



Reverse correlation reveals how observers sample visual information when estimating three-dimensional shape



Peter Scarfe^{a,*}, Paul B. Hibbard^{b,c}

^a Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB1 5QJ, United Kingdom

^b School of Psychology and Neuroscience, University of St. Andrews, St Mary's College, South Street, St. Andrews, Fife KY16 9JP, United Kingdom

^c Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom

ARTICLE INFO

Article history:

Received 12 July 2012

Received in revised form 14 March 2013

Available online 9 May 2013

Keywords:

Stereopsis

3D shape

ABSTRACT

Human observers exhibit large systematic distance-dependent biases when estimating the three-dimensional (3D) shape of objects defined by binocular image disparities. This has led some to question the utility of disparity as a cue to 3D shape and whether accurate estimation of 3D shape is at all possible. Others have argued that accurate perception is possible, but only with large continuous perspective transformations of an object. Using a stimulus that is known to elicit large distance-dependent perceptual bias (random dot stereograms of elliptical cylinders) we show that contrary to these findings the simple adoption of a more naturalistic viewing angle completely eliminates this bias. Using behavioural psychophysics, coupled with a novel surface-based reverse correlation methodology, we show that it is binocular edge and contour information that allows for accurate and precise perception and that observers actively exploit and sample this information when it is available.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Estimating three-dimensional object shape

The human visual system gains valuable information about the 3D structure of our environment from the fact that our eyes view the world from two slightly different positions (Julesz, 1971; Wheatstone, 1838). This difference means that in the shared field of vision a given point on an object typically projects to slightly different positions on the retina of each eye, producing horizontal and vertical retinal image disparities (Howard & Rogers, 2002). Horizontal disparities have been considered particularly important for depth perception because, with accurate information about object distance, they can be used to geometrically specify the full 3D structure of the environment (Hershenson, 1999; Johnston, 1991). Distance information is needed because the same pattern of horizontal disparity is consistent with infinitely many objects depending on the distance. Once distance information is known this depth ambiguity can be resolved.

Distance can be estimated from convergence (Brenner & van Damme, 1998), vertical disparity (Rogers & Bradshaw, 1993), or other cues (Hershenson, 1999), and could be used to scale horizontal disparities to specify shape accurately and unambiguously. Despite this possibility, observers typically show large systematic distance-dependent biases when estimating 3D properties of the

environment from binocular disparities. This causes the same object to have a dramatically different perceived 3D shape when placed at different distances from the observer (Johnston, 1991) and objects need to morph in shape when moving in depth to be perceived as physically constant (Scarfe & Hibbard, 2006). This has led some to question the utility of disparity as a cue to 3D shape (Pizlo, 2008; Pizlo, Li, & Steinman, 2008; Todd & Norman, 2003) and whether shape can be accurately recovered from disparity and other visual cues at all (Todd & Norman, 2003). Others have concluded that whilst the perception of 3D shape is typically biased, it can be accurately estimated, but only when observers are provided with large, continuous, perspective transformations of an object (Bingham & Lind, 2008).

This latter point raises a more general issue, namely that studies of the estimation of 3D shape from disparity often present static random dot stereograms at eye height, orientated face-onto a static observer. An example of such a stimulus, in this case a random dot stereogram of a cylinder, is shown in Fig. 1a. The aim of this type of viewing situation is to experimentally control the information available to the observer so as to constrain the ways that they could estimate 3D shape. Whilst this is in many ways a sensible experimental approach, it results in a highly unnatural viewpoint that is not at all characteristic of our natural interactions with objects in real life. It is therefore possible that the biases in perceived shape demonstrated in previous studies, using both computer generated and real world stimuli (Johnston, 1991; Watt et al., 2005), are due to the restricted information that is available to

* Corresponding author.

E-mail address: ps611@cam.ac.uk (P. Scarfe).

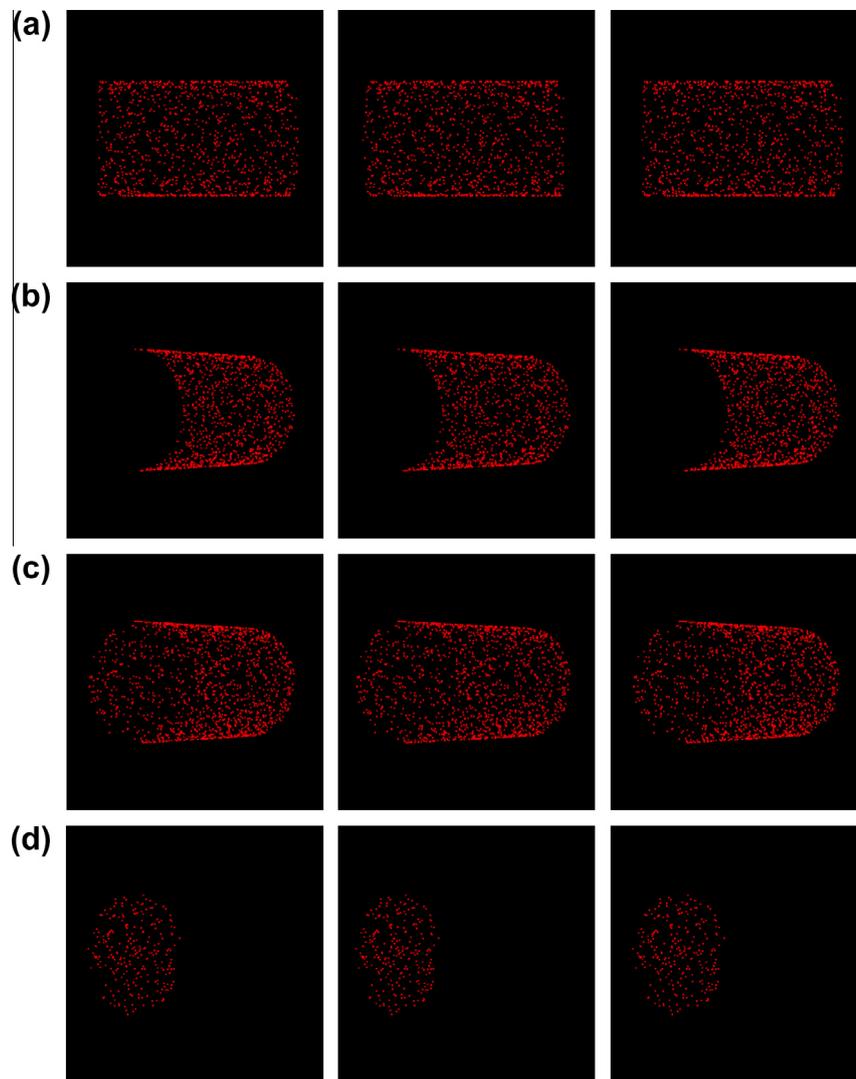


Fig. 1. Stereograms depicting the four stimulus conditions used in Experiment 1. (a) ACC condition (b) Non-Lidded condition (c) Lidded condition (d) Lid only condition. Left and middle columns for divergent fusion, middle and right columns for crossed fusion.

the observer, not some inherent inability to estimate 3D shape accurately (Todd & Norman, 2003).

The aims of the present study were therefore twofold. Firstly, we wished to determine whether veridical perception of 3D shape is possible with random dot stereograms, a type of stimulus that is routinely used to demonstrate the inaccuracies of 3D shape perception. Secondly, if accurate perception of shape is possible, we wished to determine what type of information allows for this. To achieve this we compared performance with the traditional face-on stimulus (Fig. 1a), to that obtained when cylinders were rotated around their vertical axis, either with (Fig. 1b), or without (Fig. 1c), a rendering of their “lid”. Rotating the cylinder is a fairly trivial change, but despite this fact, additional information becomes available that previous research suggests could be very informative as regards 3D shape.

1.2. Additional cues available with cylinder rotation

With cylinder rotation, observers gain a view of the abrupt disparity discontinuities present along the end contour of the cylinder. Disparity discontinuities arise where abrupt changes in the spatial gradient of disparity occur, such as at the corners or edges of a surface, or where two different surfaces abut one another

(Gillam, Blackburn, & Brooks, 2007). It is thought that the visual system might exhibit greater sensitivity to disparity discontinuities or higher derivatives of disparity gradients (Howard & Rogers, 2002; Stevens & Brookes, 1988). Consistent with this, the presence of disparity discontinuities has been shown to greatly increase both the speed of stereoscopic fusion and the accuracy of the ensuing 3D percept (in this case, stimulus slant) (Bradshaw, Hibbard, & Gillam, 2002; Gillam, Chambers, & Russo, 1988; Gillam, Flagg, & Finlay, 1984). As such it has been argued that edges and discontinuities are important primitives for stereopsis (Gillam, Chambers, & Russo, 1988). Indeed, cells in early cortical areas such as V2 of the macaque monkey have been shown to selectively respond to stereoscopic contours, edges and corners (von der Heydt, Zhou, & Friedman, 2000), and this selectivity is thought to provide valuable information to upstream cortical areas that are responsive to more complicated aspects of 3D structure (Janssen, Vogels, & Orban, 1999; Orban, Janssen, & Vogels, 2006).

