

# **Natural image statistics at depth edges modulate perceptual stability**

Zeynep Başgöze<sup>1,\*</sup>, David N. White<sup>2</sup>, Johannes Burge<sup>2,3,4</sup>, Emily A. Cooper<sup>1,5</sup>

<sup>1</sup>School of Optometry & Vision Science Program, University of California, Berkeley, Berkeley, CA USA

<sup>2</sup>Neuroscience Graduate Group, University of Pennsylvania, Philadelphia, PA USA

<sup>3</sup>Bioengineering Graduate Group, University of Pennsylvania, Philadelphia, PA USA

<sup>4</sup>Department of Psychology, University of Pennsylvania, Philadelphia, PA USA

<sup>5</sup>Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA USA

\*Corresponding Author: [bagze001@umn.edu](mailto:bagze001@umn.edu)

## 1 **Abstract**

2

3 Binocular fusion relies on matching points in the two eyes that correspond to the same physical  
4 feature in the world. However, not all world features are binocularly visible. In particular, at depth  
5 edges parts of a scene are often visible to only one eye (so-called *half occlusions*). Accurate  
6 detection of these monocularly visible regions is likely to be important for stable visual  
7 perception. If monocular regions are not detected as such, the visual system may attempt to  
8 binocularly fuse non-corresponding points, which can result in unstable percepts. We  
9 investigated the hypothesis that the visual system capitalizes upon statistical regularities  
10 associated with depth edges in natural scenes to aid binocular fusion and facilitate perceptual  
11 stability. By sampling from a large set of stereoscopic natural image patches, we found evidence  
12 that monocularly visible regions near depth edges in natural scenes tend to have features more  
13 visually similar to the adjacent binocularly visible background region than to the adjacent  
14 binocularly visible foreground. The generality of these results was supported by a parametric  
15 study of three-dimensional (3D) viewing geometry in simulated environments. In two perceptual  
16 experiments, we examined if this statistical regularity may be leveraged by the visual system.  
17 The results show that perception tended to be more stable when the visual properties of the  
18 depth edge were statistically more likely. Exploiting regularities in natural environments may  
19 allow the visual system to facilitate fusion and perceptual stability of natural scenes when both  
20 binocular and monocular regions are visible.

21

22 **Precis:** We report an analysis of natural scenes and two perceptual studies aimed at  
23 understanding how the visual statistics of depth edges impact perceptual stability. Our results  
24 suggest that the visual system exploits natural scene regularities to aid binocular fusion and  
25 facilitate perceptual stability.

26

27 **Keywords:** binocular fusion, perceptual stability, natural image statistics

28

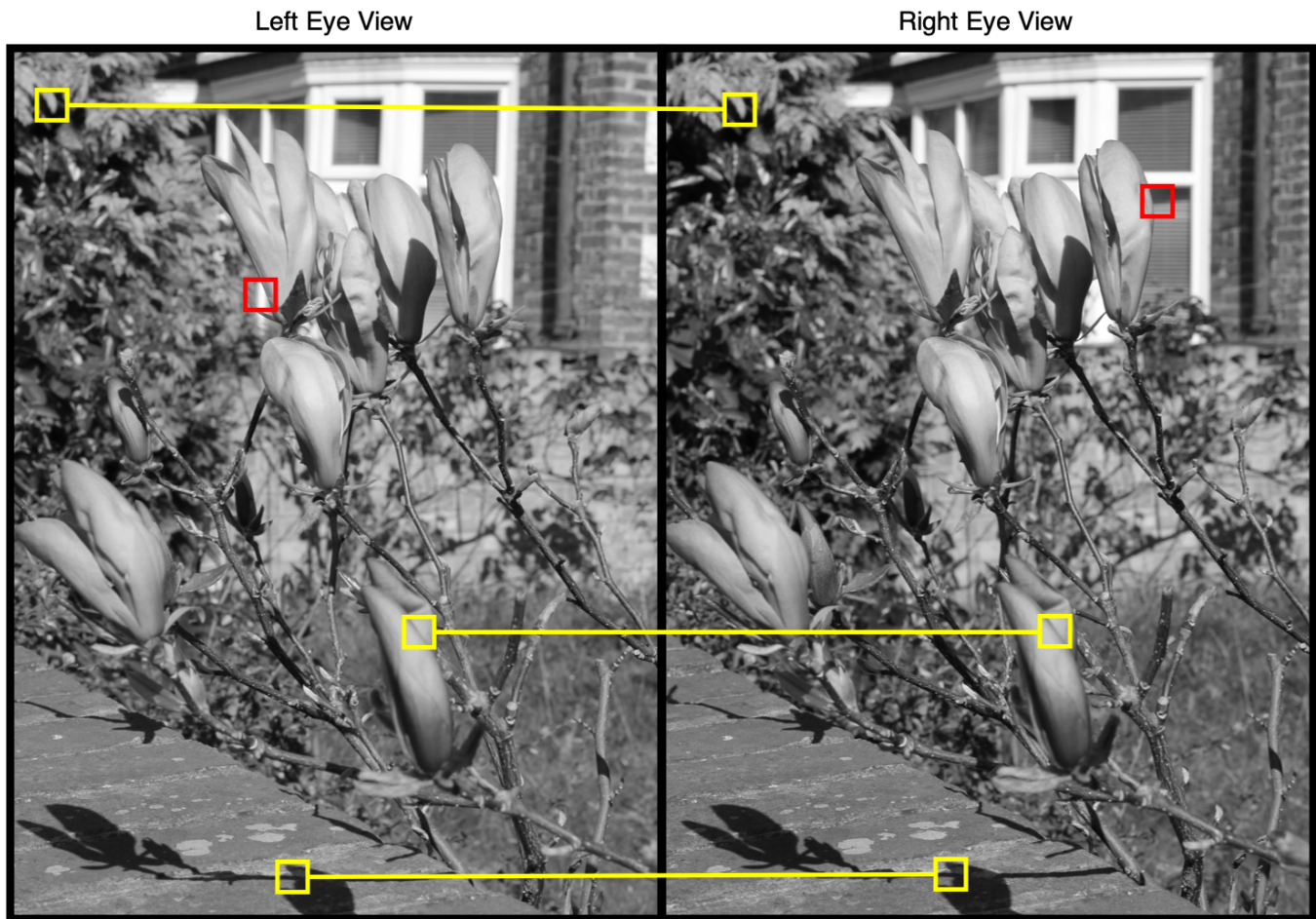
# 1 Introduction

2

3 In animals with binocular vision, the two eyes' images must be fused to obtain a single percept  
4 of the world. Because the eyes capture slightly different views, fusion requires determining which  
5 retinal points in the two eyes correspond to the same physical feature in the world (**Figure 1**,  
6 connected yellow squares). Due to ambiguities in this retinal correspondence, binocular fusion  
7 is computationally demanding. When the visual system fails to fuse the two eyes' images,  
8 observers can experience *unstable* percepts. This experience is often called binocular rivalry,  
9 because percepts can alternate between the content visible to each of the two eyes (Levelt,  
10 1965). To reduce the computational complexity of binocular fusion and obtain a stable percept  
11 of the world, the visual system is known to exploit statistical regularities that are present in  
12 binocular images of natural scenes (Burge, Fowlkes, & Banks, 2010; Burge & Geisler, 2014;  
13 Cooper, Burge, & Banks, 2011; Cooper & Norcia, 2015; Gibaldi, Canessa, & Sabatini, 2017;  
14 Goncalves & Welchman, 2017; Hibbard & Bouzit, 2005; Samonds, Potetz, & Lee, 2012;  
15 Sprague, Cooper, Tošić, & Banks, 2015).

16

17 Not all features in the world, however, create binocularly matched retinal projections. Near depth  
18 edges, parts of a scene are often visible to only one eye (**Figure 1**, isolated red squares). If  
19 these monocular regions are not recognized, the visual system may attempt to fuse regions that  
20 do not binocularly correspond, resulting in unstable percepts (Forte, Peirce, & Lennie, 2002;  
21 Hoffman & Banks, 2010; Shimojo & Nakayama, 1990). The visual system should therefore  
22 exploit statistical regularities in natural images that facilitate detection of monocular features. It  
23 is likely that strong statistical regularities are associated with monocular regions, because of the  
24 geometric constraints imposed by depth edges and the visual properties of the surfaces on either  
25 side of the depth edge (e.g., Anderson, 1994; Gillam & Borsting, 1988; Grove & Ono, 1999;  
26 Nakayama & Shimojo, 1990; Tyler, Christopher & Kontsevich, 1995; see Harris & Wilcox, 2009  
27 for an extensive review). Interestingly, although the statistical properties of binocularly visible  
28 content in natural scenes are under active study (Burge & Geisler, 2014; Chauhan, Masquelier,  
29 Montlibert, & Cottureau, 2018; Goncalves & Welchman, 2017; Hibbard, 2008; Iyer & Burge,  
30 2018; Sprague et al., 2015), the statistical properties that characterize monocularly visible  
31 regions in natural scenes have received scant attention. The current manuscript addresses this  
32 gap.



1

2 **Figure 1.** Binocular and monocular retinal projections during natural vision. A stereo view of a  
3 natural scene is shown, representing the images cast on the left and right retinas of an observer.  
4 To obtain a single percept of the world, these two views must be fused. Examples of binocularly  
5 corresponding features are highlighted with pairs of yellow squares connected by lines. Note  
6 that these features have variable binocular offsets and many similar features are visible to the  
7 two eyes, making fusion computationally challenging. Isolated red squares depict monocularly  
8 visible regions associated with depth edges. (Image courtesy of Deviant Art user aegiandyad,  
9 Creative Commons Attribution 3.0 License).

10

11 Consider three illustrative geometric configurations that can cause monocular regions in 3D  
12 scenes. The first configuration is a “background occlusion”, in which a foreground surface  
13 partially occludes an extended background surface, so that one portion of the background  
14 surface is visible to only one eye and the other portion is visible to both eyes. In this case, the  
15 appearance of the monocular region should be similar to the binocular background (**Figure 2A**).  
16 The second configuration is a “self-occlusion”, in which a foreground surface occludes some  
17 portion of itself, rendering some portion of the foreground surface visible to only one eye. In this  
18 case, the appearance of the monocular region should be similar to the binocular foreground  
19 (**Figure 2B**). The third configuration is a “hidden surface”, in which a foreground surface

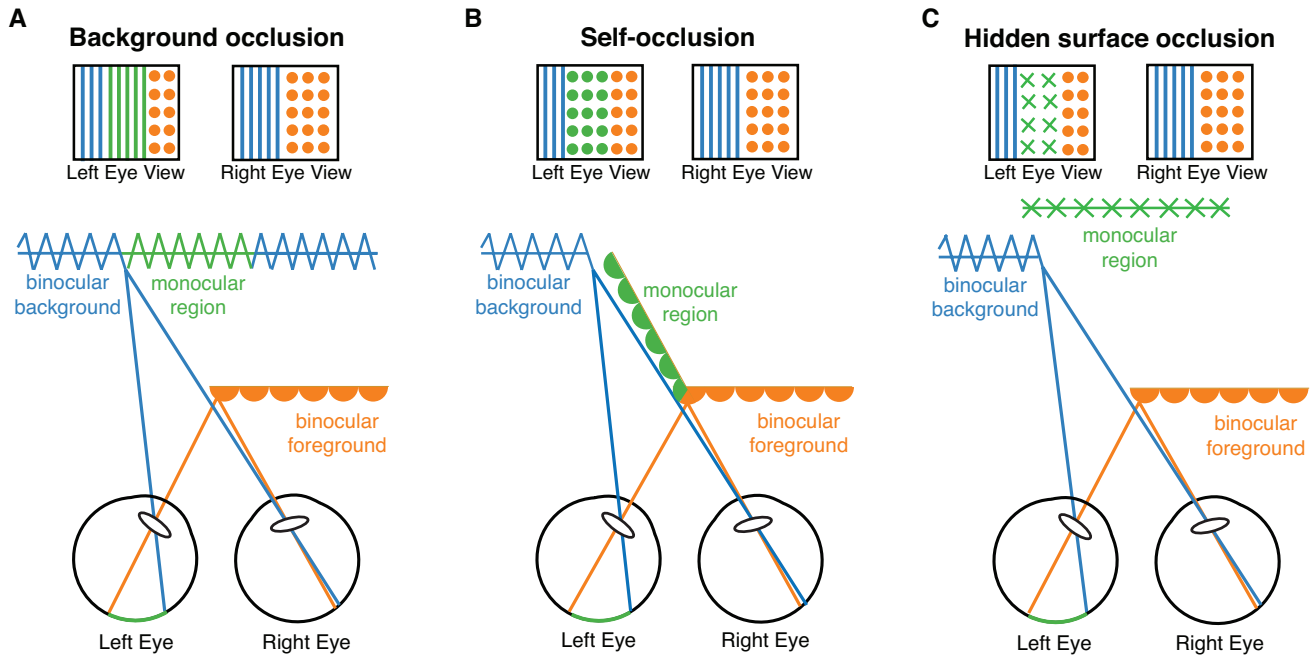
4

1 occludes a surface behind it such that the hidden surface is imaged by only one eye (**Figure**  
2 **2C**). In this last case, the monocularly visible surface in the scene is not continuous with either  
3 the adjacent binocular foreground or the adjacent binocular background; the features in the  
4 monocular image region are thus likely to be dissimilar to those in the image regions  
5 corresponding to the binocularly visible foreground and background. Additionally, monocular  
6 regions may often be comprised of a combination of multiple surfaces (“hybrid occlusions”, see  
7 Appendix **Figure A1**). For example, when the edge of the foreground surface is curved, a single  
8 monocular region can contain some portion of the foreground directly adjacent to some portion  
9 of the background. The configurations shown in **Figure 2** thus represent extremes along a  
10 continuum of possible configurations, which represent useful starting points for geometric and  
11 statistical analyses.

