

# ORDINAL JUDGMENTS OF DEPTH IN MONOCULARLY- AND STEREOSCOPICALLY-VIEWED PHOTOGRAPHS OF COMPLEX NATURAL SCENES

*Rebecca L. Hornsey<sup>1</sup>, Paul B. Hibbard<sup>1</sup> and Peter Scarfe<sup>2</sup>*

1 Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ UK.

2 School of Psychology and Clinical Language Sciences, University of Reading, Earley Gate, Whiteknights Road, Reading RG6 6AL UK.

## ABSTRACT

This study investigated the contribution of stereoscopic depth cues to the reliability of ordinal depth judgments in complex natural scenes. Participants viewed photographs of cluttered natural scenes, either monocularly or stereoscopically. On each trial, they judged which of two indicated points in the scene was closer in depth. We assessed the reliability of these judgments over repeated trials, and how well they correlated with the actual disparities of the points between the left and right eyes' views. The reliability of judgments increased as their depth separation increased, was higher when the points were on separate objects, and deteriorated for point pairs that were more widely separated in the image plane. Stereoscopic viewing improved sensitivity to depth for points on the same surface, but not for points on separate objects. Stereoscopic viewing thus provides depth information that is complementary to that available from monocular occlusion cues.

**Index Terms**— Depth perception, ordinal depth, stereopsis, S3D, complex natural scenes.

## 1. INTRODUCTION

Traditional 2D photographs contain many 'pictorial' cues that allow us to estimate depth [2]. For example, the occlusion of one object by another provides information about which of the two objects is closer to us. Perspective cues, such as texture variations, provide more detailed information, such as the slant and tilt of a surface, but only up to a scaling factor [8]. Thus, while it might be possible to accurately determine the 3D orientation at every point on the surface of an object from texture variations, without additional information (e.g. familiarity with the typical size of the object) it is not possible to establish from these cues whether the object is small and close, or larger and viewed from a greater distance.

In contrast, the information provided by binocular vision allows us to make precise metric depth judgments. By combining binocular disparities (estimates of the

differences in the locations of corresponding points into our two eyes) with knowledge of our viewing geometry (the distance between our two eyes, and the direction in which each is looking) we can in principle determine the exact three-dimensional location of all binocularly visible points in a scene.

Binocular vision also provides a striking qualitative improvement in the perception of depth. This is related to the concept of *stereopsis*. This term, literally meaning 'solid sight', refers to the characteristically vivid impression of solidity and depth, as for example is provided in stereoscopic 3D (S3D) [16,17]. It has been proposed that stereopsis is associated not with the magnitude of depth perceived in a scene, but rather with the precision and reliability with which depth is represented [7,15-17].

In the current study, we assessed the quantitative change in perceived depth that accompanies the qualitatively enhanced stereopsis experienced in stereoscopically viewed photographs of complex natural scenes. Our specific aim was to provide a detailed assessment of the contribution of binocular information to the accuracy and precision of ordinal depth in stereoscopic photographs. In our experiments, participants were asked to determine which of two points in photographs of complex natural scenes was closer to them in depth.

The choice of this particular task is an important consideration, since there are a variety of different ways in which binocular cues can be used to provide depth information, dependent on the observer's task [5]. In the simplest case, for example, one only needs to detect a disparity between one surface and another in order to break camouflage, and segment the scene into two distinct objects. To establish that one object is closer than another, one needs to be able to establish the sign of their relative disparity, but does not need to know its magnitude. In contrast, judgments requiring access to the fully specified metric depth of the scene require the estimation of the



**Fig. 1.** An example of the stereoscopic images used, arranged for uncrossed (left and centre) or crossed (centre and right) viewing.

sign and magnitude of disparity, and knowledge of the viewing geometry [5].

Depth tasks can therefore be ordered hierarchically, depending on both the complexity of the information required, and the type of depth information that this then provides. In the current study, we assessed ordinal depth judgments. This task was chosen as a case in which we might predict a clear benefit from binocular viewing [13].

