.

Landsberg, M. J., Moran-Jones, K., & Smith, R. (2006). Molecular recognition of an RNA trafficking element by heterogeneous nuclear ribonucleoprotein A2. *Biochemistry*, *45*, 3943–3951.

Liu, Y., Bovik, A. C., & Cormack, L. K. (2008). Disparity statistics in natural scenes. *Journal of Vision*, *8*(11):19, 1–14, doi:10.1167/8.11.19. [PubMed] [Article]

Lunn, P. D., & Morgan, M. J. (1997). Discrimination of the spatial derivatives of horizontal binocular disparity. *Journal of the Optical Society of America A*, *14*, 360–371.

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

McKee, S. P., & Verghese, P. (2002). Stereo transparency and the disparity gradient limit. *Vision Research*, *42*, 1963–1977.

Morgan, M. J., & Glennerster, A. (1991). Efficiency of locating centres of dot-clusters by human observers. *Vision Research*, *31*, 2075–2083.

Ogle, K. N. (1952). On the limits of stereoscopic vision. *Journal of Experimental Psychology*, *44*, 253–259.

Ogle, K. N. (1953). Precision and validity of stereoscopic depth perception from double images. *Journal of the Optical Society of America*, *43*, 906–913.

Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience*, *8*, 379–391.

Parker, A. J., & Yang, Y. (1989). Spatial properties of disparity pooling in human stereo vision. *Vision Research*, *29*, 1525–1538.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Rogers, B. J., & Cagenello, R. (1989). Disparity curvature and the perception of three-dimensional surfaces. *Nature*, *339*, 135–137.

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

Stevens, K. A., & Brooks, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research*, *28*, 371–386.

Stevenson, S. B., Cormack, L. K., & Schor, C. M. (1989). Disparity tuning in mechanisms of human stereopsis. *Vision Research*, *32*, 1685–1694.

Tsirlin, I. Allison, R. S., & Wilcox, L. M. (2008). Stereoscopic transparency: Constraints on the perception of multiple surfaces. *Journal of Vision*, *8*(5):5, 1–10, doi:10.1167/8.5.5. [PubMed] [Article]

Tsirlin, I. Allison, R. S., & Wilcox, L. M. (2012). Perceptual asymmetry reveals neural substrates underlying stereoscopic transparency. *Vision Research*, *54*, 1–11.

Tyler, C. W. (1973). Stereoscopic vision: Cortical limitations and a disparity scaling effect. *Science*, *181*, 276–278.

Tyler, C. W. (1991). Cyclopean vision. In D. Regan (Ed.), *Binocular vision* (pp. 38–74). London: MacMillan.

van Ee, R., Adams, W. J., & Mamassian, P. (2003). Bayesian modeling of cue interaction: Bistability in stereoscopic slant perception. *Journal of the Optical Society of America A*, *20*(7), 1398–1406.

van Ee, R., & Anderson, B. L. (2001). Motion

direction, speed and orientation in binocular matching. *Nature*, *410*, 690–694.

van Ee, R., & Erkelens, C. J. (1996). Temporal aspects of binocular slant perception. *Vision Research*, *36*(1), 43–51.

Vogt, F., & Wagner, A. Y. (2012). Stereo pairs in astrophysics. *Astrophysics and Space Science*, *337*, 79–92.

Wallace, J. M., & Mamassian, P. (2004). The efficiency of depth discrimination for non-transparent and transparent stereoscopic surfaces. *Vision Research*, *44*, 2253–2267.

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, doi:10.1167/5.10.7. [PubMed] [Article]

Weinshall, D. (1989). Perception of multiple transparent planes in stereo vision. *Nature*, *341*, 737–739.

Weinshall, D. (1991). Seeing "ghost" planes in stereo vision. *Vision Research*, *31*, 1731–1748.

Weinshall, D. (1993). The computation of multiple matching doubly ambiguous stereograms with transparent planes. *Spatial Vision*, *7*, 183–198.

Westheimer, G., & Tanzman, I. J. (1956). Qualitative depth localization with diplopic images. *Journal of the Optical Society of America*, *46*, 116–117.

Wheatstone, C. (1838). Contributions to the physiology of vision—Part the first. On some remarkable and hitherto unobserved phenomena of binocular vision. *Philosophical Transactions of the Royal Society*, *128*, 371–394.

Wilcox, L. M., & Hess, R. F. (1995). Dmax for stereopsis depends on size not spatial frequency content. *Vision Research*, *35*, 1061–1069.

Yang, Z., & Purves, D. (2003). Image/source statistics of surfaces in natural scenes. *Network: Computation in Neural Systems*, *14*, 371–390.

Zaroff, C. M., Knutelska, M., & Frumkes, T. E. (2003). Variation ins: Normative description, fixation disparity and the roles of aging and gender. *Investigative Ophtalmology & Vision Science*, *44*(2), 891–900. [PubMed] [Article]