12  
13 If the visual system is adapted to process depth edge configurations that are most likely to be  
14 encountered in natural scenes, then perceptual experiments may provide clues about the  
15 corresponding statistical regularities. For example, if background occlusions are the most likely  
16 cause of monocular regions, visual systems might better process images in which the monocular  
17 region is more similar to the binocularly visible background than to the binocularly visible  
18 foreground. Consistent with this hypothesis, experiments using random dot stereograms of  
19 depth edges have shown that depth processing is faster when the texture of the monocular  
20 region matches the texture of the binocularly visible background (Gillam & Borsting, 1988; Grove  
21 & Ono, 1999). Experiments have also shown that the perceived depth of the monocular region  
22 tends to match the perceived depth of the binocularly visible background (Grove, Ben Sachtler,  
23 & Gillam, 2006; Grove, Gillam, & Ono, 2002). However, not all such studies have produced  
24 consistent results. In the only study to date (that we are aware of) that has used natural images  
25 to probe the processing and perception of monocular regions, judgments about the depth  
26 structure of the scene were found to be faster for self-occlusions than for background occlusions  
27 (Wilcox & Lakra, 2007).

28  
29 In a complementary line of research, computational studies have been used to examine  
30 statistical regularities in monocular regions. For example, simulations of 3D scene geometry  
31 have shown that the probability of a point being monocularly visible varies as a function of  
32 distance in scenes composed of cluttered surfaces (Langer & Mannan, 2012; Langer, Zheng, &

1 Rezvankhah, 2016). However, the image cues that may be available to the visual system for  
2 identifying these monocular regions in typical natural environments have not been examined.  
3 Thus, given the current literature, we do not yet fully understand the visual processing and  
4 perception of monocular regions near depth edges, particularly in natural scenes.



5

6 **Figure 2.** Diagrams of depth edge configurations that lead to different types of monocular  
7 regions. A local area of a scene containing a single depth edge can be subdivided into a  
8 monocularly visible region (green), the adjacent binocularly visible background (blue), and the  
9 adjacent binocularly visible foreground (orange). Patterns above each diagram illustrate the  
10 left- and right-eye views of the scene. Note that in all configurations, the right eye views are the  
11 same, and the left eye views contains a monocular region. If the depth edge is reversed, so the  
12 foreground is on the left, the monocular region will fall in the other eye. Colors are used to  
13 indicate monocular versus binocular content, not actual color differences in the scene. **A.** In a  
14 background occlusion, the left eye sees around the corner of the foreground to expose more of  
15 the background surface. **B.** In a self-occlusion, the foreground occludes itself, and the left eye  
16 sees the side of the foreground object, but the right eye does not. **C.** In a hidden surface  
17 occlusion, there is an object or surface that the foreground hides from the right eye, but the left  
18 eye can see it. Any given depth edge may contain a combination of any of these factors (see  
19 **Figure A1**).

20

21

22 In recent decades, a substantial literature has developed that links the statistics of natural  
23 images and the geometry of natural scenes to the structure and function of the human visual  
24 (see (Geisler, 2008) for an extensive review). Here, we investigate the hypothesis that depth  
25 edges in natural scenes contain statistical regularities that the visual system capitalizes upon to  
26 facilitate perceptual stability. We sampled and analyzed a large set of stereoscopic image

1 patches containing depth edges from a natural image database with co-registered distance  
2 measurements (Burge, McCann, & Geisler, 2016). In two perceptual experiments, human  
3 observers viewed the sampled patches and rated the visual stability of the depicted scene near  
4 each depth edge. The results showed that perception was relatively more stable when the  
5 properties associated with the depth edge were statistically more likely. These conclusions were  
6 supported by a computational study of simulated 3D environments. Together, our results  
7 suggest that the visual system can exploit regularities in natural image and scene statistics to  
8 achieve perceptual stability at depth edges.

9

## 10 **Methods**

11 All aspects of the natural scene analyses, perceptual experiments, and modeling were  
12 performed using Matlab (Mathworks, Inc.). In the perceptual experiments, stimulus presentation  
13 was controlled using the PsychToolbox-3 extensions (Brainard, 1997; Kleiner, Brainard, & Pelli,  
14 2007; Pelli, 1997).

15

### 16 ***Stimulus sampling***

17 Stimuli for the scene statistics analyses and for the perceptual experiments were sampled from  
18 the natural scene dataset described in Burge et al., 2016. This dataset is composed of ~100  
19 stereo-images of natural scenes with precisely co-registered distance data. The stereo-images  
20 were photographed with an inter-camera separation of 6.5 cm, similar to a typical human  
21 interocular separation. The goal of the stimulus sampling was to select a set of stereo-image  
22 patches containing depth edges with relatively large monocular regions that could be (a) used  
23 to investigate the visual content and surrounding binocular context in a natural scene statistics  
24 analysis, and (b) used as stimuli in perceptual experiments.

25

26 To determine whether each pixel in a scene from the dataset was monocularly visible, we  
27 computed the ground truth horizontal disparity gradients at each image pixel directly from the  
28 distance maps (Iyer & Burge, 2018). It has been shown mathematically that if the disparity  
29 gradients—the change in disparity divided by the change in visual angle—all exceed a value of  
30 2.0, the pixel is visible to only one eye (Bülthoff, Fahle, & Wegmann, 1991). Visual inspection of  
31 the results confirmed that this procedure was accurate on the current range maps. Stereo-image  
32 patches ( $3.8^\circ \times 1.5^\circ$ , 195x78 pixels) centered on monocularly visible points at depth edges were

1 sampled randomly. To ensure uniqueness, we selected patches that did not overlap with each  
2 other by more than  $0.2^\circ$ .

3

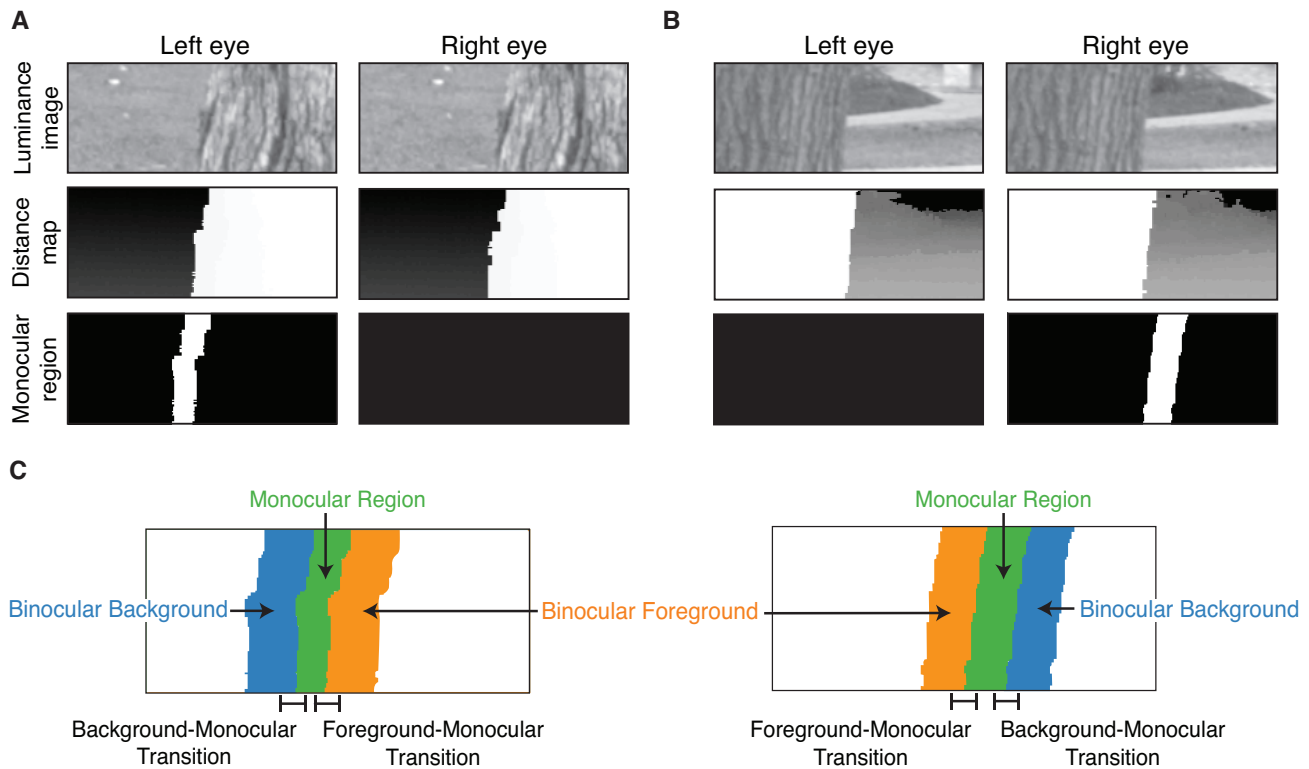
4 We vetted the sampled patches using an automated process followed by additional manual  
5 processing. To make sure the monocular region had sufficient area for analysis, the central row  
6 in the patch was constrained to contain a set of monocularly visible points between  $0.3^\circ$  and  $0.7^\circ$   
7 wide. In all other rows, the monocularly visible region was constrained to be at least  $0.1^\circ$  wide.  
8 To make sure the binocular regions were sufficiently large to analyze, the monocular region was  
9 constrained to be neighbored by at least  $0.5^\circ$  of relatively contiguous binocular background and  
10 foreground regions (see below for details). These criteria were chosen in order to obtain patches  
11 with monocular regions that were large enough to facilitate an informative analysis of visual  
12 patterns and binocular context (foreground/background) in natural scenes, but small enough that  
13 the neighboring binocular disparities would cause minimal diplopia during the perceptual  
14 experiments. Additionally, some patches contained local lowlights or highlights that resulted in  
15 luminance clipping. If more than 25% of the pixels either had no light recorded or were saturated,  
16 the patch was discarded. This sampling process resulted in 215 patches for analysis.

17

18 **Figure 3A & B** show two examples from the final patch set. Each example consists of a stereo-  
19 image patch and co-registered distance values at each pixel (**top and middle rows**), as well as  
20 binary masks labeling each pixel that was monocularly visible (**bottom rows**). In these binary  
21 masks, we removed minor artifacts that resulted from the range scanner as well as small regions  
22 of adjacent monocular pixels that intruded on the background. For example, if a tree or bush  
23 was visible in the background adjacent to a depth edge, there might be small speckles of  
24 monocular pixels associated with small partial occlusions. These were removed by manually  
25 editing the binary mask.

26





1

2 **Figure 3.** Examples of stereoscopic patches containing depth edges. **A.** Left and right eye views  
3 are shown from one of the sampled depth edges. Rows contain the luminance images (top),  
4 distance maps (middle) and monocular region binary masks (bottom). Luminance images have  
5 been histogram-adjusted for visibility. Grayscale values in the distance maps represent distance  
6 from the camera in meters, scaled to fill the full range from white (near) to black (far). Across all  
7 patches used, the distance from the cameras to the foreground region ranged from 3.1-11.1 m  
8 (average 5.4 m) and the distance from the cameras to the background ranged from 4.3-127.5 m  
9 (average 23.8). The monocular region mask indicates a single region in the left eye, consistent  
10 with the left eye seeing more of the background to the left of the tree. **B.** Same as A, but in this  
11 case the monocular region appears in the right eye. **C.** The five spatial regions of interest used  
12 for analysis are illustrated for both patches.

13

### 14 **Natural Scene Statistics**

15 To organize our analysis of the statistics of depth edges in natural scenes, we defined five  
16 regions of interest (**Figure 3C**). These regions were the monocular region (green), the adjacent  
17 binocular foreground (orange), the adjacent binocular background (blue), and two transition  
18 regions from the binocular background to the monocular region and from the binocular  
19 foreground to the monocular region (black lines). The binocular regions consisted of the  $0.5^\circ$   
20 wide image areas neighboring the monocular region to the left or right. The transition regions  
21 contained  $0.4^\circ$  wide image areas centered on the transition between the monocular region and  
22 the binocular background or foreground regions. The forthcoming results are similar for other  
23 similar region sizes.