This is firstly because previous research has found that ordinal depth judgments from pictorial cues can be unreliable and inconsistent for widely separated points in complex scenes [11]. One explanation that has been put forward to explain this is that people can only make accurate depth judgments when points are on a single slope in depth, without any maxima or minima in depth in between them [11]. As the separation between the points in the image is increased, the probability increases that a local maximum or minimum in depth will occur between the points.

In these cases, binocular disparity should still, in principle, allow observers to make accurate and reliable ordinal depth judgments. This is because the task can be performed based simply on the sign of the relative disparity between the points [5]. Variation in depth at other points in the image should not affect the ability to use this information. However, it is also clear, from studies that have used simple arrangements of dots and lines, that our sensitivity to relative depth is strongly affected by the local arrangement of image features in depth [3-4]. Our ability to make use of relative binocular disparity will therefore depend on the local spatial variations in depth, which will vary considerably in complex natural scenes.

A further complication in trying to understand how binocular cues contribute to depth perception is that the results that we obtain from sparse stimuli do not readily generalize to more complex scenes. A striking demonstration of this is the fact that simple ordinal depth judgments are more reliable for individual points in empty space than they are for the same points lying on objects defined by multiple cues [13]. The deterioration in performance in this study occurred despite the fact that the surfaces contained much more information about the



**Fig. 2.** Participants judged whether the point in the scene indicated by the red or green dot was the closer; the distance between the points ( $D$ ) and the direction of their separation ( $\theta$ ) are indicated.

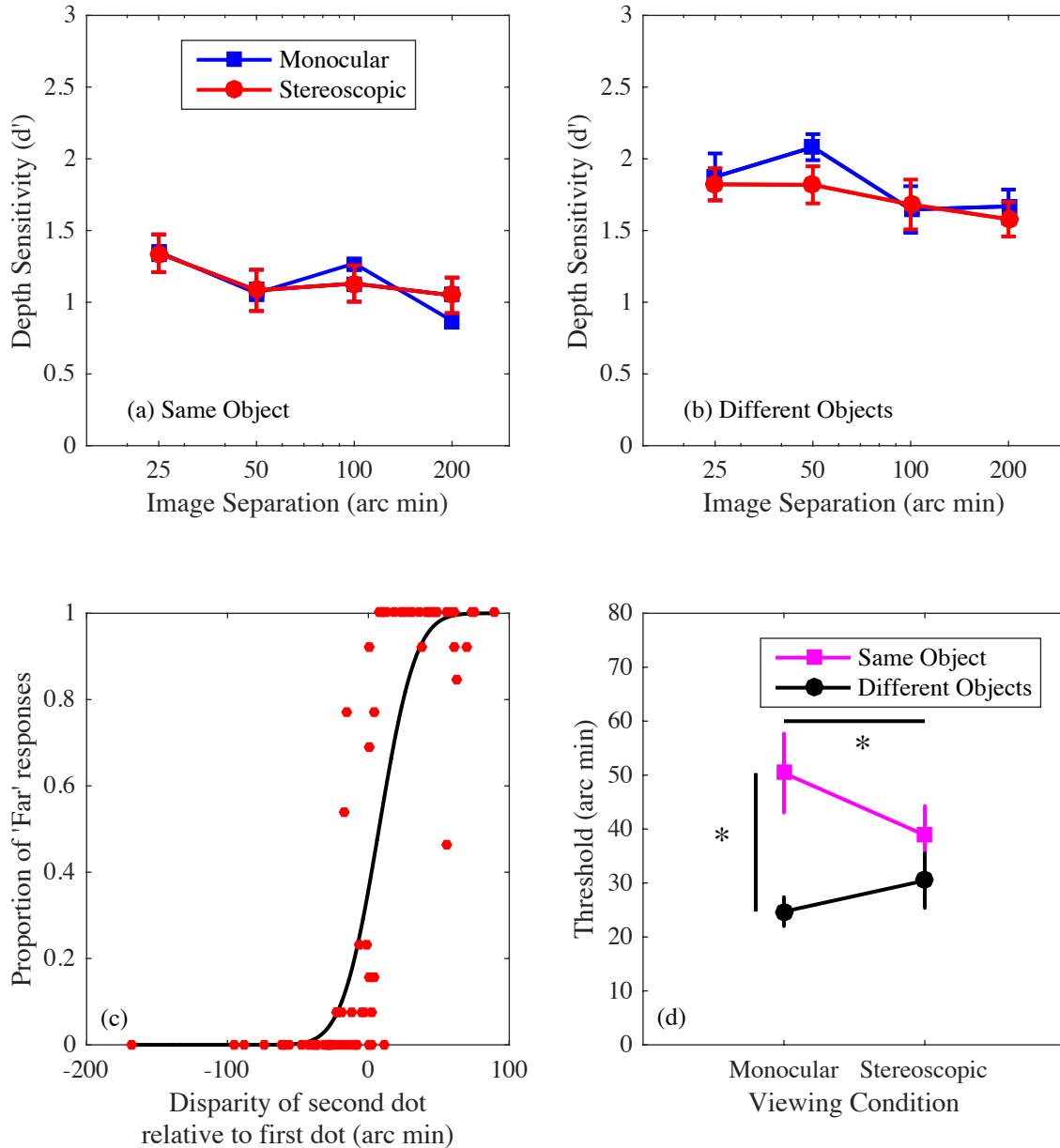
depth of the points, and could be considered more ecologically valid than sparse arrangements of dots in empty space.