In addition to information from disparity, a rotated view of a cylinder makes available, or alters, other cues useful for estimating 3D structure. The rotated cylinder’s body provides enhanced perspective cues that could be used to estimate its orientation (Hershenson, 1999; Howard & Rogers, 2002; Saunders & Backus, 2006; Saunders & Knill, 2001), and therefore possibly shape as

well. Additionally, the way texture elements in the projection of the cylinder are compressed differs greatly in a rotated, compared to a face-on, cylinder. This difference in textural compression could provide valuable information about curvature and shape (Fleming, Holtmann-Rice, & Bulthoff, 2011). These examples are in no way exhaustive, but demonstrate the substantial changes in information provided by only a small change towards a more naturalistic viewing situation.

1.3. Summary of the current study

In the present study we investigate whether the systematic distance-dependent biases found in the perception of 3D shape are in fact artefacts of forcing observers to view objects in a highly unnaturalistic viewing angle. We find that this is indeed the case. With only minor changes toward a more naturalistic viewing situation (cylinder rotation) we find that observers are able to veridically perceive 3D shape over changes in distance, where large systematic biases have been demonstrated previously (Johnston, 1991; Watt et al., 2005). Using behavioural psychophysics and a novel surface-based reverse correlation methodology, we show that the information that allows observers to do this is provided by the edges and contours of an object. We further show that, although monocular information is sufficient for accurate perception of shape, perceptual estimates are twice as precise when binocular information is available.

2. Experiment 1

The aim of Experiment 1 was to assess whether the systematic distance-dependent biases in the perception of 3D shape could be eliminated by the simple adoption of a more naturalistic viewing angle. All experiments reported in this paper were approved by the St. Andrews University Teaching and Research Ethics Committee.

2.1. Methods

2.1.1. Participants

Four observers took part in the experiment, the two authors (PBH and PS), who were experienced psychophysical observers, and two other observers, who were familiar with psychophysical research but naïve to the purposes of the experiments. All observers had normal or corrected to normal vision, and good stereopsis.

2.1.2. Apparatus

The stimuli were rendered online in OpenGL using Matlab and Psychophysics toolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007), and displayed on a 21" CRT monitor running at 100 Hz. The monitor was gamma corrected and spatially calibrated and ran at a resolution of 1024 × 864 pixels. The viewable screen measured approximately 40 cm in width and 35 cm in height. Head movements were minimized with the use of a head and chin rest, and observers were positioned such that a projection from the cyclopean eye intersected normal to the midpoint of the monitor screen.

2.1.3. Stimuli

The stimuli were random dot stereograms of elliptical hemi-cylinders, 6 cm in height and 10 cm in length. Stereoscopic presentation was achieved with the use of Crystal Eyes LCD shutter goggles that were synchronised to the refresh rate of the screen. The cylinders were rendered with anti-aliased red dots positioned with sub-pixel accuracy. Red was used as the shutter goggles have minimum cross-talk at longer wavelengths. The dot density of the cylinder

was 12 dots per square cm. The dots had a fixed radius of 1.5 pixels and were randomly positioned over the cylinder's surface. New dot coordinates were generated on each trial. All stimuli were rendered consistent with the inter-ocular distance of the observer. We specifically used stereo-defined random dot stimuli as these have consistently been shown to elicit large distance-dependent perceptual biases from observers (Bradshaw, Parton, & Glennerster, 2000; Brenner & Landy, 1999; Glennerster, Rogers, & Bradshaw, 1996; Johnston, 1991; Watt et al., 2005).

2.1.4. Procedure

Observers completed an apparently-circular-cylinder (ACC) task in which they judged whether the disparity-defined cylinder they were presented with was stretched or squashed in depth extent relative to a cylinder with a circular cross-section (Johnston, 1991). As in previous studies, cylinders were presented at eye height directly in front of the observer over a range of viewing distances. The background of the screen was black and observers viewed the monitor in complete darkness. To minimise dark adaptation observers were required to have light breaks between sessions. Cylinders were presented at distances of 40, 60, 80 and 100 cm.

The monitor was always physically positioned at the vergence-specified distance of the rendered 3D stimuli to minimise conflicting focus cues, which could otherwise signal a conflicting viewing distance. Watt et al. (2005) have shown that under these viewing conditions focus cues have no measurable effect on the accuracy of perceived slant in random dot stereograms, so we expect a minimal effect in our stimuli. Furthermore, from previous studies we can make the clear prediction that any bias shown will be *opposite* to that predicted by conflicting focus cues, as at these near distances (below approx. 1 m) observers robustly over-estimate, rather than under-estimate, object depth (see Howard and Rogers (2002) for a comprehensive summary and Scarfe and Hibbard (2011) for our previous discussion of this issue).

We used four conditions to investigate the role of edge and contour information. In the "ACC" condition (Fig. 1a) we aimed to replicate the viewing conditions in which large biases in perceived shape have been demonstrated previously by presenting the cylinder fronto-parallel to the observer (Bradshaw, Parton, & Glennerster, 2000; Glennerster, Rogers, & Bradshaw, 1996). In the "Non-Lidded" (Fig. 1b) and "Lidded" (Fig. 1c) conditions the cylinder was rotated around its central vertical axis by 50° (direction of rotation counterbalanced across observers). This provided the observer with additional information from the end contour of the cylinder, and in the case of the "Lidded" condition, its elliptical "lid". Finally, in the "Lid only" (Fig. 1d) condition we presented only the elliptical lid of the cylinder, not its body. This provided a view of the cylinder's end contour, but cues from its body were absent.

Trials were completed in 16 blocks (4 viewing conditions by 4 viewing distances). Within a block we varied the depth of the cylinders using the method of constant stimuli. There were five depth values and each was presented 20 times, in a random order. The exact depth values depended on the observer and block type and were determined on the basis of pilot experiments. Each trial started with the presentation of a red fixation point 6 pixels in diameter in the centre of the screen. This was followed by presentation of the cylinder for 2 s. The screen then went blank signalling that the observer should respond. On making a response, the presentation of the next trial was triggered. Blocks were completed in a different randomised order for each observer.

2.2. Results

Cumulative Gaussian functions were fitted to observers' responses and the point of subjective equality (PSE) and 95%

confidence intervals around this value determined in Matlab using the psignifit software package (Wichmann & Hill, 2001a, 2001b). The PSE represents the cylinder depth needed for observers to perceive it to be circular; we will refer to this as the observer's "shape setting". Shape settings across our four observers are shown in Fig. 2a. For the fronto-parallel ACC condition we were able to replicate the distance-dependent bias found in previous studies (Bradshaw, Parton, & Glennerster, 2000; Johnston, 1991; Johnston, Cumming, & Landy, 1994). As with this previous research, we too found that at these near distances (below 1 m) observers overestimated object depth relative to height, with shape being accurately estimated at a distance of around 1 m (Howard & Rogers, 2002).

In contrast, both conditions in which rotated views of the cylinder's body and end contour were available (Lidded and Non-Lidded) resulted in the elimination of this bias. This was the case whether or not the cylinder's elliptical lid was rendered. When only the elliptical end of the cylinder was rendered, distance-dependent bias was again present. We assessed the significance of our data using repeated-measures ANOVA and linear regression; this confirmed what is evident in Fig. 2a. Shape settings were significantly affected by the type of cylinder $F_{(3,18)} = 5.66, p < 0.05$, and the distance at which it was viewed $F_{(3,18)} = 24.08, p < 0.001$. There was also a significant distance-by-type interaction $F_{(9,18)} = 13.97, p < 0.0001$. Linear regression showed that this interaction arose because there was a significant effect of distance in the ACC ($p < 0.01$) and Lid-Only ($p < 0.05$) conditions, but not in the Lidded ($p = 0.51$) and Non-Lidded ($p = 0.83$) conditions.

As an extension of this experiment we assessed the magnitude of rotation needed to eliminate the bias of perceived shape, by repeating the experiment using two additional rotation angles with the Non-Lidded and Lidded cylinders, at the 40 cm viewing distance, where the largest perceptual bias was found. Three observers from the first part of Experiment 1 took part in this second part (CG, LO and PBH). In Fig. 2b we plot shape settings for rotation angles of 16.67°, 33.33° and 50°. As is clearly evident, the distance-dependent bias is completely eliminated even at the smallest rotation angle we tested. Linear regression showed that there was no effect of rotation angle on shape settings for either the Non-Lidded ($p = 0.24$), or Lidded ($p = 0.89$), conditions.

2.3. Discussion

Overall the results of Experiment 1 suggest that with only a minor change in viewing angle observers are able to veridically estimate

the 3D shape of random dot stereograms of elliptical cylinders over a distance range where significant perceptual bias has been demonstrated previously (Johnston, 1991). This suggests that large continuous perspective transformations are not a prerequisite for the accurate perception of 3D shape (Bingham & Lind, 2008).

3. Experiment 2

In Experiment 2 we use a reverse correlation technique to investigate how observers actively sample and exploit the available visual information when making their estimates of 3D shape with face on and rotated cylinders. The aim was to identify the information that allows for accurate perception of 3D shape.