1

2 From the distance maps, we calculated the mean distance of the surfaces in the monocular  
3 region, the adjacent binocular foreground region, and the adjacent binocular background region.  
4 From the images, we computed the mean luminance and the mean contrast (i.e. square root of  
5 the mean squared luminance deviation) within each of these regions as well. In the binocular-  
6 monocular transition regions, we focused on analyzing the changes in visual appearance  
7 associated with transitions between surfaces. For example, a large horizontal luminance  
8 derivative would reflect a strong vertical edge, suggesting a transition between two surfaces with  
9 different patterns. To quantify the strength of the vertical luminance edge between the monocular  
10 region and adjacent binocular regions, we used 5-tap derivative filters to compute the mean  
11 magnitude of the horizontal luminance derivative (Farid & Simoncelli, 1997). As a control  
12 measure, we repeated this analysis for vertical derivatives, which are not expected to be  
13 particularly strong at horizontal surface transitions.

14

15 Prior to statistical analyses, all of the image-based measurements described above were  
16 normalized. Specifically, the mean luminance within each region and the mean magnitude of  
17 luminance derivatives within each transition region were normalized by the mean luminance of  
18 the entire patch. The mean contrast within each region was normalized by the mean contrast of  
19 the entire patch. Two-tailed Wilcoxon signed rank tests were used to examine statistically  
20 significant differences in the visual properties of different regions. Non-parametric effect sizes  
21 for these tests were calculated as  $z/\sqrt{n}$ , where  $z$  refers to the  $z$  score of the Wilcoxon test  
22 statistic and  $n$  refers to the number of differences used to calculate the test statistic (Fritz, Morris,  
23 & Richler, 2012).

24

25 It is important to note that the natural scene dataset used here does not contain objects that are  
26 closer than 3 m. This limitation of the dataset should be considered when interpreting the results  
27 of our analysis because prior work suggests that the prevalence of monocularly visible regions  
28 depends on the distribution of object distances in a scene (Langer & Mannan, 2012). That said,  
29 we later report the results of a geometric simulation that parametrically examined how object  
30 distance affects the geometric causes of monocular regions (see Results). The findings from the  
31 simulation suggest that the main conclusions of this natural scene analysis are likely to hold not  
32 just for surfaces that are far away, but also for surfaces that are up close.

1

## 2 ***Perceptual Experiments***

3 Two perceptual experiments were conducted to examine how the visual properties of natural  
4 depth edges affect perceptual stability. Experiment 1 was conducted in an exploratory manner  
5 and used to develop a set of hypotheses and appropriate statistical tests. Experiment 2 tested  
6 these hypotheses with a second set of observers, a larger, more diverse set of stimuli and a  
7 higher dynamic range display. We first describe the methods in the two experiments that were  
8 the same, and then indicate the differences between them.

9

### 10 **General Methods.**

11 *Observers.* All observers had normal or corrected-to-normal visual acuity and normal  
12 stereoacuity, as determined by the Randot stereo test (detection of 70 arcsecs or less).  
13 Observers gave informed consent prior to starting the experiment. The procedures were  
14 approved by the institutional review board at the University of California, Berkeley and were  
15 consistent with the Declaration of Helsinki. Experiment 1 had 10 observers (5 female, mean age  
16 = 31.0 ± 2.6 years) and Experiment 2 had 15 observers (9 female, mean age = 27.7 ± 5.4 years).

17

18 *Stimulus Presentation.* Patches from the dataset described above were presented to observers  
19 stereoscopically on gamma-linearized displays. Observers were positioned via a forehead and  
20 chin rest at a viewing distance of approximately 60cm (1 pixel subtended ~0.02-0.03°). The  
21 stimulus was designed to simulate viewing a portion of a natural scene through a window (**Figure**  
22 **4**). Specifically, stereoscopic scene patches were viewed through a virtual window that was  
23 specified by disparity to be at the display distance. The window subtended the same visual area  
24 as each patch (3.75°x1.5°). Using recently developed techniques described in Iyer & Burge  
25 (2018), left and right eye image patches were cropped at corresponding points such that the  
26 depth edge had five arcminutes of uncrossed disparity with respect to the screen. This procedure  
27 ensured that all depicted surfaces in the scene were stereoscopically defined to be behind the  
28 plane of the display and the frame of the window.

29

30 On each trial, observers viewed a patch for 5 seconds through the virtual window, and were then  
31 asked on a separate response screen to evaluate the visual stability of the patch using a  
32 continuous slider response. To explain the task, they were shown an example of a rivalrous

1 stimulus and given descriptions of types of visual instability (e.g., shimmering or alternating  
2 between two states). Each response was recorded on a 100-point scale, with 100 indicating the  
3 most unstable. We refer to these responses as *instability ratings*. Pilot testing confirmed that  
4 even if there was local instability at the depth edge, observers could fuse these stimuli as a  
5 whole and experience depth percepts of the foreground and background (White & Burge, 2019).  
6 For the current study we chose a visual stability rating task, because previous work suggests  
7 that stability is a salient aspect of perception when binocular fusion fails (Hoffman & Banks,  
8 2010; Shimojo & Nakayama, 1990). All observers first completed an initial practice session in  
9 which they observed and rated example patches. For Experiment 2, this practice session  
10 included all stimuli in order to facilitate better response reliability during the main task. Data from  
11 the practice session were not included in analysis. Following the practice session, observers  
12 completed two repetitions per patch. Patches were not repeated until all stimuli had been seen  
13 once. We calculated the test-retest reliability via the squared Pearson correlation coefficients ( $r^2$ )  
14 between each observer's ratings from the 1<sup>st</sup> and 2<sup>nd</sup> presentations of each patch. Observers  
15 with  $r^2$  values below 0.1 were excluded from further analysis (3 observers in Experiment 1, and  
16 none in Experiment 2).

17

## 18 **Experiment 1**

19 *Stimuli.* We selected a subset ( $n = 92$ ) of the patches from the natural scene statistics analysis  
20 for an exploratory study. To reduce the number of features that varied across different patches,  
21 the mean luminance and contrast of each patch were fixed prior to displaying them to the  
22 observers. The mean luminance of each patch was set to 30% of the maximum luminance output  
23 of the display. The normalized root-mean-squared (RMS) contrast of each patch when presented  
24 onscreen was fixed to 0.25. Note that the luminance and contrast within and across the regions  
25 of interest still varied in the patches – fixing the global mean luminance and global mean contrast  
26 simply allowed us to concentrate our analyses on the impact of local image properties. Stimuli  
27 were presented stereoscopically on a liquid crystal display (LCD) (ViewSonic V3D231, 50.9 x  
28 28.6cm, 1920 x 1080 pixels), which supports passive stereoscopic presentation via polarizing  
29 filters. Observers wore polarizing glasses so that stimuli could be viewed dichoptically  
30 (specifically, alternating rows of pixels on the screen were visible to each eye). In this  
31 experiment, the viewing window was surrounded by a 1/f noise field. The maximum luminance

1 through the filters was approximately  $63 \text{ cd/m}^2$ , and stimulus settings were selected to eliminate  
2 any visible crosstalk.

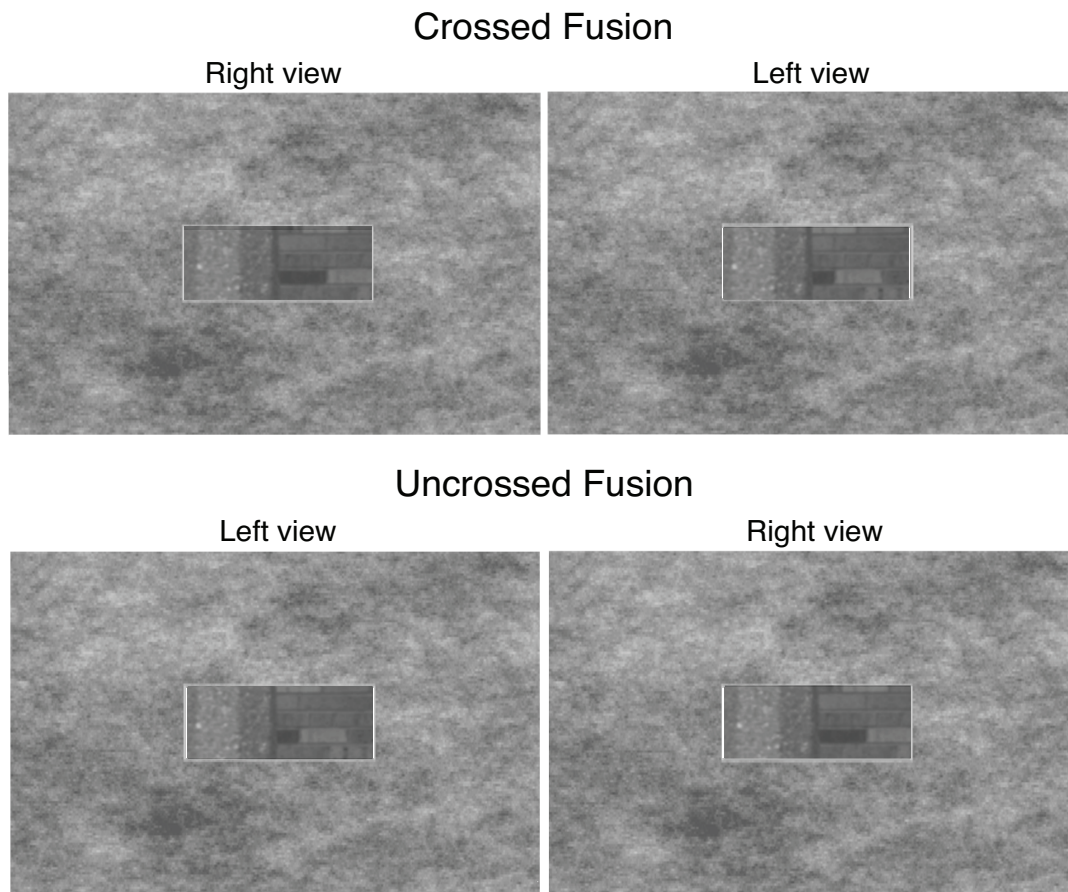
3

#### 4 **Experiment 2**

5 *Stimuli.* In this second experiment, we used a larger set of patches to obtain a more diverse set  
6 of stimuli ( $n = 125$ ). In addition, we did not fix the mean luminance or mean contrast of the stimuli,  
7 as we did in Experiment 1. By allowing these values to vary, we could test whether the findings  
8 from Experiment 1 generalized to more diverse stimuli (i.e., they are not specific to the luminance  
9 and contrast values selected for that experiment). The results were highly similar across  
10 experiments. Stimuli were presented stereoscopically on a mirror haploscope with two LCDs to  
11 provide a higher luminance range (LG 32UD99-W, 69.6 x 39.3 cm, 1920 x 1080 pixels, maximum  
12 luminance of approximately  $330 \text{ cd/m}^2$ ). The stimulus was adjusted so that the vergence demand  
13 of the virtual window was equal to the monitor distance. The virtual window was surrounded by  
14 a  $1/f$  noise field, as in Experiment 1, but the texture was removed from the area immediately  
15 around the window in order to reduce visual clutter.

16

17



1

2

3 **Figure 4.** Example stimulus. Left and right eye views are shown (oriented for crossed fusion in  
4 the upper row, and uncrossed fusion in the lower row). Each stereoscopic patch was placed in  
5 a window surrounded by  $1/f$  noise, on an uncrossed disparity pedestal so that all content  
6 appeared behind the window frame. Stimulus is shown as presented in Experiment 1, and the  
7 surrounding background is cropped for visibility. Images have been histogram adjusted for  
8 visibility.

9

10

### 11 **Modeling instability ratings**

12 We used the same statistical analysis in both experiments. First, we defined a set of image  
13 features, the values of which varied across the stimuli. Then, we defined a set of linear mixed  
14 models that used these features as predictors for patch instability ratings. In each model,  
15 different observers were modeled with random intercepts, and the usefulness of different  
16 features for predicting the responses was examined (see Results for details). Initial analyses of  
17 Experiment 1 were conducted in an exploratory manner based on the results of the scene  
18 statistics analysis. Based on these exploratory analyses, we hypothesized that the strength of  
19 the vertical edges at the binocular-monocular transition regions modulated the perceptual

1 stability of the depth edges. We then conducted a second experiment to test this hypothesis on  
2 a larger independent sample of perceptual data.

3  
4 In a follow up analysis, we asked whether higher contrast in the patches was associated with  
5 higher perceptual instability, based on predictions from previous work (L. Liu, Tyler, & Schor,  
6 1992). For each observer, we converted instability ratings to z-scores by subtracting the mean  
7 response and dividing by the response standard deviation. We calculated Pearson correlation  
8 coefficients between the contrast in each region and the mean z-score for each patch. For  
9 correlation values, 95% confidence intervals are calculated via bootstrapping.