In the current study, we therefore directly assessed the sensitivity of observers to ordinal depth differences in natural scenes, and how this is affected by the availability of binocular disparity information. The purpose of the study was to understand how binocular cues contribute to the perception of depth in complex natural scenes. We manipulated (i) the distance between the points to be compared in the image plane (ii) whether the points were on the same object, or a different object and (iii) whether the scenes were viewed monocularly or stereoscopically. The first manipulation was to test the prediction that ordinal depth judgments would become unreliable for points that are widely separated in the image. The second manipulation was to test whether the presence of monocular occlusion cues allowed for reliable ordinal depth judgments. The final manipulation was to test the contribution that binocular disparity can make to a simple depth judgment task in complex natural scenes.

## 2. METHODS

### 2.1 Apparatus

Stimuli were presented on a VIEWPIXX 3D monitor, viewed from a distance of 96cm. The monitor screen was 52cm wide and 29cm tall. The screen resolution was 1920x1080 pixels, with a refresh rate of 120Hz. Each pixel subtended 1 arc min. Stimuli were presented at 8 bit resolution. Stereoscopic presentation was achieved using a 3DPiXX IR emitter and NVIDIA 3D Vision LCD shutter glasses. Participants' responses were recorded using a RESPONSEPiXX response box. Stimuli were generated and presented using MATLAB and the Psychophysics Tool box extensions [1,10].



**Fig. 3.** Depth sensitivity, measured as  $d'$  values, for points that are (a) on the same object or (b) on different objects. Data are plotted separately for points viewed monocularly and stereoscopically in each case, as a function of the distance between the two points in the image plane. In each graph, the mean is calculated over all participants and all relevant point pairs, and the error bars represent  $\pm 1$ SEM, calculated over 20 point pairs in each case. (c) An example of a psychometric function in which the proportion of 'second dot far' responses are plotted against the relative disparity between the two dots. Data are for a single participant, with stereoscopic viewing, for all pairs in which the points are on different objects. (d) Depth discrimination thresholds, calculated as the distance between the 25% and 75% points on the psychometric functions. The mean threshold, across participants, is plotted separately for monocular and stereoscopic viewing, for points on the same object or on different objects. \* indicates significantly different values. Error bars show  $\pm 1$ SEM.

## 2.2 Participants

7 participants (3 male, 4 female) completed the experiment, including the experimenters RH and PH.

## 2.3 Stimuli

Stimuli consisted of four stereoscopic photographs of complex natural scenes, containing 'still-life'

	Viewing Condition					
	Monocular			Stereoscopic		
Predictor	Standardised $\beta$	t	Sig.	Standardised $\beta$	t	Sig.
Boundary	-0.363	-4.797	p<0.001	-0.332	-0.422	p<0.001
Separation	-0.264	-3.716	p<0.001	-0.206	-0.2779	p=0.006
Depth	0.253	3.181	p=0.002	0.231	2.784	p=0.006

**Table 1.** Results of the linear regression analysis, used to predict depth sensitivity ( $d'$ ) from (i) the presence or absence of an object boundary between the points (ii) their separation in the image plane and (iii) their actual depth separation, estimated from disparities measured from the stereoscopic image pairs.

arrangements of vegetables and plants. An example image pair is provided in Fig. 1. Images were taken with an inter-camera distance of 6.5cm; full details of the camera setup and calibration are described elsewhere [6].