3.1. Methods

3.1.1. Participants

The same observers from Experiment 1 took part in Experiment 2.

3.1.2. Apparatus

The same apparatus was used as described in Experiment 1.

3.1.3. Stimuli

We used the same type of stimuli as Experiment 1, but only the ACC and Non-Lidded conditions. Observers were again presented with a series of cylinders and were asked to decide if these were stretched or squashed in depth extent relative to a cylinder of circular cross-section. Cylinders were presented at the 40 cm viewing distance only, as it was at this distance that the greatest difference between conditions was observed. All cylinders were tailored to be perceptually circular to the observer. We achieved this by using the shape setting data from Experiment 1. In addition to this, we randomly perturbed the depth coordinate of the dots defining the cylinder on each trial. This was done prior to rotation in the case of the Non-Lidded condition. The depth noise added to each point was randomly chosen from a Gaussian distribution centred on zero with a standard deviation set equal to the observer's just noticeable difference (JND) for this condition from Experiment 1, multiplied by the radius of a physically circular cylinder.

3.1.4. Procedure

Observers were presented with a series of disparity-defined cylinders, which had "noisy" depth coordinates, but which would be

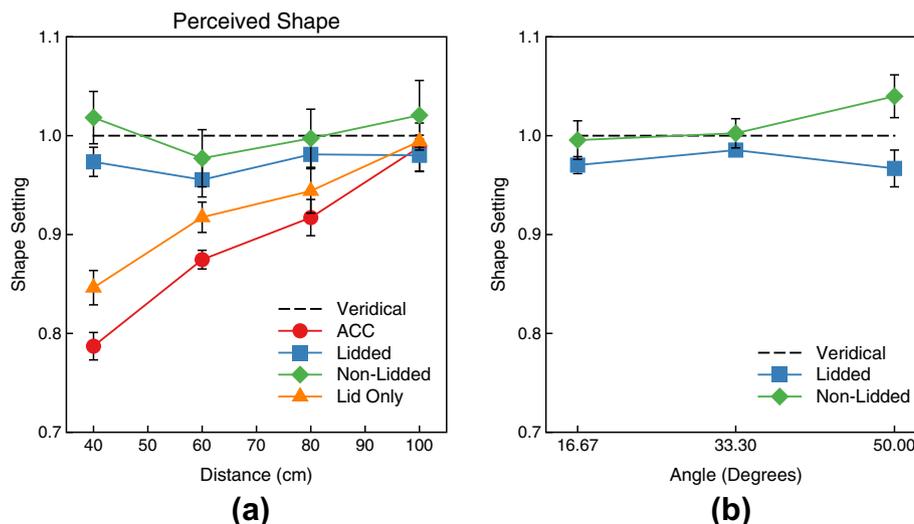


Fig. 2. (a) Mean shape settings made by observers over the four distances tested, (b) mean shape settings made by observers over the three rotation angles tested at the 40 cm viewing distance. The horizontal dashed line shows veridical perception of shape and the error bars show standard error of the mean (SEM).

on average perceptually circular. Each type of cylinder was presented on 2000 trials split into 20 blocks of 100. In those instances where observers completed more than one block in a sitting, they were required to take light breaks between blocks to minimise dark adaptation. The timings and trial presentation were otherwise identical to Experiment 1. By recording (a) the dot coordinates of the cylinders, (b) the value of noise added to each coordinate, and (c) the observer's response, we were able to build classification images of the cylinder's surface that revealed those surface regions from which observers actively utilised information when making their perceptual estimates (Ahumada, 1996; Beard & Ahumada, 1998).

3.2. Results

3.2.1. Construction of the surface-based classification images

To create the classification images for a given observer, we divided the trials for each cylinder type (ACC and Non-Lidded) into those where the observer responded that the cylinder was stretched in depth extent relative to a circular cylinder, and those where the observer responded that the cylinder was squashed in depth extent relative to a circular cylinder. This gave us four sets of trials (1) ACC stretched in depth extent, (2) ACC squashed in depth extent, (3) Non-Lidded stretched in depth extent, and (4) Non-Lidded squashed in depth extent. For each trial we had a record of the X , Y and Z coordinates of the dots that had defined the cylinder's surface, and the noise value that had been added to the Z coordinate of each dot.

We defined a pixel grid across a circular cylinder's surface with 19 bins for angular position ($\pm 90^\circ$) and 19 bins for the X position (± 5 cm). Four pixel bin maps were created, one for each of the four sets of trials. Each cylinder was on average perceptually circular for the observer, so we could use the X coordinate for the dot and its angular position on the cylinder's surface, defined by the dots Y and Z coordinates, to bin the dot coordinates to create the classification images. For each cylinder dot we determined which pixel the dot coordinate fell in, and binned the Z noise value for the dot into the appropriate pixel map. For each map we then

calculated the average Z noise value in each pixel, after having filtered out outlying noise values.

These were defined as values more than three standard deviations away from the mean of the Gaussian distribution from which we drew the noise (for that observer and that cylinder type). This insured that the pixel averages were not unduly influenced by rare, large, outlier noise values (on average 0.56% across observers). This gave us four pixel maps showing the average noise per pixel for each of our four sets of trials. For the two cylinder types we then took the absolute difference between the pixel maps for the "stretched in depth extent" and "squashed in depth extent" judgements. This gave us two pixel difference maps, which we refer to as our *classification images* for each type of cylinder. These give us a linear approximation of the observer's decision rules when making their shape estimates.

The classification images were then normalised such that their values were scaled to lie between 0 and 1. The maximum and minimum values used to do this scaling were calculated separately for each type of cylinder (ACC or Non-Lidded). This allowed us to see those parts of each type of cylinder that were most important when estimating 3D shape. The classification images were then scaled in size by a factor of 50 using nearest neighbour interpolation and smoothed by convolving the image with a 60 pixel square 2D Gaussian with a standard deviation of 20 pixels. The images were then normalised again, as described before. Because two observers had viewed cylinders rotated by $+50^\circ$ and two by -50° , we left-right flipped the classification images of LO and PS to make the cylinder orientation comparable across observers.

3.2.2. Information sampling across a cylinder's surface

Classification images for the ACC and Non-Lidded cylinders are shown in Figs. 3 and 4 respectively. In both instances we present the images in flat-map form and texture-mapped onto the surface of a rotated cylinder. Mean classification images across observers, for both types of cylinder, are shown in Fig. 5a and b. As can be seen there are striking differences in how disparity is utilised in the two viewing conditions. For the Non-Lidded cylinders the "hot spots" are primarily clustered along the cylinder's end

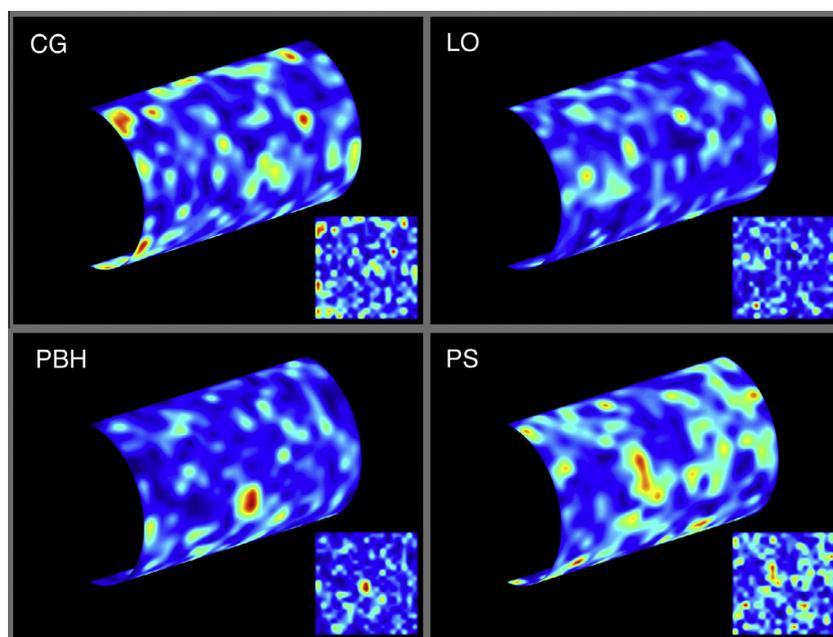


Fig. 3. Classification images for the ACC cylinders for each observer. Images are shown texture-mapped onto a 3D representation of a rotated cylinder, and in flat map form in the insets. Note that for this condition the cylinders were viewed fronto-parallel to the observer as in Fig. 1a.