10

## 11 **Results**

### 12 ***Natural Scene Statistics***

13 We analyzed the statistical regularities of scene and image properties in the immediate vicinity  
14 of depth edges in natural scenes, with an emphasis on the transitions between monocular and  
15 binocular regions. If robust regularities are present, the visual system might exploit these  
16 regularities to determine which regions of the scene are monocularly and binocularly visible.

17

18 First, we analyzed the difference between the mean distance to the surfaces in the monocularly  
19 visible region of the scene and the mean distances to the surfaces in the adjacent binocularly  
20 visible background and foreground regions. This analysis was possible because each image in  
21 the dataset had laser-based distance measurements co-registered to each image pixel (Burge  
22 et al, 2016). When a monocular region primarily consists of the background occluded by the  
23 foreground (**Figure 2A**), the distance of monocularly visible points from the camera, on average,  
24 should be more similar to the distance of the binocular background than the distance of the  
25 binocular foreground. Conversely, when a monocular region primarily consists of a self-occlusion  
26 (**Figure 2B**), the distance of the monocularly visible points, on average, should be more similar  
27 to the binocular foreground. We found that the distance of surfaces in the monocular regions  
28 tended to be substantially more similar to the adjacent binocular background than the  
29 foreground. **Figure 5A** shows histograms of the distance between each monocular region and  
30 the adjacent binocular foreground (orange) and the adjacent binocular background (blue) – that  
31 is, the mean distance in the monocular region minus the mean distance in the adjacent binocular  
32 regions. A value of zero indicates that points in the monocular region had the same average

1 distance as points in the relevant binocular region. Positive and negative values indicate that the  
2 distance of points in the monocular region was farther and closer, respectively, than the distance  
3 of points in the binocular region. The median absolute distance of points in the monocular region  
4 from the binocular foreground was 9.59 m, while the median absolute distance of points in the  
5 monocular region from the binocular background was 0.10 m. The absolute distance of points in  
6 the monocular region to points in the binocular background was significantly smaller than the  
7 distances to points in the binocular foreground ( $p \ll 0.001$ ,  $z = 12.71$ , *effect size* = 0.87). This  
8 result suggests that, in this dataset, the vast majority of monocular regions were caused primarily  
9 by background occlusions.

10

11 We reasoned that if a monocular region associated with a depth edge is caused by a background  
12 occlusion (i.e., is an extension of the neighboring binocular background), then there should be  
13 reliable information about this fact in the luminance pattern within the *transition regions* (see  
14 **Figure 3C**) (Y. Liu, Cormack, & Bovik, 2009). Specifically, when a monocular region is an  
15 extension of the neighboring binocular background there should be a relatively strong vertical  
16 luminance edge at the foreground-monocular transition region, because this will coincide with  
17 the depth edge. Relatedly, there should be a relatively weak vertical edge at the background-  
18 monocular transition region, because the binocular background and the monocular region are  
19 likely to contain the same surfaces. On the other hand, if the foreground surface occludes itself—  
20 a self-occlusion (**Figure 2B**)—there will be no depth edge in the foreground-monocular transition  
21 region, and there should therefore be a relatively weak vertical luminance edge in the  
22 foreground-monocular transition region. A stronger luminance edge is thus expected in the  
23 background-monocular transition region (**Figure 2B**). Note that these inferences are based on  
24 the assumption that background and foreground surfaces have visually distinct textures, and  
25 that the statistics of these textures are relatively stationary within a surface. These assumptions  
26 have not been previously tested empirically.

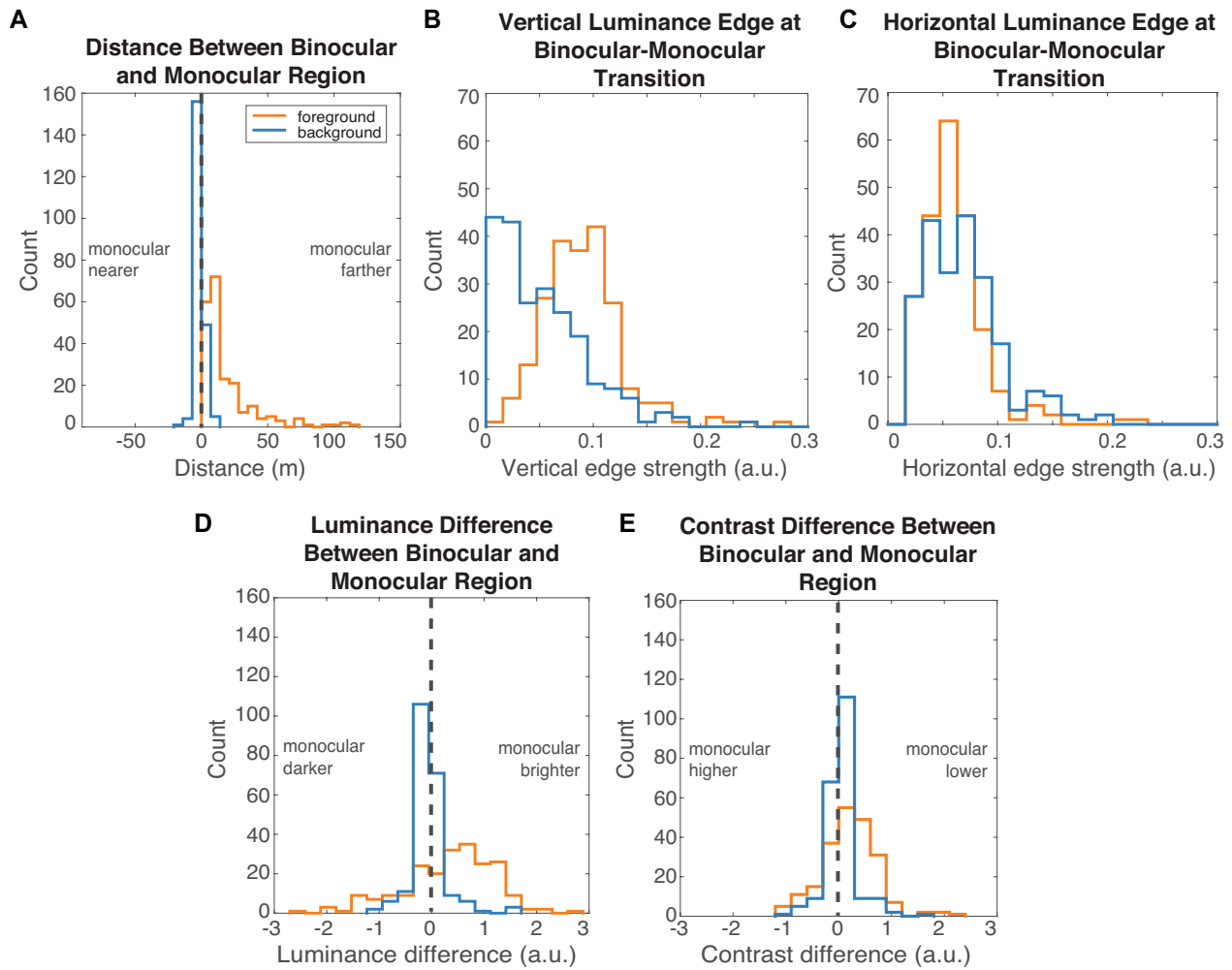
27

28 The current data suggest that these assumptions are sufficiently accurate for natural image  
29 patches to meet these expectations. The vertical luminance edge strength (calculated as  
30 described in the Methods) in the foreground-monocular transition region tended to be higher  
31 (0.09) than at the vertical luminance edge strength in the background-monocular transition (0.05)  
32 (**Figure 5B**;  $p \ll 0.001$ ,  $z = 10.7$ , *effect size* = 0.73). If this pattern is tied to the surface in the



1 monocular region, we hypothesized that this pattern should be *specific* to vertical luminance  
2 edges, because edges that are vertical indicate a horizontal transition between surfaces. To test  
3 this hypothesis, we repeated this analysis for horizontal luminance edge strength. With  
4 horizontal luminance edges, we found a small but statistically significant difference in the  
5 opposite direction. The horizontal edge strength at the foreground-monocular transition was  
6 slightly but significantly lower (0.057) than the horizontal edge strength at the background-  
7 monocular transition (0.064) (**Figure 5C**;  $p \ll 0.001$ ,  $z = 3.43$ , *effect size* = 0.23). We speculate  
8 that this effect occurs because the abrupt transition between surfaces at the foreground-  
9 monocular transition precludes any strong horizontal continuity. Importantly, with these  
10 analyses, we cannot rule out the possibility that some of these monocular regions were caused  
11 by hybrid occlusions: a background occlusion combined with a thin self-occlusion at the curved  
12 edge of a foreground object, for example. Such hybrid occlusions would still provide a vertical  
13 luminance edge within the defined transition region. In the forthcoming geometric simulation, we  
14 examine the prevalence of hybrid occlusions directly.

15  
16 To further test the conclusions of the transition analysis (i.e., that monocular regions tend to  
17 primarily belong to the binocular background, and therefore share similar visual features), we  
18 examined the luminance and contrast within the monocular regions relative to the binocular  
19 regions (**Figure 5D and E**). Consistent with the previous analysis, the results showed that these  
20 visual properties were more similar to the binocular background than to the binocular foreground.  
21 Specifically, the median absolute luminance difference from the background (0.08) was  
22 approximately ten times lower than the median difference from the foreground (0.74) ( $p \ll 0.001$ ,  
23  $z = 11.6$ , *effect size* = 0.79). Recall that luminance is represented in arbitrary linear units. The  
24 median absolute contrast difference between the monocular region and the background (0.10)  
25 was more than three times lower than the median absolute contrast difference between the  
26 monocular region and the foreground (0.36) ( $p \ll 0.001$ ,  $z = 8.95$ , *effect size* = 0.61). Indeed,  
27 the strength of the vertical luminance edge between two regions was highly correlated with both  
28 the luminance difference ( $r = 0.57$ , 95% confidence interval 0.49-0.64,  $p \ll 0.001$ ) and the  
29 contrast difference ( $r = 0.51$ , 95% confidence interval 0.42-0.61,  $p \ll 0.001$ ). We speculate that  
30 these measures are highly correlated because they all reflect the same underlying image  
31 structure in the neighborhood of depth edges.



1

2 **Figure 5.** Properties of monocular regions and the surrounding binocular context. Histograms  
3 show the frequency of: (A) the distances between binocular foregrounds/backgrounds and  
4 monocular regions within a patch, (B) the vertical luminance edge strength at the transitions from  
5 binocular foregrounds/backgrounds to monocular regions, (C) the horizontal luminance edge  
6 strength at the transitions from binocular foregrounds/backgrounds to monocular regions, (D)  
7 the differences in luminance between binocular foregrounds/backgrounds and monocular  
8 regions, and (E) the differences in contrast between binocular foregrounds/backgrounds and  
9 monocular regions. Distance was calculated in units of meters. Vertical and horizontal edge  
10 strengths represent the normalized magnitude of the derivative filters in the respective transition  
11 regions (i.e., foreground-monocular and background-monocular). Luminance is represented in  
12 arbitrary units, and contrast is represented as normalized RMS pixel luminance. A.u. indicates  
13 arbitrary units.

14

15 Collectively, these natural scene analyses suggest that monocularly visible regions in natural  
16 scenes are associated with robust statistical regularities. The results suggest that monocular  
17 regions in natural scenes are more likely to be composed primarily of an occluded extension of  
18 the binocularly visible background than an occluded portion of the binocularly visible foreground,  
19 at least within scenes that are similar to those in this particular dataset (e.g., relatively far object

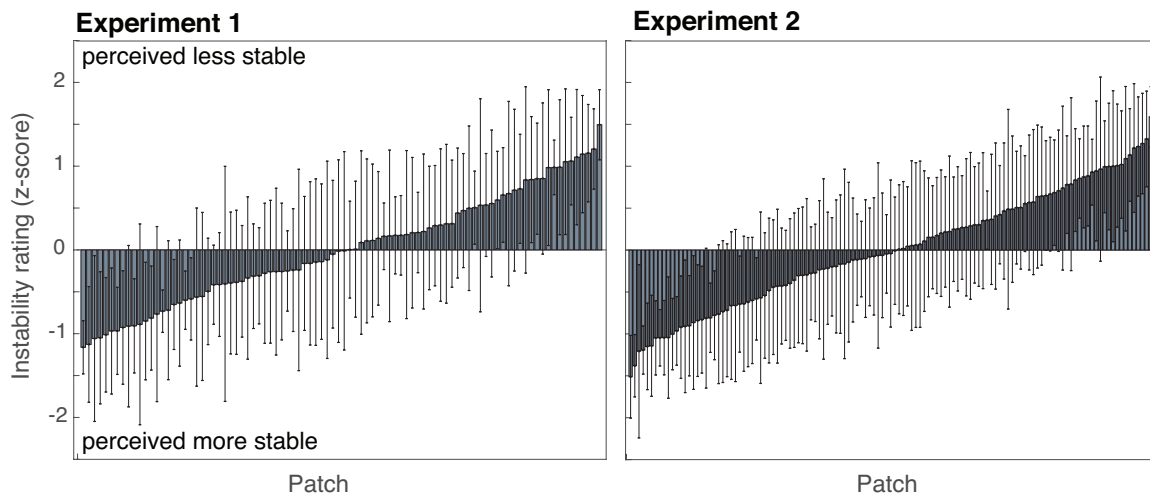
1 distances, relatively sparse object density). According, the monocular regions we analyzed  
2 tended to visually similar to the adjacent binocular background. In particular, the transition from  
3 the foreground to the monocular region was associated with a strong vertical luminance edge,  
4 whereas the transition from the monocular region to the background was not. If the visual system  
5 is adapted to these statistical regularities in the natural environment, the visual similarity of the  
6 monocular region with the background, and perhaps the visual dissimilarity of the monocular  
7 region with the foreground, should be associated with increased perceptual stability of depth  
8 edges in natural scenes. We similarly predict that greater perceptual instability will result from  
9 stimuli in which these statistical regularities in the luminance images are violated. To test these  
10 predictions, we performed a pair of perceptual experiments.