The experiment was performed under monocular and stereoscopic viewing conditions. In the monocular condition, the photograph was presented to the participant's dominant eye, and a black screen to their other eye. In the stereoscopic condition, the correct image from the stereo pair was presented to each eye. Each block of trials consisted of monocular-only, or stereoscopic-only trials.

## 2.4 Procedure

On each trial, one of the photographs was presented, with two small dots (one red, one green) superimposed (Fig. 2). The dots were presented only to the participant's dominant eye. The participant's task was to decide whether the point in the scene indicated by the red or the green dot was the closer, and indicate this by pressing the corresponding button on the response box. The separation of the points (distance  $D$  in Fig. 2) was 25, 50, 100 or 200 arc min, and the direction of separation in the image plane (angle  $\theta$  in Fig. 2) varied between pairs of points. The points had been preselected by author PH so that they lay either on the same object, or on two different objects. For each photograph, 5 points on the same object, and 5 on different objects, were used. This created a total of 160 [4 (photographs) x 4 (separations) x 2 (same object or different) x 5 (samples)] distinct stimuli. Within a block of trials, each stimulus was presented once, in a randomized order. 10 blocks of trials were presented for each viewing condition (monocular and stereoscopic).

## 3. RESULTS

For each participant, and for each pair of points, we obtained a score of the proportion of times that the first point was chosen as the closer. These proportions were converted to  $d'$  values, as a measure of sensitivity to the

difference in depth for each point pair [9,12]. The average magnitude of these, across participants, was then calculated for each of the 160 pairs of points. Here, a score of 0 indicates no preference (that each point was chosen as the closer on 50% of trials) and larger values indicate more reliable responses.  $d'$  values could not be directly calculated when the proportion of times that the first point was chosen as closer was 1 or 0, since the inverse cumulative error function returns a value  $\pm\infty$  in these cases. To avoid this, values of 0 and 1 were replaced with 0.05 and 0.95 prior to the calculation of  $d'$  values [14]. This meant that the maximum magnitude of  $d'$ , for cases in which the participant gave the same response on all 10 trials, was 2.33.

Mean  $d'$  values are plotted in Fig. 3a,b as a function of the separation of the points in the image (on the x-axis), whether viewing was monocular (blues squares) or stereoscopic (red circles) and whether they belonged to the same (Fig. 3a), or different (Fig. 3b) objects. For each viewing condition, a linear regression was performed, with the mean  $d'$  value for each point as the dependent variable, and the surface boundary condition, image-plane separation and depth separation as predictors. The results of this regression are shown in Table 1. For both viewing conditions, the reliability of responses increased with increase depth separation, decreased with increasing image separation, and was greater for points on separate objects.

To determine the specific contribution of binocular cues, we calculated depth discrimination thresholds for points on the same or different objects, for monocular and stereoscopic viewing. This was done by plotting the percentage of 'second dot further' responses against the image disparity for all 80 points (separately for the sets of 'same object' and 'different object' points), for monocular and stereoscopic viewing. We then fit a cumulative Gaussian curve to these data, and calculated discrimination thresholds as half the distance between the 25% and 75% points on this curve. An example is shown in Fig 3c. Mean thresholds are shown in Fig. 3d. A two-way ANOVA showed a significant effect of boundary

condition ( $F(1,6)=7.304$ ;  $p=0.035$ ). This showed that thresholds were lower, and therefore that reliability was better, when the two points lay on separate objects. There was no overall difference between monocular and stereoscopic viewing ( $F(1,6)=5.236$ ;  $p=0.062$ ), but there was a significant interaction between the two factors ( $F(1,6)=7.80$ ;  $p=0.031$ ). Paired-sample t-tests showed that thresholds were significantly lower under stereoscopic viewing for points on the same surface ( $t(6)=2.61$ ;  $p=0.040$ ), but not for points on separate surfaces ( $t(6)=0.260$ ;  $p=0.803$ ). Paired-sample t-tests also showed that thresholds were lower for points on two separate objects under monocular viewing ( $t(6)=3.126$ ;  $p=0.020$ ), but not stereoscopic viewing ( $t(6)=1.932$ ;  $p=0.102$ ). These results show that stereoscopic cues increased the reliability of ordinal depth judgments for points on the same surface, but not for points on separate objects.