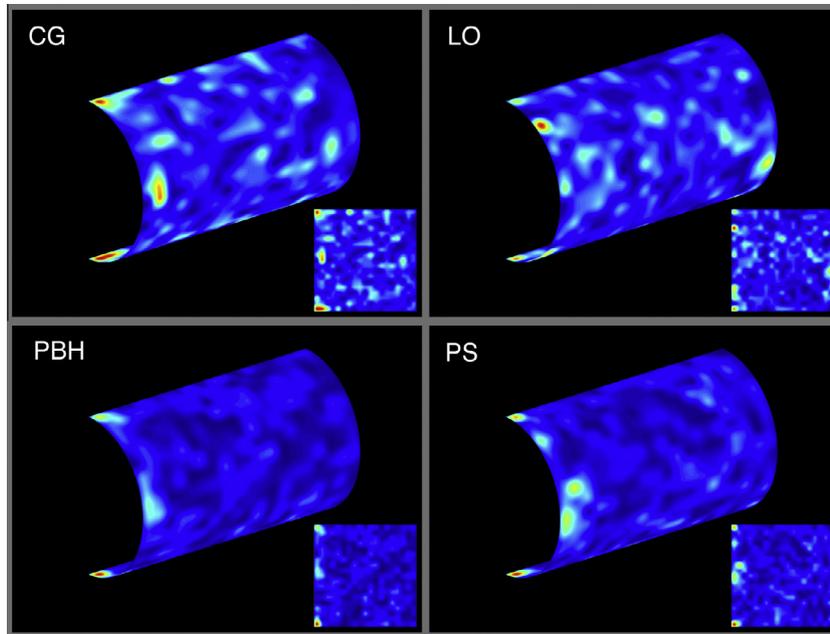


Fig. 4. Classification images for the rotated cylinders for each observer (conventions as in Fig. 3).

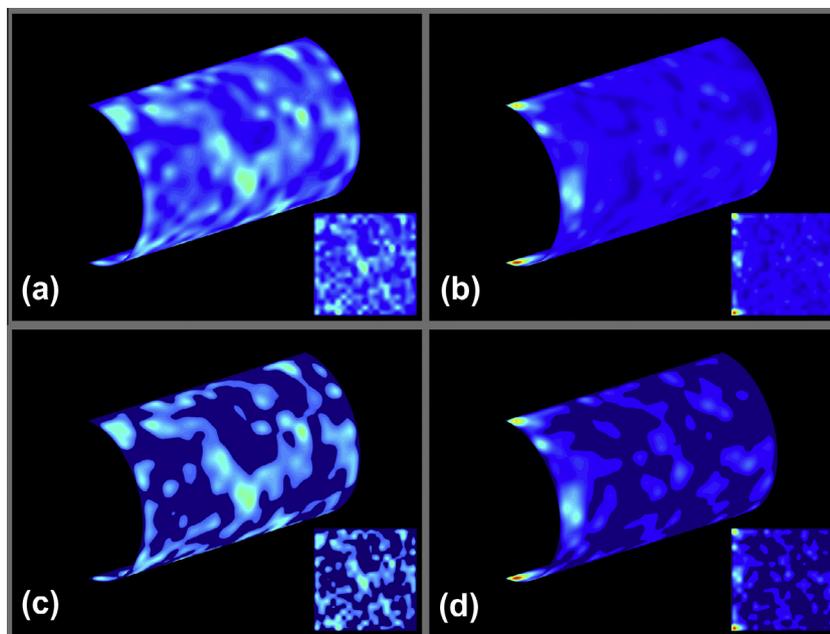


Fig. 5. Mean classification images for the (a) ACC (fronto-parallel) and (b) Non-Lidded (rotated) cylinders. Thresholded versions of these are shown in (c) and (d). See accompanying text for details.

contour, particularly at the upper and lower corners. By comparison, the hotspots for the ACC cylinders are more dispersed, but broadly clustered around fixation. These differences are more evident in Fig. 5c and d, where we have thresholded the mean classification image for the two types of cylinder to highlight the location of the most prominent hotspots (threshold set as the mean of each classification image).

In order to assess the likelihood with which we would have obtained the hotspots shown in Fig. 5 by chance we simulated the whole experiment and analysis 1000 times with simulated observers who randomly guessed on each trial. Our simulated observers

had the same PSE's and JND's as the observers in the experiment and we used the same scaling values for normalisation as were used for our real observers; this allowed us to produce simulated classification images directly comparable to the experimental data. These are shown in Fig. 6. As can be seen the simulated classification images for both the ACC and Non-Lidded condition are a uniform blue, indicating that that the hotspots found in the experimental classification images are unlikely to arrive by chance. None of the 1000 simulated classification images produced the same or greater number of 'hot' pixels, or reached the same or greater mean pixel 'hotness' of the experimental classification

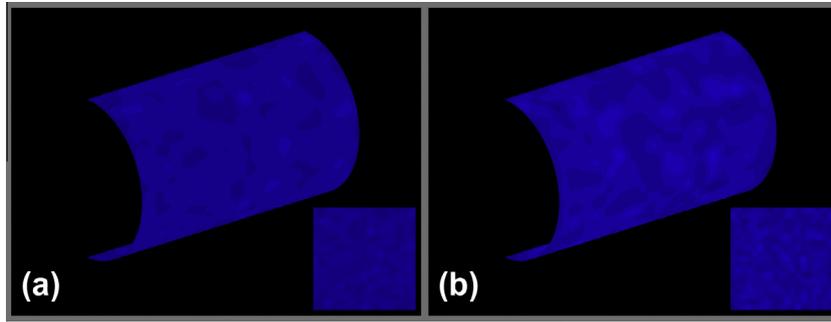


Fig. 6. Mean classification images produced by a simulation of observers' performance in the experiment if they were all guessing at chance on each trial for the (a) ACC and (b) Non-Lidded cylinders.

images. This, together with the consistency of the classification images across observers (Figs. 3 and 4), shows that the hotspots represent clear differences in how visual information was utilised by observers across conditions.

3.2.3. Where on the cylinder's surface is most informative about 3D shape?

In an ideal situation we would like to have a full computational understanding of how observers estimate 3D shape from the information contained in our stimuli. Suffice to say that given the variety of types information available, this is not possible. One place to start however is to analyse how the information carried by horizontal and vertical retinal image disparities varies across the cylinder's surface. Numerous studies have shown these cues to be important in the estimation of 3D depth and shape (Howard & Rogers, 2002), particularly in random dot stereograms, which are designed, in part, to isolate these cues (Julesz, 1971; Zabulis & Backus, 2004). To do this we performed Monte Carlo simulations of the stimuli used in the reverse correlation to find out where across the cylinder's surface the most informative disparity information is available.

First, for each of our four observers (with their specific interocular distance and JND's) and for each type of cylinder (ACC and Non-Lidded), we defined a 38 by 38 grid of evenly spaced points across a circular cylinder's surface. The size, distance and rotation angle of the simulated cylinders was matched to that used in the reverse correlation. The horizontal and vertical disparity of each point on the cylinder's surface was then calculated. For vertical disparity we used Fick coordinates (Read, Phillipson, & Glennerster, 2009). We refer to these values as our 'baseline values'. For 5000 subsequent simulation runs, for each cylinder type and for each observer (i.e. 40,000 runs in total), we randomly perturbed each of the original surface points in depth within the same range used for the reverse correlation (i.e. scaled to each observers JND). The

resultant horizontal and vertical retinal image disparities were then calculated and the absolute difference between these and our baseline taken.

The 5000 'disparity difference maps' for each observer were then averaged, resulting in one map for horizontal disparity and one for vertical disparity per observer. These average disparity difference maps were then combined, by calculating an overall disparity magnitude, θ_o , given by:

$$\theta_o = \sqrt{\theta_H^2 + \theta_V^2}$$

where θ_H is the point's horizontal disparity and θ_V the point's vertical disparity. This vectorization gives an overall picture of how much disparity in general varies with changes in local depth. The map values were then normalised to lay between 0 and 1, enlarged by a factor of 25 using bicubic interpolation, then normalised again, allowing for direct comparison with the classification images. This gives the final disparity difference maps, for the ACC (Fig. 7a) and Non-Lidded (Fig. 7b) cylinders. By comparing these to the classification images in Fig. 5 it is clear that, for both types of cylinder, the hot spots identified during the reverse correlation lay in those areas of the cylinder that are most informative about changes in disparity. This suggests that, in making their decisions about 3D shape, observers actively sample, or weight, those local areas of an object's surface where changes in object depth result in the greatest changes in the measurable disparity signal.

3.2.4. Comparing the two types of cylinder directly

Whilst the analysis so far has served to emphasise the information availability and use for each type of cylinder (face onto the observer or rotated), it did so at the expense of comparing both types of cylinder with a common scale. This is because the normalisation of image values was carried out separately for each type of cylinder. Therefore, to compare the cylinders on a common scale, new experimental classification images and simulated disparity

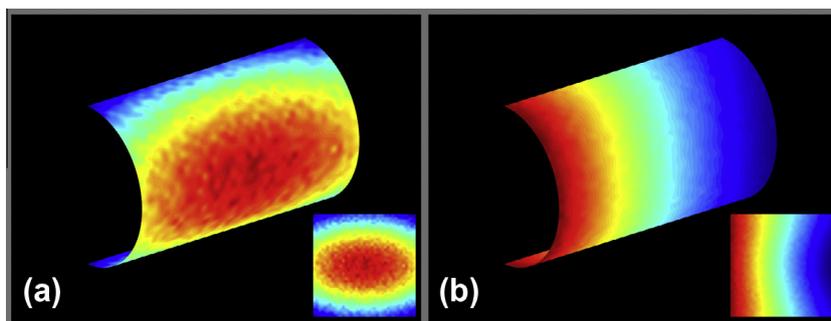


Fig. 7. "Disparity difference" maps for the (a) ACC and (b) Non-Lidded cylinders, showing where disparity cues change most with local changes in object depth.