11

## 12 ***Perceptual Experiments***

13 As a reminder, we conducted two perceptual experiments in which observers viewed a subset  
14 of the natural stereo-image patches from the previous analysis and rated the perceived instability  
15 of the depth edge. The perceptual experiments were designed to determine whether the  
16 statistical regularities associated with depth edges in natural scenes predict perceptual  
17 instability. Before examining this question, we first established the consistency of ratings within  
18 and across observers (recall that each observer rated each patch twice). The mean correlation  
19 of the observer ratings from the first and second stimulus presentations was 0.45 in Experiment  
20 1 and 0.61 in Experiment 2, suggesting reasonable consistency within observers. To obtain  
21 mean instability ratings for each patch, we thus averaged the first and second ratings of each  
22 observer. **Figure 6** shows the mean and standard deviation of these z-scored ratings across  
23 observers for each patch, ranked from highest to lowest. Across patches, the perceived stability  
24 varied considerably. The standard deviation of instability ratings across observers was relatively  
25 consistent across patches and was also relatively small compared to the change in stability  
26 across the patch ensemble. Together, these results suggest that meaningful comparisons of  
27 perceived stability can be made across this stimulus ensemble.

28



1

2 **Figure 6.** Instability ratings for each natural stereoscopic patch stimulus in both Experiments.  
3 Each bar represents the mean z-score for a patch across observers, ordered from lowest to  
4 highest. Error bars represent the standard deviation of the mean.

5

6 Now, we turn to an analysis of the factors that may contribute to perceptual instability. Before  
7 examining the impact of image-based properties on perceptual instability, we first determined  
8 how much of the variation in perceived instability could be attributed to the relative disparity  
9 between the foreground and background surfaces. Depth edges with a large binocular disparity  
10 between the foreground and background may tend to appear unstable when they exceed  
11 Panum's fusional area, for the simple reason that they are difficult to fuse. With large disparities,  
12 the background is likely to appear diplopic when the foreground is fused and vice versa. In the  
13 stimulus set, the magnitude of the depth steps across the edges corresponded to relative  
14 disparities ranging from 0.3-0.8°. We found that relative disparity was significantly correlated  
15 with the mean instability ratings in both experiments (Experiment 1:  $r = 0.54$ , 95% confidence  
16 interval 0.32-0.70,  $p \ll 0.001$ ; Experiment 2:  $r = 0.59$ , 95% confidence interval 0.47-0.70,  $p \ll$   
17 0.001), accounting for 29% and 34% of the total variance in mean instability ratings, respectively.  
18 Hence, as expected, larger relative disparities are associated with more perceptual instability.

19

20 To model the effects of the image-based properties on instability ratings, we fit the raw response  
21 data with a set of mixed linear regression models. Based on the analyses of natural scene  
22 statistics, we formulated two nested models: the *Disparity-Only model* and the *Edge model*. The  
23 Disparity-Only model included the magnitude of relative disparity as the only predictor. The Edge  
24 model included the vertical luminance edge strength in each of the two binocular-monocular  
25 transition regions as two additional predictors. We chose these two predictors because the

1 natural scene statistics analysis suggested that a strong vertical luminance edge at the  
 2 foreground-monocular transition is likely in natural depth edges. We thus hypothesized that  
 3 strong foreground-monocular vertical edges should be associated with higher perceptual  
 4 stability, but that background-monocular vertical edges should not.

5

6 **Table 1.** Regression model results. For each experiment, results are shown for the two nested  
 7 models: the Disparity-Only model and the Edge model. Confidence intervals (95%) on each of  
 8 the coefficient estimates are indicated in parentheses. T-statistics were used to assess whether  
 9 these estimates were significantly different from zero. Significance values less than 0.05 are  
 10 bolded. Random intercepts were fit per observer in each experiment, with a standard deviation  
 11 of 9.1 and 8.0 in Experiments 1 and 2, respectively.

12

Experiment 1					
	Fixed Effects	Coefficient estimate	T statistic	p value	r <sup>2</sup>
<b>Disparity-Only Model</b>	Intercept	10.9 (0.7,21.0)	2.1	<b>0.04</b>	0.23
	Disparity range	73.3 (58.2,88.5)	9.5	<b>&lt;&lt;0.001</b>	
<b>Edge Model</b>	Intercept	10.0 (-1.0,21.1)	1.8	0.08	0.30
	Disparity range	78.8 (64.3,93.3)	10.7	<b>&lt;&lt;0.001</b>	
	Background-monocular transition	560.4 (423.5,697.2)	8.0	<b>&lt;&lt;0.001</b>	
	Foreground-monocular transition	-334.6 (-490.4,-178.8)	-4.2	<b>&lt;&lt;0.001</b>	
Experiment 2					
	Fixed Effects	Coefficient estimate	T statistic	p value	r <sup>2</sup>
<b>Disparity-Only Model</b>	Intercept	11.8 (6.3,17.4)	4.2	<b>&lt;&lt;0.001</b>	0.27
	Disparity range	77.7 (70.0,85.4)	19.8	<b>&lt;&lt;0.001</b>	
<b>Edge Model</b>	Intercept	5.9 (0.3,11.6)	2.1	<b>0.04</b>	0.35
	Disparity range	81.8 (74.4,89.1)	21.9	<b>&lt;&lt;0.001</b>	
	Background-monocular transition	175.9 (148.5,203.4)	12.6	<b>&lt;&lt;0.001</b>	
	Foreground-monocular transition	-55.8 (-83.1,-28.5)	-4.0	<b>&lt;&lt;0.001</b>	

13

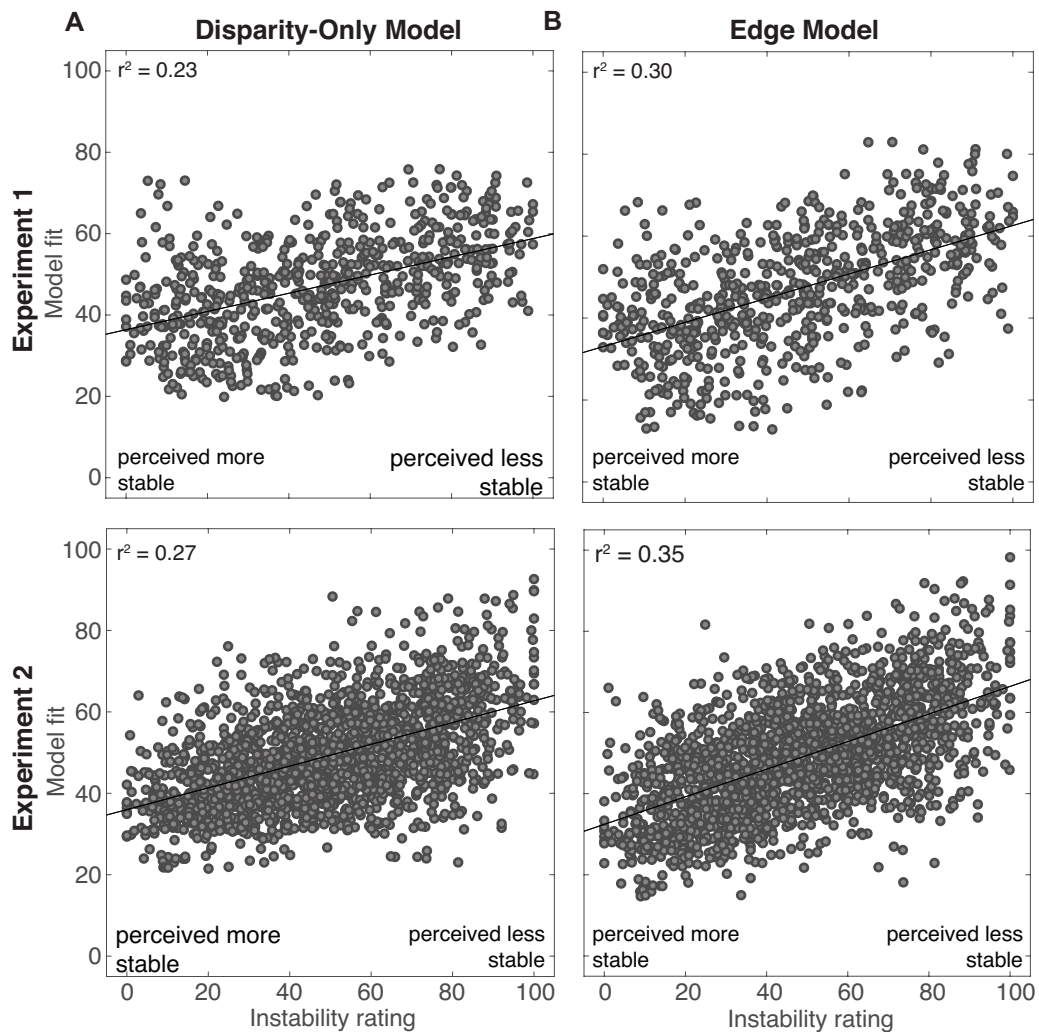
14 The results suggest that higher perceptual stability is associated with image patterns that are  
 15 more likely given the natural image and scene statistics. **Table 1** summarizes these results for  
 16 each experiment. The observed versus the fitted ratings are plotted in **Figure 7**. The Disparity-  
 17 Only model and the Edge model both indicate a significant effect of disparity magnitude. In the  
 18 Edge Model, vertical luminance edge strength in the binocular-monocular transition region was  
 19 also a significant predictor of the ratings: stronger vertical edges in the foreground-monocular  
 20 transition region were associated with greater perceptual stability (lower ratings) and stronger

1 vertical edges in the background-monocular transition region were associated with greater  
2 perceptual *instability* (higher ratings). A vertical edge at an atypical location relative to the  
3 monocular region may cause the visual system to misinterpret monocular regions as being  
4 binocular, and the increased instability may thus result because the visual system tries and fails  
5 to fuse monocular regions with binocular regions.

6 A likelihood-ratio test indicated that the Edge Model was a significantly better fit to the data  
7 (Experiment 1:  $LR = 61.61$ ,  $\chi^2 df = 2$ ,  $p \ll 0.001$ ; Experiment 2:  $LR = 194.52$ ,  $\chi^2 df = 2$ ,  $p \ll$   
8  $0.001$ ). This result was supported by a comparison of each model's Akaike Information Criterion  
9 (AIC), which accounts for the increased number of parameters in the Edge Model (Experiment  
10 1: Disparity-Only Model = 5922.9, Edge Model = 5892.1; Experiment 2: Disparity-Only Model =  
11 16649, Edge Model = 16458). We also computed the proportion of the response variance  
12 explained by the Edge Model as compared to the Disparity-Only Model. The Edge Model  
13 explained 30% and 35% of the response variance in Experiments 1 and 2, while the Disparity-  
14 Only model explained 23% and 27%. These percentages should be interpreted relative to the  
15 expected amount of explainable variance in the modeled data, which is limited by the variability  
16 in repeated presentations of the same stimulus to the same observer. Using the method  
17 described in Hsu, Borst, & Theunissen (2004), we estimated that the explainable variance in the  
18 mean responses in the first experiment was 50% and in the second experiment was 62% (see  
19 also Holdgraf et al., 2017). These results are approximate, because we only had two repeats  
20 per observer, but we can conclude that the Edge Model captures a reasonable portion of the  
21 explainable variance within and across observers in the current data. Thus, the results of this  
22 analysis suggest that image cues at the binocular-monocular transition regions of a depth edge  
23 affect perceptual stability, and that this effect is consistent with observed statistical regularities  
24 in real depth edges from natural scenes.

25

26



1

2 **Figure 7.** Actual versus fitted instability ratings for the Disparity-Only model (A) and Edge model  
3 (B). Each panel shows the instability ratings of all observers on the abscissa, and the ratings  
4 produced by the best fit model on the ordinate. The top row shows the data from Experiment 1,  
5 and the bottom row shows the data from Experiment 2.