#### 4. DISCUSSION

We used judgments of ordinal depth separation to establish the contribution of binocular cues to the perception of depth in stereoscopic photographs. Reliability was greater for point pairs on separate objects, and for those with a greater depth separation, and decreased with increasing separation in the image plane. Binocular cues contributed to increasing the reliability of judgments only for points on the same surface, not for points on separate objects. This may be attributed to the additional monocular occlusion cues that are available in the latter case. Binocular cues contributed most to depth judgments within objects, and monocular cues to depth judgments between objects. Stereoscopic viewing thus provides depth information that is complementary to that available from monocular occlusion cues.

#### 5. REFERENCES

[1] D.H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10:433–436, 1997

[2] J.E. Cutting, and P.M. Vishton. Perceiving layout: The integration, relative dominance, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Handbook of perception and cognition: Vol. 5. Perception of space and motion*, pages 69–117. New York, NY: Academic Press, 1995

[3] M. Fahle, and G. Westheimer, G. Local and global factors in disparity detection of rows of points. *Vision Research*, 28:171–178, 1988.

[4] A. Glennerster, S.P. McKee, and M.D. Birch. Evidence for surface-based processing of binocular disparity. *Current Biology*, 12:825–828, 2002.

[5] A. Glennerster, B.J. Rogers, and M.F. Bradshaw. Stereoscopic depth constancy depends on the subject's task. *Vision Research*, 36:3441–3456, 1996.

[6] P.B. Hibbard. Binocular energy responses to natural images. *Vision Research*, 48:1427–1439, 2008.

[7] P.B. Hibbard. Can appearance be so deceptive? Representationalism and binocular vision. *Spatial Vision* 21:549–559, 2008.

[8] J.M. Hillis, S.J. Watt, M.S. Landy, and M.S. Banks. Slant from texture and disparity cues: optimal cue combination, *Journal of Vision*, 4:967–992, 2004.

[9] F.A.A. Kingdom, and A. Prins. *Psychophysics: a practical introduction*, London, Academic Press, 2009.

[10] M. Kleiner, D. Brainard, and D. Pelli, D. What's new in psychtoolbox-3? *Perception*, 36(ECVP Supplement):s14, 2007.

[11] J.J. Koenderink and A.J. van Doorn. Relief: Pictorial and otherwise, *Image and Vision Computing* 13:321–334, 1995.

[12] N.A. Macmillan, and C.D. Creelman *Detection theory: a user's guide*, Mahwah, New Jersey, Lawrence Erlbaum Associates, 2005.

[13] J.F. Norman, and J.T. Todd. Stereoscopic discrimination of interval and ordinal depth relations on smooth surfaces and in empty space, *Perception*, 27:257–272, 1998.

[14] K. Tsukida, and M.R. Gupta. M. R. (2011). How to analyze paired comparison data, *Technical Report UWEEETR-2011-0004*, Seattle, WA: University of Washington, 2011.

[15] D. Vishwanath. D The Epistemological Status of Vision Science and its Implications for Design, *Axiomathes*, 15:399–486, 2005.

[16] D. Vishwanath D. Towards a new theory of stereopsis, *Psychological Review*, 121:151–178, 2014.

[17] D. Vishwanath, and P.B. Hibbard. Seeing in 3D with just one eye: Stereopsis in the absence of binocular disparities. *Psychological Science*, 24:1673–1685, 2013.

#### 6. ACKNOWLEDGMENTS

This work was funded by a British Academy Mid-Career Fellowship to PH, and a University of Essex Undergraduate Research Opportunity Programme award to RH.

© IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.