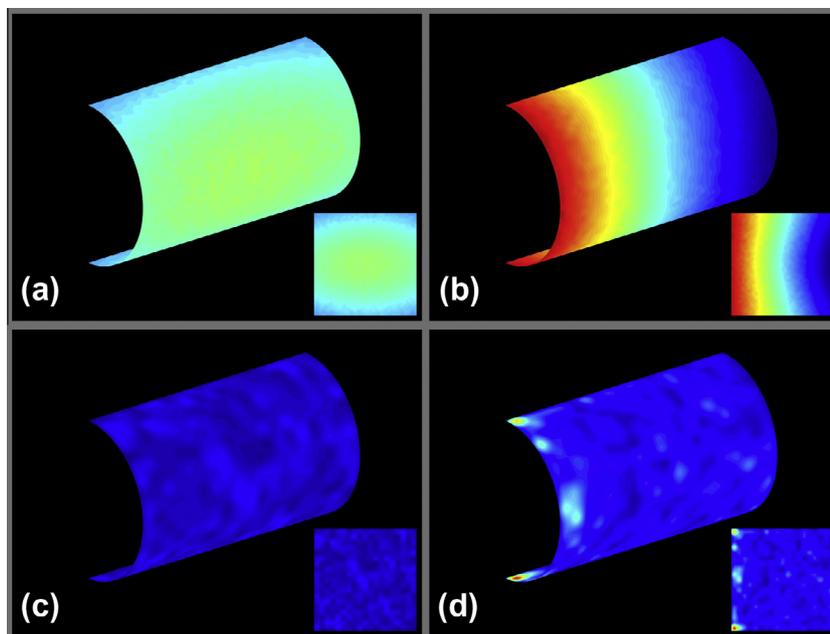


Fig. 8. Disparity difference maps for the (a) ACC and (b) Non-Lidded cylinders, normalized with values calculated across both types of cylinder, and mean experimental classification images for the (c) ACC and (d) Non-Lidded cylinders, again normalized with values calculated across both types of cylinder. See accompanying text for details.

difference maps were created using the maximum and minimum values calculated across *both types of cylinder* for normalisation (Fig. 8). From these images it is clear that disparity varies more with changes in local depth for the Non-Lidded compared to ACC cylinders (Fig. 8a and b) and this is reflected in greater differential use of disparity information by observers (Fig. 8c and d). This also explains why the hotspots in the original classification images for the ACC cylinders are less localised than the Non-Lidded cylinders.

3.3. Discussion

Overall our results and analysis suggest that observers are highly attuned to the quality of information available across an object's surface. This is consistent with previous work on cue combination for the perception of slant where it has been shown that observers weight disparity and texture cues depending on the local reliability of each (Hillis et al., 2004). This implies that if we were to present *only* the most informative parts of the cylinder's surface, where (a) disparity varies most greatly with changes local depth, and (b) observers actively sample information when estimating shape, observers' perceptions of 3D shape should be unaltered from when the full cylinder's surface is visible. This would also rule out the possibility that during the reverse correlation experiment observers were simply tracking those areas of the surface where disparity varied most greatly on a trial-by-trial basis and not exploiting the information that best allowed them to accurately estimate shape.

4. Experiment 3

The aim of Experiment 3 was to investigate whether observers' perception of shape is equally veridical across changes in distance if we present only the most informative parts of a cylinder's surface, as defined by the results and analysis of Experiment 2. This is most readily testable in our Non-Lidded cylinders as these show the largest localised variation in disparity across the cylinder's surface and the greatest differential sampling of disparity. Moreover, it is with these stimuli that we have been able to show veridical

perception of depth where systematic distortions have been previously reported (Johnston, 1991).

4.1. Methods

4.1.1. Participants

Three observers took part in Experiment 2, one of the authors (PBH) and two others (AM and RW) who familiar with psychophysical research but naïve to the purposes of the experiment. All observers had normal or corrected to normal vision, and good stereopsis.

4.1.2. Apparatus

The apparatus was the same as that used previously.

4.1.3. Stimuli

The stimuli for this experiment consisted of shortened versions of the Non-Lidded stimuli used in Experiment 1. For the first three conditions, rather than being 10 cm in length, the cylinders were 1 cm (Fig. 9a), 2 cm (Fig. 9b) or 3 cm (Fig. 9c) in length. As before, the cylinders were rotated by 50° (positive for AM and negative for PBH and RW). In all cases the nearest part of the cylinder to the observer was rendered. In all other respects, these "puck" stimuli were identical to that described previously. In a final condition, we used an outline stimulus where only a solid line tracing the cylinder's nearest end contour was rendered (Fig. 9d).

4.1.4. Procedure

The procedure was identical to that described in Experiment 1.

4.2. Results

As in Experiment 1, Cumulative Gaussian functions were fitted to observers' responses and the point of subjective equality (PSE) and 95% confidence intervals around this value determined in Matlab using the *psignifit* software package (Wichmann & Hill, 2001a, 2001b). The PSE represents the cylinder depth needed for observers to perceive it to be circular; we will refer to this as the

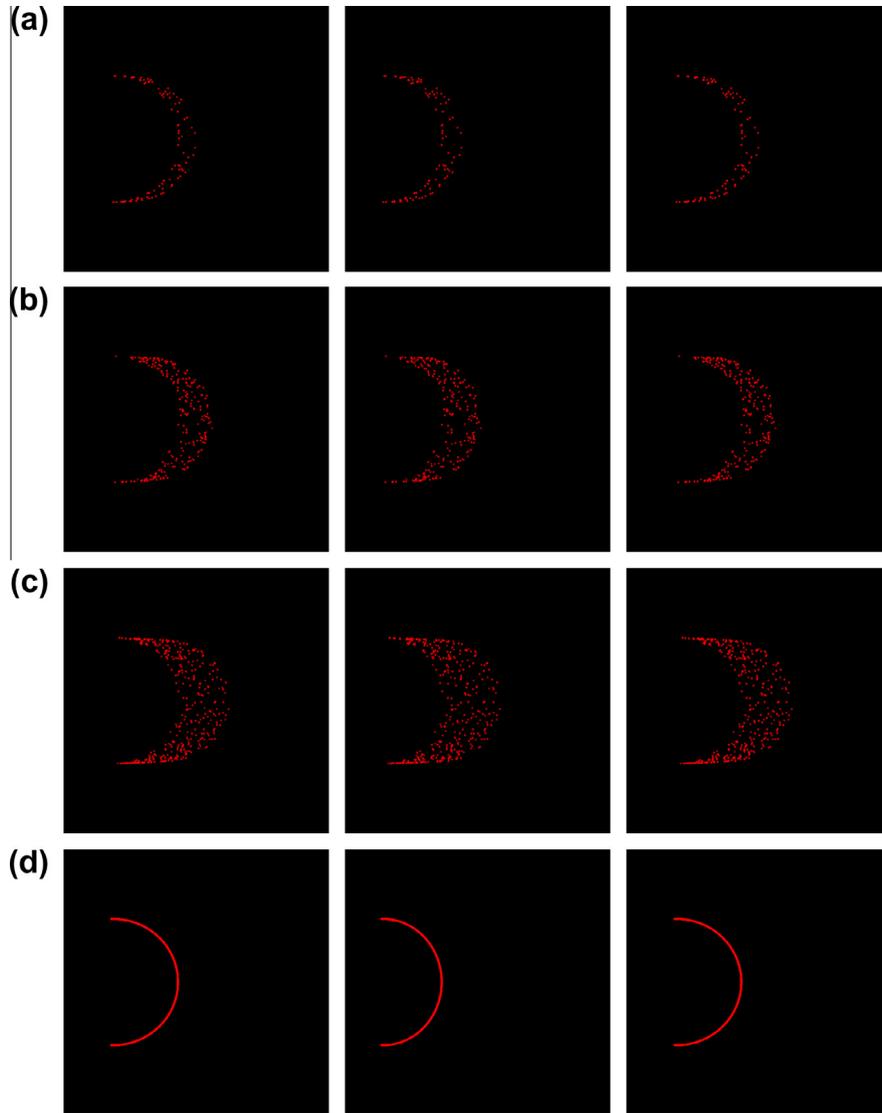


Fig. 9. Stereograms depicting the four stimulus conditions used in Experiment 3. (a) 1 cm Puck (b) 2 cm Puck (c) 3 cm Puck, and (d) Outline. Left and middle columns for divergent fusion, middle and right columns for crossed fusion.