6

7 To further explore this result, we examined to what extent luminance or contrast differences  
8 between the monocular and binocular regions contributed to the mediating effects of transition  
9 region vertical edges on perceptual stability. We created two additional models, each with two  
10 predictors in addition to disparity magnitude: the *Luminance model* and the *Contrast model*. In  
11 the Luminance model, we used the luminance difference between the monocular region and the  
12 background, and between the monocular region and the foreground (**Figure 5D**) as predictors.  
13 In the Contrast model, we replaced these with the contrast differences (**Figure 5E**). Both of these  
14 signals can contribute to the strength of the vertical edge. The results suggest that the luminance  
15 differences, and to some extent contrast differences, contributed to the influence of binocular-  
16 monocular vertical edges on perceptual stability. Specifically, greater luminance differences

1 between the monocular region and the binocular foreground were associated with more stability  
 2 (lower ratings), and greater luminance differences between the monocular region and the  
 3 binocular background were associated with more instability (higher ratings). Greater contrast  
 4 differences between the monocular region and the binocular background were associated with  
 5 more instability (this effect was only statistically significant in Experiment 2). The full results are  
 6 reported in **Table 2**. In practice, the identification of binocular corresponding and non-  
 7 corresponding points is likely to also rely on higher order image features. However, these results  
 8 provide some insight into the nature of the surface differences that may facilitate binocular fusion.  
 9

10 **Table 2.** Regression model results for luminance and contrast differences. Data are formatted  
 11 the same as in Table 1.

<b>Experiment 1</b>					
	<b>Fixed Effects</b>	<b>Coefficient estimate</b>	<b>T statistic</b>	<b>p value</b>	<b>r<sup>2</sup></b>
<b>Luminance Model</b>	Intercept	23.53 (12.60,34.45)	4.23	<<0.001	0.31
	Disparity range	71.78 (57.34,86.22)	9.76	<<0.001	
	Background-monocular difference	37.10 (17.93,56.28)	3.80	<<0.001	
	Foreground-monocular difference	-42.63 (-53.78,-31.49)	-7.51	<<0.001	
<b>Contrast Model</b>	Intercept	7.46 (-3.20,18.12)	1.37	0.17	0.24
	Disparity range	76.71 (61.36,92.06)	9.81	<<0.001	
	Background-monocular difference	8.00 (-0.43,16.43)	1.86	0.06	
	Foreground-monocular difference	1.45 (-4.16,7.05)	0.51	0.61	

<b>Experiment 2</b>					
	<b>Fixed Effects</b>	<b>Coefficient estimate</b>	<b>T statistic</b>	<b>p value</b>	<b>r<sup>2</sup></b>
<b>Luminance Model</b>	Intercept	15.51 (9.80,21.22)	5.33	<<0.001	0.29
	Disparity range	74.97 (67.37,82.57)	19.35	<<0.001	
	Background-monocular difference	9.98 (6.64,13.32)	5.86	<<0.001	
	Foreground-monocular difference	-5.50 (-7.22, -3.78)	-6.27	<<0.001	
<b>Contrast Model</b>	Intercept	10.21 (4.55,15.88)	3.53	<<0.001	0.28
	Disparity range	78.27 (70.60,85.94)	20.02	<<0.001	
	Background-monocular difference	5.74 (1.96,9.52)	2.98	0.003	
	Foreground-monocular difference	0.64 (1.97,3.25)	0.48	0.63	

12  
 13  
 14



1 In a final analysis, we examined how perceptual instability experienced by our observers when  
2 viewing natural depth edges relates to stimulus features that are known to exacerbate binocular  
3 rivalry in simple stimuli. Previous studies on binocular rivalry that used synthetically mismatched  
4 stimuli (e.g., sinusoidal gratings with different orientations in the two eyes), have reported that  
5 low contrast stimuli tend to be more perceptually stable than their high contrast counterparts.  
6 Sometimes, stimuli that cause unstable percepts at high contrast are fused into stable unified  
7 percepts at low contrast. For example, when orthogonal gratings are dichoptically presented at  
8 high contrast, they cause strikingly unstable percepts that alternate between the two grating  
9 orientations. When the same orthogonal gratings are presented at low contrast, they are  
10 perceived as a stable plaid (L. Liu et al., 1992). Does the relationship between contrast and  
11 perceptual instability hold in natural images when monocular regions are caused by depth  
12 edges? First, we examined the correlation between the instability ratings and the image contrast  
13 in the binocular background, monocular, and binocular foreground regions in Experiment 2. As  
14 expected, binocular instability tended to be higher when contrast was higher. Higher instability  
15 ratings were associated with higher contrast in the monocular region ( $r = 0.20$ , 95% confidence  
16 interval 0.07-0.33,  $p = 0.02$ ) and with higher contrast in the adjacent binocular foreground ( $r =$   
17  $0.31$ , 95% confidence interval 0.15-0.45,  $p \ll 0.001$ ). This relationship was weaker and not  
18 statistically significant in the binocular background ( $r = 0.17$ , 95% confidence interval 0.04-0.30,  
19  $p = 0.06$ ). Next, we examined the correlation between instability ratings and contrast in the  
20 transition regions. Instability ratings were higher with higher contrast in the background-  
21 monocular transition region ( $r = 0.19$ , 95% confidence interval 0.05-0.32,  $p = 0.03$ ), but the same  
22 was not true for the foreground-monocular transition region ( $r = -0.097$ , 95% confidence interval  
23  $-0.25-0.09$ ,  $p = 0.28$ ). The absence of a correlation between perceptual instability and contrast  
24 in the foreground-monocular transition region may be connected to the natural scene statistics.  
25 Because *dissimilarity* between the monocular region and the binocular foreground is highly likely,  
26 a high contrast in this transition region provides information about the existence of an edge in  
27 the natural environment, which may facilitate rather than impede fusion.

28  
29 Taken together, these results suggest that depth edges with features that are statistically more  
30 likely are also likely to be more perceptually stable. Because monocular regions tend to be  
31 similar to the adjacent binocular background than to the adjacent binocular foreground in natural

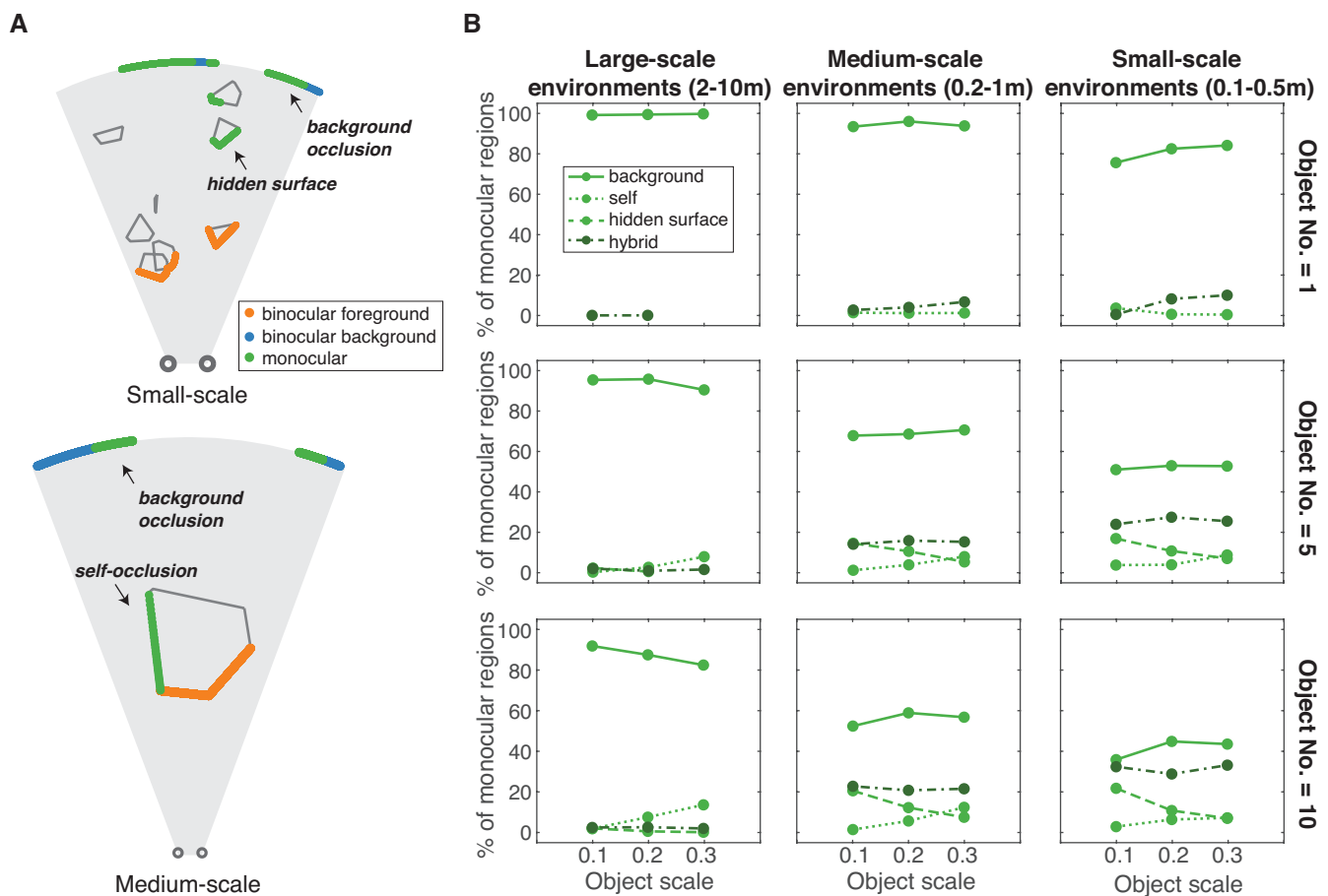
1 scenes, the visual system might be inclined to correctly identify these regions as monocular  
2 when they share features that are similar to the binocular background instead of the foreground.

### 4 ***Geometric Simulation***

5 The results from the natural scene statistics analysis suggest that natural depth edges have  
6 robust statistical regularities that could be exploited by the visual system (i.e., monocularly visible  
7 regions tend to share visual properties with the binocular background). The perceptual studies  
8 suggest that stimuli adhering to this statistical regularity are more likely to be perceived as stable,  
9 and stimuli that violate it are more likely to be perceived as unstable. One potential caveat with  
10 these results is that the sampled patches used in this study were drawn from a set of outdoor  
11 scenes that were photographed such that the closest objects were 3m away. If the results  
12 depend on viewing distance, these may not generalize to near viewing distances. In scenes with  
13 close by objects, it is possible that self-occlusions and hidden surfaces may occur more  
14 frequently than in the scenes that we analyzed. If so, the conclusions that we have drawn may  
15 apply only to distant visual scenes.

16  
17 Prior work has used geometric modeling as an additional tool to explore potential statistical  
18 regularities in binocular and monocular visibility in 3D scenes (Hansard, 2012; Langer &  
19 Mannan, 2012; Langer et al., 2016). This approach is appealing because it allows the modeler  
20 to parametrically vary properties of a scene (e.g., object distance and clutter). For example,  
21 Langer & colleagues asked how the probability of points that are visible only monocularly varies  
22 as a function of object distance (Langer & Mannan, 2012; Langer et al., 2016). They found that  
23 the probability that a given point is monocularly visible increases monotonically as a function of  
24 its distance from the viewer, and that this increase in probability is more rapid when the objects  
25 are smaller. However, because the simulated scenes were populated only with identical planar  
26 surfaces, self-occlusions were not present in these simulations. Thus, the relative frequency of  
27 the different causes of monocular regions (the topic of the current investigation) could not be  
28 examined. Building on this prior work, we performed a parametric study of simulated 3D scenes  
29 populated with objects capable of self-occlusion. Specifically, we examined how changes in  
30 viewing distance and the properties of the 3D environment change the proportion of monocular  
31 regions tend to be due to background occlusions, self-occlusions, or hidden surfaces. This

1 simulation also allowed us to explore the prevalence of hybrid occlusions that contain multiple  
 2 monocularly visible surfaces.  
 3  
 4 Each simulation included a binocular observer positioned within a different environment. Each  
 5 environment was parameterized in terms of its scale (i.e. maximum object distance), its density  
 6 (i.e. total number of objects), and the size of the objects within it (as a proportion of the scale).  
 7 Objects were defined as polygons with a random number of sides (up to 7), and were positioned  
 8 randomly within the virtual observer's central visual field. Ray tracing was used to determine  
 9 whether each visible scene point was binocularly or monocularly visible. We limited our analyses  
 10 to the horizontal plane of regard. **Figure 8A** illustrates two example simulations.



11  
 12 **Figure 8.** Geometric simulation. **A.** Example diagrams of two simulated environments are shown  
 13 in a top-down view. The small circles at the bottom illustrate the location of a virtual observer's  
 14 eyes (interpupillary distance of 6.5 cm), and the shaded area represents the visible environment.  
 15 Object size was first specified as the diameter of a circle bounding a polygon, as a proportion of  
 16 the maximum distance in the current environment (0.1 in upper, 0.3 in lower panel). Random  
 17 vertices were then placed along each circle, so the objects' final sizes ranged depending on the  
 18 vertex locations. The number of vertices was randomly selected to be between 2 and 8, resulting  
 19 in polygons with between 1 and 7 sides. A minimum object distance was set as 20% of the

1 maximum distance for each environment. The number of objects for each simulation was varied  
2 as well. **B.** We ran 200 simulations per parameter combination and determined monocular pixels  
3 by ray tracing over the central 40° field of view at a resolution of 60 rays per degree. We  
4 identified all unique monocular regions and categorized each region as being a background  
5 occlusion, self-occlusion, hidden surface occlusion, or hybrid occlusion. A hybrid occlusion was  
6 one that contained less than 80% of any of the three other categories. The effect of  
7 environmental scale (maximum distances of 10, 1, 0.5 m) is shown in different columns, and  
8 number of objects (1, 5, 10) in different rows. Each panel shows the percentage of monocular  
9 regions in each category, as a function of object scale (0.1, 0.2, 0.3). These scales correspond  
10 to sizes that range from 1-3m, 0.1-0.3m, and 0.05-0.15m for the large-, medium-, and small-  
11 scale environments, respectively. Data points at 0% are not plotted. Across all parameter  
12 combinations, 95% confidence intervals, as determined by the binomial distribution, ranged from  
13 6-11% (not plotted). The different line styles correspond to the four different kinds of monocular  
14 regions: background occlusion, self-occlusion, hidden surface occlusion, and hybrid occlusion.  
15

16 In each simulation, we identified all monocular regions occurring at depth edges within the  
17 observer's central visual field (40°), and categorized each monocular region as being a  
18 background occlusion, self-occlusion, hidden surface, or hybrid occlusion (see **Figure 2** and  
19 **Figure A1**). A hybrid occlusion was defined as a continuous monocular region that contained  
20 less than 80% of any of the three other categories. The results indicate that monocular regions  
21 are most likely to be background occlusions across diverse environments with different scales,  
22 object densities, and object sizes. Each subpanel of **Figure 8B** summarizes the results for a  
23 given environment scale (columns) and object density (rows), and the abscissa of each panel  
24 depicts variations in object scale.

25  
26 We start by examining the results from the large-scale environments, which are most similar to  
27 the scenes used in the natural scene statistics (left column). Consistent with the results from our  
28 analysis of natural stereo-images, the majority of monocular regions in these simulated  
29 environments tended to be background occlusions (solid lines). Other types of occlusions were  
30 present, but highly infrequent by comparison. These results suggest a strong constraint on the  
31 content of monocular regions when objects are relatively far away from the observer. It would  
32 be plausible for this constraint to weaken or reverse when objects are nearby. However, the  
33 observation that most monocular regions were background occlusions generalized to smaller  
34 scale environments in this simulation as well (right and center column). These smaller  
35 environments produced more monocular regions overall, with background occlusions still  
36 consistently the most prevalent. Not surprisingly, however, the relative proportion of other types  
37 of occlusions was higher in the small-scale environments compared to larger-scale

1 environments. For example, for small-scale environments with many objects, background  
2 occlusions and hybrid occlusions had relatively similar frequency. Naturally, as the number of  
3 objects increased (upper to lower row), the number of self-occlusions, hidden surfaces, and  
4 hybrid occlusions also increased. Interestingly, hidden surfaces and hybrid occlusions were  
5 often more frequent than a self-occlusion in this simulation.

6  
7 Lastly, our results are consistent with the simulations presented by Langer & Mannan (2012), in  
8 that we observe more monocular regions when the environments are populated with smaller  
9 objects (Langer & Mannan, 2012). For example, in the small-scale environments with the many  
10 small objects, over 30% of visible points were monocular, whereas in the large-scale  
11 environments less than 3% of visible points were monocular. The assumptions made for these  
12 simulations clearly do not fully capture the complex structure of natural scenes. Nevertheless,  
13 the assumptions allowed us to parametrically explore how different environmental factors may  
14 affect the prevalence and properties of monocularly visible surfaces. These simulations suggest  
15 that the monocularly visible regions of real-world scenes are much more likely to belong to  
16 background surfaces.

17

## 18 **Discussion**

19 The images that are cast on the retinas during natural viewing carry incomplete information about  
20 the three-dimensional scene in front of the observer. However, the visual system often exploits  
21 statistical regularities in these images to achieve accurate percepts of the environment.  
22 Perception of depth edges may be facilitated by assumptions about the configurations that are  
23 likely to occur in natural scenes. The current work provides evidence that there are indeed strong  
24 statistical regularities near depth edges in natural scenes, and that these regularities have a  
25 systematic effect on perceptual stability.

26

27 To examine these perceptual effects, we measured human perceptual stability in two  
28 experiments using stimuli taken from natural scenes and analyzed the results with mixed linear  
29 regression models. Models that included image features (e.g., edge strength, luminance  
30 differences) in addition to disparity as predictors for perceptual instability explained 7-8% more  
31 variance than a Disparity-Only model. This is a promising result given the relatively small total  
32 effect size and the limitations of the current approach. For example, it is likely that there are

1 stimulus factors (e.g. higher-order textural properties) that are useful for predicting instability and  
2 are not well captured by the current set of predictors. In addition, in the current perceptual  
3 experiment, all stimuli were presented at a fixed distance (relative to the foreground), even  
4 though the natural scene patches were sampled from a range of distances. An exploratory  
5 analysis suggested that patches derived from farther distances may tend to be more perceptually  
6 stable, perhaps because of differences in textural or perspective cues. The interaction between  
7 binocular disparity and textural cues at depth edges represents an interesting direction for future  
8 work.

9  
10 Our current set of models explained only around 1/3 of the variance in mean instability ratings,  
11 leaving a substantial proportion of the variance unaccounted for. Indeed, repeated presentations  
12 of the same stimulus could elicit quite variable instability ratings within observers. With additional  
13 training procedures or a two-alternative forced choice design, this stimulus-independent  
14 variability in our response measure may be further reduced. Another potential next step might  
15 be to employ more complex textural descriptors derived from natural scenes, or to use automatic  
16 techniques for finding the most useful image features for the task (Burge & Jaini, 2017; Geisler,  
17 Najemnik, & Ing, 2009; Jaini & Burge, 2017). Doing so may reveal additional stimulus factors  
18 that are useful for predicting perceptual instability. Finally, image manipulation may also be used  
19 causally manipulate perceptual instability, for example, by introducing luminance and contrast  
20 differences between adjacent image regions. However, locally manipulating natural images  
21 without producing visible artifacts can be challenging.

22  
23 The current work has focused on local statistics and low-level image features. In addition to  
24 these local statistics, the global structure of a scene is likely to play a role in the perceptual  
25 stability of depth edges. For example, a previous study that used photographs of real world  
26 objects as stimuli (tabletop scenes with boxes) suggested that correct depth ordering  
27 judgements were facilitated for self-occlusions, as compared to background occlusions (Wilcox  
28 & Lakra, 2007). Such stimuli may produce a strong expectation that the side of the object should  
29 be visible. Recent work also suggests that the detection of depth edges in natural images may  
30 be facilitated by taking global layout information (e.g., elevation) into account (Ehinger, Graf,  
31 Adams, & Elder, 2017). More global contextual cues may thus lead the visual system to interpret  
32 a given scene as more likely to contain one type of occlusion over another, but a thorough

1 understanding of local image cues is clearly important. Early, feedforward stages of visual  
2 processing rely on localized computations that are likely made more informative by incorporating  
3 low-level statistical regularities from the environment. For example, Anderson & Nakayama  
4 (1994) proposed that binocular combination relies on both disparity-tuned receptive fields and  
5 receptive fields tuned to detect monocular regions at depth edges at the earliest stages of visual  
6 processing. In the search for neural correlates of binocular combination, it has been  
7 demonstrated that binocular populations in V1 include neurons that are sensitive to interocular  
8 differences that do not typically occur for binocular points during natural vision (e.g., phase  
9 disparities) (Ohzawa, DeAngelis, & Freeman, 1990; Read & Cumming, 2007). However, the  
10 ability of these populations to detect monocular points is unknown. Recent work on the natural  
11 statistics of depth edges may provide guidance for investigating the predicted properties of  
12 neurons optimized to encode the combination of monocular and binocular features at depth  
13 edges (Iyer & Burge, 2018).

14  
15 Inspired by the proposed mechanisms for biological vision, some computer vision approaches  
16 also aim to simultaneously estimate the disparity of binocular points and locations of monocular  
17 points (Wang & Zickler, 2019; Weng, Ahuja, & Huang, 1988). In the field of computer vision,  
18 monocular regions have been of interest for their role in facilitating detection of object boundaries  
19 when stereo image pairs are available. Recent work in this area has begun to incorporate  
20 assumptions about typical scene geometry into this procedure (Wang & Zickler, 2019). However,  
21 they have only considered variations of background occlusions, rather than detecting self-  
22 occlusions or hidden objects. Based on the primary results presented in the current study, it  
23 seems probable that this approach from computer vision would work well on natural stereo  
24 images, given that background occlusions are the most likely type of occlusion. Nonetheless,  
25 the results of our simulation suggest that hidden object scenarios may also contain useful cues  
26 to depth edges, particularly in cluttered scenes.

27  
28 Looking forward, a systematic exploration of how visual context across multiple scales impacts  
29 binocular combination may improve the ability to predict when natural stimuli are most easily  
30 fused. In addition, a complete understanding of binocular fusion during natural vision should take  
31 into account complex local depth relationships, which can include nested monocular regions  
32 associated with multiple half-occluded surfaces (see Assee & Qian, 2007). This understanding

1 would be facilitated by additional datasets from a wider variety of natural scene types (e.g.,  
2 indoors, outdoors, natural, man-made, etc.). Finally, the temporal dynamics of natural vision,  
3 caused by motion in the scene as well as observer motion, likely have a strong modulatory effect  
4 on both visual statistics and perceptual stability. As we learn more about the combined features  
5 of depth and other cues during natural binocular vision, we can expect that additional sources  
6 of information may be discovered that facilitate stable percepts.

7

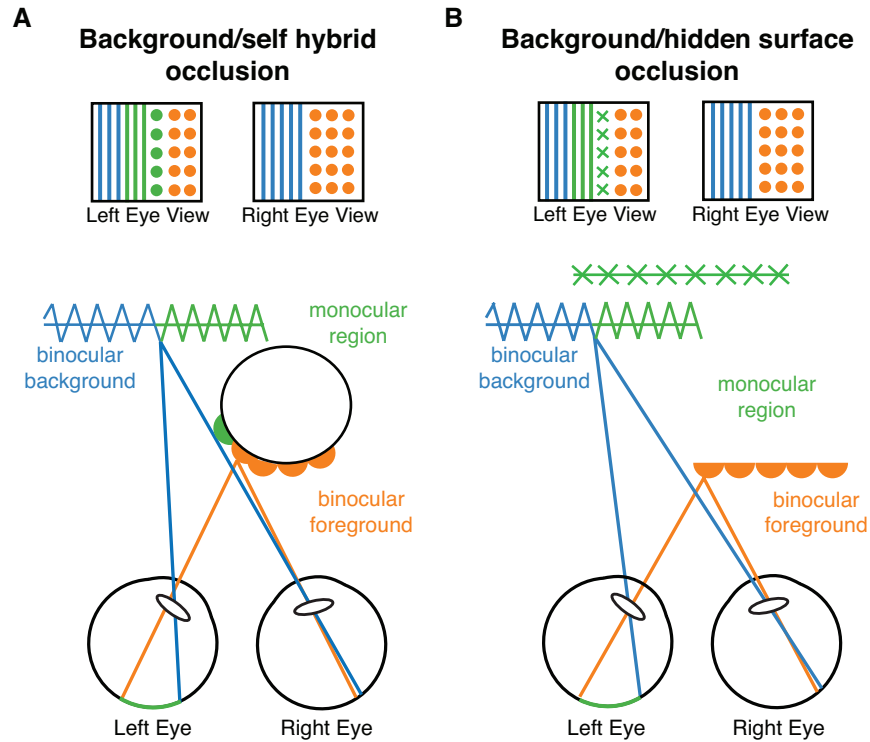
## 8 **Acknowledgements**

9 The authors would like to thank Steve Cholewiak for equipment assistance. EAC and ZB were  
10 supported by Google. JB and DNW were supported by NIH grant R01-EY028571 from the  
11 National Eye Institute and the Office of Behavioral and Social Sciences Research. JB was also  
12 supported by startup funds from the University of Pennsylvania.



# 1 Appendix

2



3

4 **Figure A1.** Additional diagrams illustrating cases in which the monocular region has multiple  
5 visible surfaces (hybrid occlusions). **A.** In a background/self hybrid occlusion, the left eye sees  
6 both a portion of the foreground and background that are hidden from the right eye. **B.** In a  
7 background/hidden surface hybrid occlusion, the background is interrupted by a hidden  
8 surface, which is also monocularly visible.