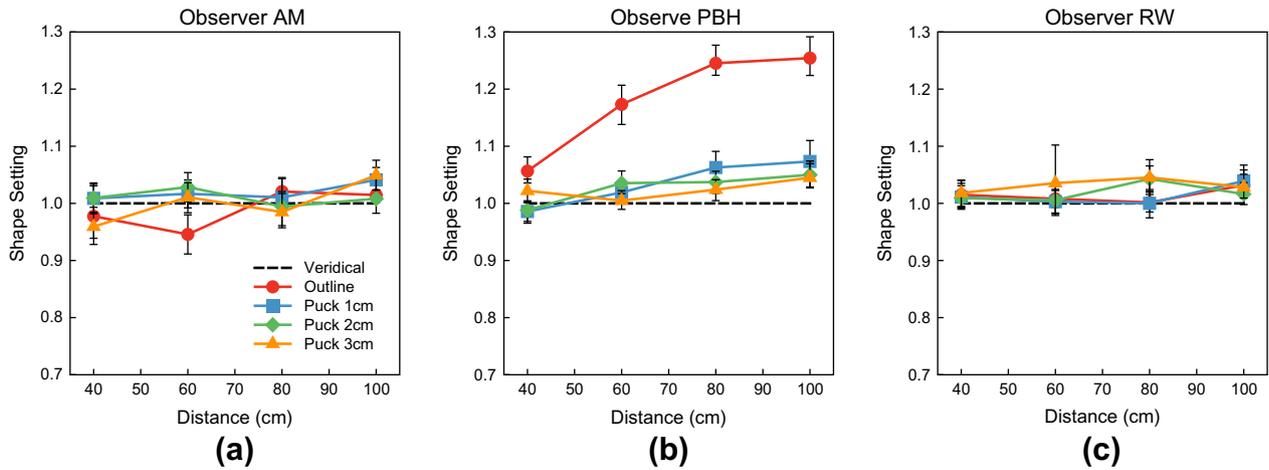


Fig. 10. Shapes settings for the shortened “puck” cylinders, and the outline only stimulus. Error bars show 95% confidence intervals bootstrapped from the psychometric function fit.

observer's "shape setting". Shape settings for the four conditions are shown in Fig. 10, with a separate graph for each observer. As can be seen, all three observers were able to veridically estimate shape over the full distance range with shortened versions of the Non-Lidded cylinders (Puck 1 through 3). Two out of three observers were also able to estimate shape veridically with only an outline of the cylinder's end contour. Observer PBH was unable to do this and reported no clear impression of depth.

4.3. Discussion

Observer's veridical estimation of shape over changes in distance with shortened cylinders is consistent with the results of Experiments 1 and 2, which showed that it is in these surface regions of the cylinder where disparity varies most with changes in local depth, and that observers preferentially sample visual information from these areas when estimating 3D shape. That two out of three observers could veridically estimate shape with only an outline stimulus was somewhat surprising given the biases in perceived shape shown in the "Lidded" stimuli from Experiment 1. We suggest that the outline stimulus used in the present experiment provided enhanced information about the cylinder's end contour allowing for accurate estimation of shape compared to that of the lidded stimuli used in Experiment 1. In Experiment 1 the cylinder's end contour, had to be *inferred* from the relatively sparse stereogram of the cylinder's lid. This explanation is consistent with previous work showing that, even *with* an oblique view of a cylinder (which gave veridical perception on shape in Experiment 1), observer's perception of shape can still be biased so long as end contour information is selectively degraded (Scarfe & Hibbard, 2006).

5. Experiment 4

The results of Experiments 1 through 3 suggest that when visual information from the edges and contours is available in binocularly viewed cylinders, observers actively exploit this information and in doing so eliminate the large systematic distance-dependent biases of perceived shape documented previously. Experiment 4 sought to determine the extent to which this improvement is due to monocular or binocular visual information. Edge and contour information is available in both situations, so it is possible that observers are using monocular cues alone. If this were the case observers should be equally accurate and precise when estimating shape with a single eye's view as compared to binocular viewing.

5.1. Methods

5.1.1. Participants

Observers PBH and LO took part in this experiment along with two new observers (RW and SN) who were both naïve to the purposes of the experiment. All observers had normal or corrected to normal vision, and good stereopsis.

5.1.2. Apparatus

The apparatus used in this experiment was identical to that described previously.

5.1.3. Stimuli

The stimuli were the same as those used in the Lidded condition of Experiment 1, except that they were now presented monocularly. To do this the observers wore an eye patch over the left eye and the right eye's image alone was presented on each frame.

5.1.4. Procedure

The procedure was identical to that used in Experiments 1 and 3. As observers RW and SN were new observers, in addition to the monocularly viewed Lidded stimuli of the current experiment, they also collected data for the Lidded and ACC conditions of Experiment 1. All aspects of data collection for these conditions were identical to that described previously. We specifically chose monocular presentation to compare to binocular, rather than bi-ocular (same image to both eyes) for four reasons, all of which would impact the interpretation of our results; (1) bi-ocular presentation never occurs in real life, (2) bi-ocular presentation gives a strong cue to stimulus flatness, especially in random dot stereograms, (3) bi-ocular presentation has been shown not to lead to an improvement in the apparent depth afforded by monocular cues (Bradshaw et al., 2004), and (4) bi-ocular presentation can result in distortions in apparent depth (Koenderink, van Doorn, & Kappeers, 1994). All of these problems are avoided by comparing monocular to binocular viewing.

5.2. Results

Cumulative Gaussian functions were fitted to observers' responses and the point of subjective equality (PSE) and 95% confidence intervals around this value determined in Matlab using the psignifit software package (Wichmann & Hill, 2001a, 2001b). Mean shape settings for the Monocular Lidded condition are shown along with those for the Lidded and ACC conditions in Fig. 11a, in addition to this we show the average slope of the psychometric

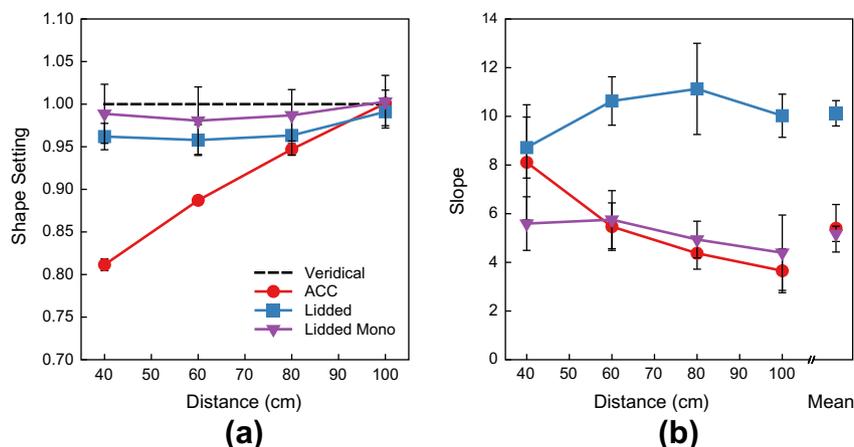


Fig. 11. (a) Mean shape settings and (b) psychometric function slopes, for the ACC, Lidded and Lidded Monocular cylinders. Error bars show standard error of the mean (SEM).

function fitted to observer's data in Fig. 11b, as well as the mean across all distances. These slopes provide a measure of observers' precision in each condition.

Two things are immediately apparent from the data. Firstly, observers can gain the same level of veridicality (lack of bias) in estimating the shape of "Lidded" cylinders whether the information about contours and edges is carried by luminance alone (Monocular Lidded) or disparity and luminance (Lidded). ANOVA showed a main effect of the distance at which the cylinder was viewed $F_{(3,18)} = 26.82, p < 0.001$ and no main effect of cylinder type $F_{(2,18)} = 3.49, p = 0.1$. These main effects were modified by a significant distance-by-type interaction $F_{(6,18)} = 17.03, p < 0.0001$. Linear regression showed that this interaction arose because there was a significant effect of distance in the ACC condition ($p < 0.01$) but not the Lidded (0.21) and Lidded Monocular (0.33) conditions.

Secondly, the precision with which observers could estimate shape accurately with Lidded cylinders was greatly impaired when binocular information was absent. ANOVA showed that slopes were significantly affected by the type of cylinder $F_{(2,18)} = 26.46, p < 0.001$ but not the distance at which they were viewed $F_{(3,18)} = 0.92, p = 0.47$. There was no distance by type interaction $F_{(6,18)} = 2.47, p = 0.06$. Slopes for the ACC and Lidded Monocular condition did not differ ($p = 0.71$), however slopes for the Lidded cylinders were significantly greater than both the ACC ($p < 0.01$) and Lidded Monocular ($p = 0.01$) cylinders. Indeed, compared to the Lidded cylinders, slopes were half as steep when only monocular information was available.

5.3. Discussion

The results of Experiment 4 suggest a specific role for binocularly defined edges and contours in estimating 3D shape. Monocular edges and contours allow for accurate performance, but at the cost of halving the precision with which 3D shape can be estimated. This further emphasises that even with a highly impoverished stimulus (compared to real life), accurate and precise estimation of shape is possible with only a trivial change to a more naturalistic viewing situation.

6. General discussion

6.1. Summary of results

The utility of disparity as a cue to 3D shape has been questioned due to the systematic distance-dependent biases that observers exhibit (Johnston, 1991; Watt et al., 2005). This has led some to question whether the veridical perception of 3D shape is possible at all (Todd, 2004; Todd & Norman, 2003), whereas others have suggested that accurate perception of shape is possible, but only if observers are presented with large continuous perspective transformations of an object (Bingham & Lind, 2008). In the present study we have shown that both of these conclusions need revising. Observers were able to accurately and precisely estimate 3D shape from random dot stereograms of elliptical cylinders, a stimulus that has routinely been used to demonstrate perceptual distortions, so long as binocular information from the edges and contours of the cylinders was available. We provided this information with a trivial stimulus manipulation: slightly rotating an object to a more naturalistic viewing angle, so that it was no longer orientated face-onto the observer.

We used a novel surface-based reverse correlation methodology, along with additional behavioural experiments, to show that it was specifically *binocular* edge and contour information that allowed for accurate and precise performance. Observers' shape judgements were equally unbiased when edge and contour information was provided monocularly, but they could estimate shape

with only half the precision. Overall, our results suggest that the systematic distortions in perceived 3D shape demonstrated repeatedly over the last few decades occur only in a highly restricted, reduced cue situation (Johnston, 1991; Watt et al., 2005). This is consistent with previous research questioning the extent to which inferences about real-world performance can be made on the basis of experiments conducted in reduced cue experiments (Mon-Williams & Bingham, 2008).

6.2. Models of sensory cue combination

At present a full computational understanding of how cues are combined to estimate 3D shape is a far away goal. In this section we consider two models of sensory cue combination and how they may account for our data. Our aim is to highlight aspects of our data that cannot readily be accounted for by these models.

6.2.1. Linear weighted averaging

The weighted averaging model of cue combination suggests that the optimisation criterion the human sensory system adopts when combining cues is minimising the variance of the combined-cues estimate (Landy et al., 1995). In its simplest form, if we consider two cues to shape, one monocular S_M and one binocular S_B , the combined cues estimate of shape, S_C , is given by:

$$S_C = w_M S_M + w_B S_B \quad (1)$$

Here, the weights given to monocular and binocular information, w_M and w_B , are determined by the relative reliability of each cue. Reliability is defined as the inverse of the variance associated with each cue (for a more detailed discussion see Landy et al., 1995). It provides a good account of cue combination in some situations (Ernst & Banks, 2002; Ernst & Bulthoff, 2004; Helbig & Ernst, 2007), but not others (Fetsch et al., 2009; Oruc, Maloney, & Landy, 2003; Rosas et al., 2005; Rosas, Wichmann, & Wagemans, 2007; Serwe, Drewing, & Trommershauser, 2009; Zalevski, Henning, & Hill, 2007).

One reason why weighted averaging may fail to predict combined cues performance is if the weights assigned to single cue estimates do not reflect the reliability of those same cues in the combined cues setting. This can occur if discrimination thresholds are a poor measure of the underlying reliability of a cue (Todd, Christensen, & Guckes, 2010) or if cues are not conditionally independent (Oruc, Maloney, & Landy, 2003). This latter case may occur if the use of a cue is yoked to the mode of viewing. For example, texture cues defining an edge or contour may be better measured binocularly than monocularly. This would mean that there would be no such thing as a "pure" texture cue, only "texture-as-viewed-monocularly" and "texture-as-viewed-binocularly". This is a more general problem for all types of cue, especially when arising from the same sensory modality. As such, there may never be a pure "single cue estimate" of a property (see Zabulis & Backus, 2004 for an excellent discussion of this issue).

In many instances it is also unclear what people are estimating when making their judgements. When it was proposed, the stated goal of weighted averaging was to build a full metric depth map of the scene (Landy et al., 1995). This suggested the units the observer uses when combining cues is one of depth, and that all cues should be "promoted" to these units before combination (Burge, Fowlkes, & Banks, 2010; Landy et al., 1995). However, weighted averaging has been applied to numerous properties such as slant (Hillis et al., 2004), angle between planes (Watt et al., 2005), size (Glennester et al., 2006), 3D location (Svarverud, Gilson, & Glennester, 2010), curvature (Curran & Johnston, 1994) and shape (Helbig & Ernst, 2007). An added problem is that in many instances the same task can be completed in multiple ways, so it is difficult, *a priori*, to know what computations the brain is

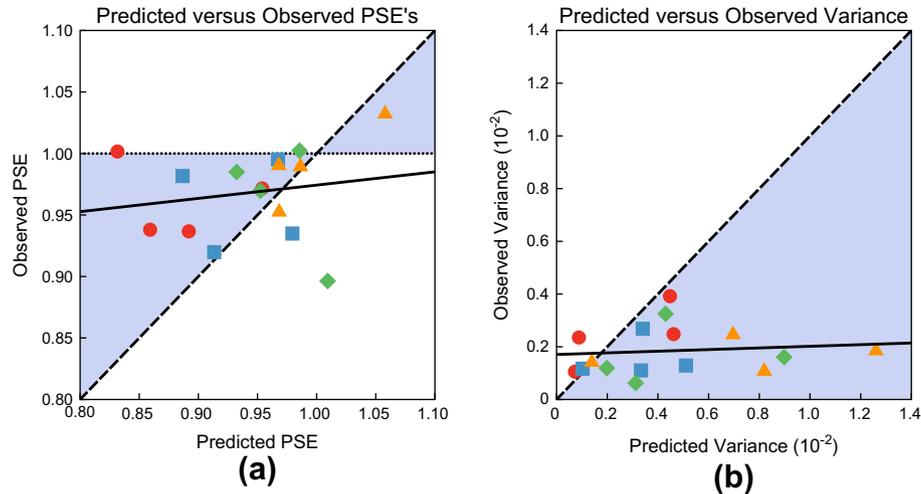


Fig. 12. Predicted versus observed PSE's and variances for the weighted averaging model. Points should lie on the dashed diagonal line if they correspond to that predicted by weighted averaging. Shaded regions on both graphs show where points would lie if observers are more veridical (PSE graph) and more precise (variance graph) than predicted by weighted averaging. For the PSE graph, points should lie on the dotted line if shape is perceived veridically.

applying (Curran & Johnston, 1994; Todd & Norman, 2003). In some ways the reverse correlation methodology we applied might aid in understanding the cues observers use in making perceptual estimates and therefore how to frame, or constrain, the application of cue combination models.

In taking a first pass analysis of the weighted averaging model we fit it to data from the ACC, Monocular Lidded and Binocular Lidded conditions. In the terminology of Eq. (1), these were used for our “single-cue binocular”, “single-cue monocular” and “combined-cue” estimators respectively. The “ACC” condition has been used previously to demonstrate the binocular estimation of shape (e.g. Johnston, 1991), whereas the “Binocular Lidded” condition adds primarily edge and contour information, and the “Monocular Lidded” condition then removes the disparity. We are fully aware that, as with any application of the weighted averaging model, these conditions are unlikely to provide a “pure” measure of each estimator (Zabulis & Backus, 2004). We simply want to see where weighted averaging may fail and why. Predicted versus observed combined-cues shape percepts and variances, along with a linear fit to these data (solid black line), are plotted in Fig. 12. As can be seen, the weighted averaging model provides a poor fit to our data; observers are both more accurate and more precise than weighted averaging predicts.

This is shown by the data predominantly falling away from the unity line and into the shaded regions of both graphs. The predictions of weighted averaging are most clearly violated in terms of the combined-cues variance. This pattern of results could occur if edges and contours act at quantitatively different cues when viewed binocularly versus monocularly (Gillam, Blackburn, & Brooks, 2007; Stevens & Brookes, 1988). An alternative possibility is that we might have overestimated the weight assigned to disparity from the ACC condition. This seems unlikely because, as our analysis has shown, disparity is more informative about depth and shape when the cylinders are rotated. The mechanism of promotion (Landy et al., 1995) also seems unlikely to account for the failure as observers could estimate shape veridically in both the Binocular Lidded and Monocular Lidded conditions. Overall this suggests that defining and measuring “single cue estimates” of a property may be very difficult in practice (Zabulis & Backus, 2004).

6.2.2. The intrinsic constraints model

Domini and colleagues have proposed the intrinsic constraints (IC) model of sensory cue combination, in part to cope with one of the problems we have been discussing: the non-independence

of sensory cues (Domini, Caudek, & Tassinari, 2006; Tassinari & Domini, 2008). Two central tenets of this model are that as more cues are added, the signal to noise ratio (SNR) of the combined-cues estimate increases, and that larger metric depth values are assigned to greater SNR's (Domini, Shah, & Caudek, 2011; Domini, Caudek, & Tassinari, 2006). Our data are consistent with the first prediction, but inconsistent with the latter. With rotation of our cylinders, the precision of observers' perceptual estimates increased, consistent with an increased SNR with more cues. However, the amount of depth our observers perceived in the stimuli *decreased*, opposite to that predicted by the intrinsic constraints model. This arose because observers overestimated the depth of our ACC stimuli, but perceived shape veridically with the additional information provided by binocular edges and contours when the cylinders were rotated. This pattern of results is consistent with previous data on the combination of stereo and motion cues to shape (Scarfe & Hibbard, 2011). Here it was shown that when stereo and motion cues are combined, the depth of the combined cues percept could either increase or decrease, dependent on the bias and variability of the individual cues.