9

10

11

## 12 References

13 Anderson, B. L. (1994). The role of partial occlusion in stereopsis. *Nature*, 367, 365–368.

14 <https://doi.org/10.1038/nature02371.1>.

15 Assee, A., & Qian, N. (2007). Solving da Vinci stereopsis with depth-edge-selective V2 cells.

16 *Vision Research*, 47(20), 2585–2602. <https://doi.org/10.1016/j.visres.2007.07.003>

17 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.

18 <https://doi.org/10.1163/156856897X00357>

19 Bülthoff, H., Fahle, M., & Wegmann, M. (1991). Perceived depth scales with disparity gradient.

20 *Perception*, 20(2), 145–153. <https://doi.org/10.1068/p200145>

21 Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the

- 1 figure-ground cue of convexity affects human depth perception. *Journal of Neuroscience*,  
2 30(21), 7269–7280. <https://doi.org/10.1523/JNEUROSCI.5551-09.2010>
- 3 Burge, J., & Geisler, W. S. (2014). Optimal disparity estimation in natural stereo images.  
4 *Journal of Vision*, 14(2), 1. <https://doi.org/10.1167/14.2.1>
- 5 Burge, J., & Jaini, P. (2017). Accuracy Maximization Analysis for Sensory-Perceptual Tasks:  
6 Computational Improvements, Filter Robustness, and Coding Advantages for Scaled  
7 Additive Noise. *PLoS Computational Biology*, 13(2), e1005281.  
8 <https://doi.org/10.1371/journal.pcbi.1005636>
- 9 Burge, J., McCann, B. C., & Geisler, W. S. (2016). Estimating 3D tilt from local image cues in  
10 natural scenes. *Journal of Vision*, 16(13), 2. <https://doi.org/10.1167/16.13.2>
- 11 Chauhan, T., Masquelier, T., Montlibert, A., & Cottureau, B. R. (2018). Emergence of binocular  
12 disparity selectivity through hebbian learning. *Journal of Neuroscience*, 38(44), 9563–  
13 9578. <https://doi.org/10.1523/JNEUROSCI.1259-18.2018>
- 14 Cooper, E. A., Burge, J., & Banks, M. S. (2011). The vertical horopter is not adaptable, but it  
15 may be adaptive. *Journal of Vision*, 11(3), 1–19. <https://doi.org/10.1167/11.3.1>
- 16 Cooper, E. A., & Norcia, A. M. (2015). Predicting Cortical Dark/Bright Asymmetries from  
17 Natural Image Statistics and Early Visual Transforms. *PLoS Computational Biology*, 11(5),  
18 e1004268. <https://doi.org/10.1371/journal.pcbi.1004268>
- 19 Ehinger, K. A., Graf, E. W., Adams, W. J., & Elder, J. H. (2017). Local depth edge detection in  
20 humans and deep neural networks. In *Proceedings - 2017 IEEE International Conference*  
21 *on Computer Vision Workshops, ICCVW 2017*. <https://doi.org/10.1109/ICCVW.2017.316>
- 22 Farid, H., & Simoncelli, E. P. (1997). Optimally rotation-equivariant directional derivative  
23 kernels. In *Lecture Notes in Computer Science (including subseries Lecture Notes in*  
24 *Artificial Intelligence and Lecture Notes in Bioinformatics)*. [https://doi.org/10.1007/3-540-](https://doi.org/10.1007/3-540-63460-6_119)  
25 [63460-6\\_119](https://doi.org/10.1007/3-540-63460-6_119)
- 26 Forte, J., Peirce, J. W., & Lennie, P. (2002). Binocular integration of partially occluded  
27 surfaces. *Vision Research*, 42(10), 1225–1235. [https://doi.org/10.1016/S0042-](https://doi.org/10.1016/S0042-6989(02)00053-6)  
28 [6989\(02\)00053-6](https://doi.org/10.1016/S0042-6989(02)00053-6)
- 29 Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use,  
30 calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 30.  
31 <https://doi.org/10.1037/a0024338>
- 32 Geisler, W. S. (2008). Visual Perception and the Statistical Properties of Natural Scenes.

- 1        *Annual Review of Psychology*, 59, 167–192.  
2        <https://doi.org/10.1146/annurev.psych.58.110405.085632>
- 3        Geisler, W. S., Najemnik, J., & Ing, A. D. (2009). Optimal stimulus encoders for natural tasks.  
4        *Journal of Vision*, 9(13), 1–16. <https://doi.org/10.1167/9.13.1>
- 5        Gibaldi, A., Canessa, A., & Sabatini, S. P. (2017). The active side of stereopsis: Fixation  
6        strategy and adaptation to natural environments. *Scientific Reports*, 7, 44800.  
7        <https://doi.org/10.1038/srep44800>
- 8        Gillam, B., & Borsting, E. (1988). The role of monocular regions in stereoscopic displays.  
9        *Perception*, 17(5), 603–608. <https://doi.org/10.1068/p170603>
- 10        Goncalves, N. R., & Welchman, A. E. (2017). “What Not” Detectors Help the Brain See in  
11        Depth. *Current Biology*, 27(10), 1403–1412.e8. <https://doi.org/10.1016/j.cub.2017.03.074>
- 12        Grove, P. M., Ben Sachtler, W. L., & Gillam, B. J. (2006). Amodal completion with background  
13        determines depth from monocular gap stereopsis. *Vision Research*, 46(22), 3771–3774.  
14        <https://doi.org/10.1016/j.visres.2006.06.020>
- 15        Grove, P. M., Gillam, B., & Ono, H. (2002). Content and context of monocular regions  
16        determine perceived depth in random dot, unpaired background and phantom  
17        stereograms. *Vision Research*, 42(15), 1859–1870. [https://doi.org/10.1016/S0042-](https://doi.org/10.1016/S0042-6989(02)00083-4)  
18        [6989\(02\)00083-4](https://doi.org/10.1016/S0042-6989(02)00083-4)
- 19        Grove, P. M., & Ono, H. (1999). Ecologically invalid monocular texture leads to longer  
20        perceptual latencies in random-dot stereograms. *Perception*, 28(5), 627–639.  
21        <https://doi.org/10.1068/p2908>
- 22        Hansard, M. (2012). Binocular projection of a random scene. In *BMVC 2012 - Electronic*  
23        *Proceedings of the British Machine Vision Conference 2012* (pp. 1–11).  
24        <https://doi.org/10.5244/C.26.90>
- 25        Harris, J. M., & Wilcox, L. M. (2009). The role of monocularly visible regions in depth and  
26        surface perception. *Vision Research*, 49(22), 2666–2685.  
27        <https://doi.org/10.1016/j.visres.2009.06.021>
- 28        Hibbard, P. B. (2008). Binocular energy responses to natural images. *Vision Research*, 48,  
29        1427–1439. <https://doi.org/10.1016/j.visres.2008.03.013>
- 30        Hibbard, P. B., & Bouzit, S. (2005). Stereoscopic correspondence for ambiguous targets is  
31        affected by elevation and fixation distance. *Spatial Vision*, 18(4), 399–411.
- 32        Hoffman, D. M., & Banks, M. S. (2010). Focus information is used to interpret binocular

- 1 images. *Journal of Vision*, 10(5), 13. <https://doi.org/10.1167/10.5.13>
- 2 Holdgraf, C. R., Rieger, J. W., Micheli, C., Martin, S., Knight, R. T., & Theunissen, F. E. (2017).  
3 Encoding and decoding models in cognitive electrophysiology. *Frontiers in Systems*  
4 *Neuroscience*, 11, 61. <https://doi.org/10.3389/fnsys.2017.00061>
- 5 Hsu, A., Borst, A., & Theunissen, F. E. (2004). Quantifying variability in neural responses and  
6 its application for the validation of model predictions. *Network: Computation in Neural*  
7 *Systems*, 15(2), 91–109. [https://doi.org/10.1088/0954-898X\\_15\\_2\\_002](https://doi.org/10.1088/0954-898X_15_2_002)
- 8 Iyer, A. V., & Burge, J. (2018). Depth variation and stereo processing tasks in natural scenes.  
9 *Journal of Vision*, 18(6), 4. <https://doi.org/10.1167/18.6.4>
- 10 Jaini, P., & Burge, J. (2017). Linking normative models of natural tasks to descriptive models of  
11 neural response. *Journal of Vision*, 17(12), 1–26. <https://doi.org/10.1167/17.12.16>
- 12 Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3 ? *Perception*,  
13 36(ECVP Abstract Supplement). <https://doi.org/10.1068/v070821>
- 14 Langer, M. S., & Mannan, F. (2012). Visibility in three-dimensional cluttered scenes. *Journal of*  
15 *the Optical Society of America A*, 29, 1794–1807. <https://doi.org/10.1364/josaa.29.001794>
- 16 Langer, M. S., Zheng, H., & Rezvankhah, S. (2016). Depth discrimination from occlusions in  
17 3D clutter. *Journal of Vision*, 16(11), 1–18. <https://doi.org/10.1167/16.11.11>
- 18 Levelt, W. (1965). *On binocular rivalry*. (Doctoral dissertation, Van Gorcum Assen).
- 19 Liu, L., Tyler, C. W., & Schor, C. M. (1992). Failure of rivalry at low contrast: Evidence of a  
20 suprathreshold binocular summation process. *Vision Research*, 32(8), 1471–1479.  
21 [https://doi.org/10.1016/0042-6989\(92\)90203-U](https://doi.org/10.1016/0042-6989(92)90203-U)
- 22 Liu, Y., Cormack, L. K., & Bovik, A. C. (2009). Luminance, disparity, and range statistics in 3D  
23 natural scenes. In *Human Vision and Electronic Imaging XIV*.  
24 <https://doi.org/10.1117/12.811459>
- 25 Nakayama, K., & Shimojo, S. (1990). Da vinci stereopsis: Depth and subjective occluding  
26 contours from unpaired image points. *Vision Research*, 30(11), 1811–1825.  
27 [https://doi.org/10.1016/0042-6989\(90\)90161-D](https://doi.org/10.1016/0042-6989(90)90161-D)
- 28 Ohzawa, I., DeAngelis, G., & Freeman, R. (1990). Stereoscopic depth discrimination in the  
29 visual cortex: neurons ideally suited as disparity detectors. *Science*, 249(4972), 1037–  
30 1041. <https://doi.org/10.1126/science.2396096>
- 31 Pelli, D. G. (1997). The VideoToolbox software for visual psychphysics: transforming numbers  
32 into movies. *Spatial Vision*, 10(4), 437–442.

- 1 Read, J. C. A., & Cumming, B. G. (2007). Sensors for impossible stimuli may solve the stereo  
2 correspondence problem. *Nature Neuroscience*, *10*(10), 1322–1328.  
3 <https://doi.org/10.1038/nn1951>
- 4 Samonds, J. M., Potetz, B. R., & Lee, T. S. (2012). Relative luminance and binocular disparity  
5 preferences are correlated in macaque primary visual cortex, matching natural scene  
6 statistics. *Proceedings of the National Academy of Sciences of the United States of*  
7 *America*, *109*(16), 6313–6318. <https://doi.org/10.1073/pnas.1200125109>
- 8 Shimojo, S., & Nakayama, K. (1990). Real world occlusion constraints and binocular rivalry.  
9 *Vision Research*, *30*(1), 69–80. [https://doi.org/10.1016/0042-6989\(90\)90128-8](https://doi.org/10.1016/0042-6989(90)90128-8)
- 10 Sprague, W. W., Cooper, E. A., Tošić, I., & Banks, M. S. (2015). Stereopsis is adaptive for the  
11 natural environment. *Science Advances*, *1*(4), e1400254.  
12 <https://doi.org/10.1126/sciadv.1400254>
- 13 Tyler, C. W., & Kontsevich, L. L. (1995). Mechanisms of Stereoscopic Processing:  
14 Stereoattention and Surface Perception in Depth Reconstruction. *Perception*, *24*(2), 127–  
15 153. <https://doi.org/10.1068/p240127>
- 16 Wang, J., & Zickler, T. (2019). Local detection of stereo occlusion boundaries. In *The IEEE*  
17 *Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3818–3827).
- 18 Weng, J., Ahuja, N., & Huang, T. S. (1988). Two-view Matching. In *[1988 Proceedings] Second*  
19 *International Conference on Computer Vision* (pp. 64–73). IEEE.  
20 <https://doi.org/10.1109/CCV.1988.589972>
- 21 White, D. N., & Burge, J. (2019). Depth Estimates in Half Occlusions in Natural Scenes. In  
22 *Vision Sciences Society 19th Annual Meeting*. <https://doi.org/10.1167/19.10.64c>
- 23 Wilcox, L. M., & Lakra, D. C. (2007). Depth from binocular half-occlusions in stereoscopic  
24 images of natural scenes. *Perception*, *36*(6), 830–839. <https://doi.org/10.1068/p5708>  
25