7. Conclusions

In summary, we have shown that human observers are fully capable of accurately and precisely estimating the 3D shape of random dot stereograms of elliptical cylinders, a stimulus that has routinely been used to demonstrate systematic distance-dependent biases in perceived shape over the last few decades. Accurate perception of shape was achieved by the simple adoption of a more naturalistic viewing angle. Along with previous research, this suggests that caution needs to be taken when making inferences from un-naturalistic reduce-cue experiments, to real world performance (Mon-Williams & Bingham, 2008). Using surface-based reverse correlation, we showed that the specific information allowing for accurate and precise estimation of 3D shape was that provided by binocular edges and contours. We assessed observer's performance in terms of two contemporary models of cue-combination and conclude that, at present, these models are unable to fully account for human perceptual performance.

Acknowledgment

We would like to thank an anonymous reviewer for suggesting we conduct Experiment 3.

References

- Ahumada, A. J. J. (1996). Perceptual classification images from Vernier acuity masked by noise. *Perception*, 26, 18.
- Beard, B., & Ahumada, A. J. J. (1998). A technique to extract relevant image features for visual tasks. *SPIE Proceedings*, 3299, 79–85.
- Bingham, G. P., & Lind, M. (2008). Large continuous perspective transformations are necessary and sufficient for accurate perception of metric shape. *Perception & Psychophysics*, 70(3), 524–540.
- Bradshaw, M. F., Elliott, K. M., Watt, S. J., Hibbard, P. B., Davies, I. R. L., & Simpson, P. J. (2004). Binocular cues and the control of prehension. *Spatial Vision*, 17(1–2), 95–110.
- Bradshaw, M. F., Hibbard, P. B., & Gillam, B. (2002). Perceptual latencies to discriminate surface orientation in stereopsis. *Perception & Psychophysics*, 64(1), 32–40.
- Bradshaw, M. F., Parton, A. D., & Glennerster, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Research*, 40(27), 3725–3734.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Brenner, E., & Landy, M. S. (1999). Interaction between the perceived shape of two objects. *Vision Research*, 39(23), 3834–3848.
- Brenner, E., & van Damme, W. J. M. (1998). Judging distance from ocular convergence. *Vision Research*, 38(4), 493–498.
- Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-scene statistics predict how the figure-ground cue of convexity affects human depth perception. *Journal of Neuroscience*, 30(21), 7269–7280.
- Curran, W., & Johnston, A. (1994). Integration of shading and texture cues: testing the linear model. *Vision Research*, 34(14), 1863–1874.
- Domini, F., Caudek, C., & Tassinari, H. (2006). Stereo and motion information are not independently processed by the visual system. *Vision Research*, 46(11), 1707–1723.
- Domini, F., Shah, R., & Caudek, C. (2011). Do we perceive a flattened world on the monitor screen? *Acta Psychologica*, 138(3), 359–366.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433.
- Ernst, M. O., & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4), 162–169.
- Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *Journal of Neuroscience*, 29(49), 15601–15612.
- Fleming, R. W., Holtmann-Rice, D., & Bulthoff, H. H. (2011). Estimation of 3D shape from image orientations. *Proceedings of the National Academy of Sciences of the USA*, 108(51), 20438–20443.
- Gillam, B., Blackburn, S., & Brooks, K. (2007). Hinge versus twist: the effects of 'reference surfaces' and discontinuities on stereoscopic slant perception. *Perception*, 36(4), 596–616.
- Gillam, B., Chambers, D., & Russo, T. (1988). Postfusional latency in stereoscopic slant perception and the primitives of stereopsis. *Journal of Experimental Psychology: Human Perception and Performance*, 14(2), 163–175.
- Gillam, B., Flagg, T., & Finlay, D. (1984). Evidence for disparity change as the primary stimulus for stereoscopic processing. *Percept Psychophys*, 36(6), 559–564.
- Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research*, 36(21), 3441–3456.
- Glennerster, A., Tcheang, L., Gilson, S. J., Fitzgibbon, A. W., & Parker, A. J. (2006). Humans ignore motion and stereo cues in favor of a fictional stable world. *Current Biology*, 16(4), 428–432.
- Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research*, 179(4), 595–606.
- Hershenson, M. H. (1999). *Visual space perception: a primer*. London: MIT Press.
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, 4(12), 967–992.
- Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth: Depth perception*. Toronto: I Porteous.
- Janssen, P., Vogels, R., & Orban, G. A. (1999). Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proceedings of the National Academy of Sciences of the USA*, 96(14), 8217–8222.
- Johnston, E. B. (1991). Systematic Distortions of Shape from Stereopsis. *Vision Research*, 31(7–8), 1351–1360.
- Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research*, 34(17), 2259–2275.
- Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago: Chicago University Press.
- Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, 36, 14–14.
- Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. (1994). On so-called paradoxical monocular stereoscopy. *Perception*, 23(5), 583–594.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination – In defense of weak fusion. *Vision Research*, 35(3), 389–412.
- Mon-Williams, M., & Bingham, G. P. (2008). Ontological issues in distance perception: Cue use under full cue conditions cannot be inferred from use under controlled conditions. *Percept Psychophys*, 70(3), 551–561.
- Orban, G. A., Janssen, P., & Vogels, R. (2006). Extracting 3D structure from disparity. *Trends in Neurosciences*, 29(8), 466–473.
- Oruc, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, 43(23), 2451–2468.
- Pizlo, Z. (2008). *3D Shape: It's unique place in visual perception*. USA: MIT Press.
- Pizlo, Z., Li, Y., & Steinman, R. M. (2008). Binocular disparity only comes into play when everything else fails; a finding with broader implications than one might suppose. *Spatial Vision*, 21(6), 495–508.
- Read, J. C., Phillipson, G. P., & Glennerster, A. (2009). Latitude and longitude vertical disparities. *Journal of Vision*, 9(13), 11–37 (article no. 11).
- Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature*, 361(6409), 253–255.
- Rosas, P., Wagemans, J., Ernst, M. O., & Wichmann, F. A. (2005). Texture and haptic cues in slant discrimination: Reliability-based cue weighting without statistically optimal cue combination. *Journal of the Optical Society of America A-Optics Image Science and Vision*, 22(5), 801–809.
- Rosas, P., Wichmann, F. A., & Wagemans, J. (2007). Texture and object motion in slant discrimination: Failure of reliability-based weighting of cues may be evidence for strong fusion. *Journal of Vision*, 7(6), 3.
- Saunders, J. A., & Backus, B. T. (2006). Perception of surface slant from oriented textures. *Journal of Vision*, 6(9), 882–897.
- Saunders, J. A., & Knill, D. C. (2001). Perception of 3D surface orientation from skew symmetry. *Vision Research*, 41(24), 3163–3183.
- Scarfe, P., & Hibbard, P. B. (2006). Disparity-defined objects moving in depth do not elicit three-dimensional shape constancy. *Vision Research*, 46(10), 1599–1610.
- Scarfe, P., & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory estimates. *Journal of Vision*, 11(7).
- Serwe, S., Drewing, K., & Trommershauser, J. (2009). Combination of noisy directional visual and proprioceptive information. *Journal of Vision*, 9(5), 14–21 (article no. 28).
- Stevens, K. A., & Brookes, A. (1988). Integrating stereopsis with monocular interpretations of planar surfaces. *Vision Research*, 28(3), 371–386.
- Svarverud, E., Gilson, S. J., & Glennerster, A. (2010). Cue combination for 3D location judgements. *Journal of Vision*, 10(1), 1–13 (article no. 5).
- Tassinari, H., & Domini, F. (2008). The intrinsic constraint model for stereo-motion integration. *Perception*, 37(1), 79–95.
- Todd, J. T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences*, 8(3), 115–121.
- Todd, J. T., Christensen, J. C., & Guckes, K. M. (2010). Are discrimination thresholds a valid measure of variance for judgments of slant from texture? *Journal of Vision*, 10(2), 18–21 (article no. 20).
- Todd, J. T., & Norman, J. F. (2003). The visual perception of 3-D shape from multiple cues: Are observers capable of perceiving metric structure? *Perception & Psychophysics*, 65(1), 31–47.
- von der Heydt, R., Zhou, H., & Friedman, H. S. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Research*, 40(15), 1955–1967.
- Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, 5(10), 834–862.
- Wheatstone, C. (1838). On some remarkable, and hitherto unobserved phenomena of binocular vision. *Philosophical Transactions of the Royal Society, London*, 128, 371–394.
- Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8), 1293–1313.
- Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, 63(8), 1314–1329.
- Zabulis, X., & Backus, B. T. (2004). Starry night: A texture devoid of depth cues. *Journal of the Optical Society of America A-Optics Image Science and Vision*, 21(11), 2049–2060.
- Zalevski, A. M., Henning, G. B., & Hill, N. J. (2007). Cue combination and the effect of horizontal disparity and perspective on stereoacuity. *Spatial Vision*, 20(1–2), 107–138.