ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Vision Research 46 (2006) 1599–1610

Vision Research

www.elsevier.com/locate/visres

Disparity-defined objects moving in depth do not elicit three-dimensional shape constancy

P. Scarfe *, P.B. Hibbard

School of Psychology, University of St. Andrews, St. Andrews, Fife KY16 9JP, UK

Received 19 April 2005; received in revised form 1 November 2005

Abstract

Observers generally fail to recover three-dimensional shape accurately from binocular disparity. Typically, depth is overestimated at near distances and underestimated at far distances [Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research, 31*, 1351–1360]. A simple prediction from this is that disparity-defined objects should appear to expand in depth when moving towards the observer, and compress in depth when moving away. However, additional information is provided when an object moves from which 3D Euclidean shape can be recovered, be this through the addition of structure from motion information [Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America A, 2*, 343–349], or the use of non-generic strategies [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 and Psychophysics, 65*, 31–47]. Here, we investigated shape constancy for objects moving in depth. We found that to be perceived as constant in shape, objects needed to contract in depth when moving toward the observer, and expand in depth when moving away, countering the effects of incorrect distance scaling (Johnston, 1991). This is a striking example of the failure of shape constancy, but one that is predicted if observers neither accurately estimate object distance in order to recover Euclidean shape, nor are able to base their responses on a simpler processing strategy.

Keywords: Binocular disparity; Depth constancy; Stereopsis; Cue interaction; Motion parallax

1. Introduction

Our two eyes receive slightly different views of the world as a consequence of their lateral separation in the head. The binocular disparities between the two images include horizontal and vertical differences in the location of corresponding points across the two images, and monocular occlusion (regions seen by only one eye). Binocular horizontal disparity is a particularly useful source of depth information as, once suitably scaled, it can be used to specify the full three-dimensional structure and location of objects. Numerous cues are available that may be used to scale horizontal disparity, such as vergence, vertical disparity, and other cues to distance (Hershenson, 1999; Rogers & Bradshaw, 1993).

However, despite the possibility that scaled horizontal disparity may provide accurate three-dimensional shape information, the perception of shape from binocular cues is often far from perfect. Typically, misperceptions of the shape and size of objects are consistent with an overestimation of close distances, and an underestimation of far distances. As a result, objects tend not to exhibit shape constancy. That is, the perceived three-dimensional shape of an object when presented at two different distances may differ, even when the physical shape of the object has not changed (Bradshaw, Parton, & Glennerster, 2000; Glennerster, Rogers, & Bradshaw, 1996; Johnston, 1991). Studies in this area have generally sought to isolate binocular cues through the use of computer-generated stimuli. These are typically presented on a monitor at relatively

* Corresponding author. Tel.: +44 (0)1334461995, fax: +44 (0)133 4463042.

E-mail address: ps18@st-andrews.ac.uk (P. Scarfe).

0042-6989/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2005.11.002

short viewing distances (<1.5 m), where binocular cues are most reliable. Distortions of shape, size and distance are, however, also observed using real-world stimuli (Bradshaw et al., 2000; Cuijpers, Kappers, & Koenderink, 2000; Loomis, Philbeck, & Zahorik, 2002), including during viewing in naturalistic open field environments (Koenderink, van Doom, & Lappin, 2000; Koenderink, van Doom, Kappers, & Todd, 2002; Loomis & Philbeck, 1999; Wagner, 1985).

The lack of three-dimensional shape constancy was demonstrated in a highly influential study by Johnston (1991). In this study, observers were asked to judge the apparent three-dimensional shape of disparity-defined hemi-cylinders. In performing this task, observers were subject to systematic inaccuracies, which depended on the distance at which the hemi-cylinders were presented. For near viewing distances (less than around 1 m) observers overestimated the depth of an object relative to its height, whereas for far viewing distances depth was underestimated relative to height. This lack of shape constancy raises intriguing questions regarding shape constancy for objects that are moving in depth. One simple prediction is that an object of constant physical shape will appear to stretch in depth as it moves towards an observer, and to compress in depth as it moves away from an observer. Conversely, in order to maintain perceptual constancy, an object would have to squash in depth as it moved towards the observer, and stretch in depth as it retreated.

While these predictions follow directly from the lack of shape constancy demonstrated by Johnston (1991), there are at least two reasons for suspecting that such a result would not be observed, and that observers would maintain accurate shape constancy for objects moving in depth. First, the retinal motion generated by the motion of the object in depth provides additional information as to its three-dimensional shape. Depth information provided by binocular disparity and image motion in conjunction provides a unique interpretation of the three-dimensional shape of an object (Richards, 1985). Although most studies on the combination of disparity and motion have focused on rotational object motion (e.g., Brenner & Landy, 1999; Johnston, Cumming, & Landy, 1994) or observer movement (e.g., Bradshaw et al., 2000), motion in depth could be used in a similar way to accurately determine shape.

Evidence that the human visual system combines disparity and motion information is equivocal. Numerous studies have shown that these cues are not always combined to allow the accurate recovery of three-dimensional shape even though this is theoretically possible (Bradshaw et al., 2000; Norman, Todd, Perotti, & Tittle, 1996; Tittle, Todd, Perotti, & Norman, 1995; Todd & Norman, 2003; Todd, Tittle, & Norman, 1995). In some circumstances the addition of motion is seen to improve shape from stereo, but in most cases this is a partial improvement and does not result in veridical perception (Bradshaw, Parton, & Eagle, 1998; Brenner & Landy, 1999; Durgin, Proffitt, Olson, & Reinke, 1995; Johnston et al., 1994; Rogers & Collett, 1989; Tittle & Braunstein, 1993). Stereo-motion combination thus seems to occur under restricted and at present undefined circumstances (Brenner & Landy, 1999; Landy & Brenner, 2001). Given that studies showing optimal combination of disparity and motion are in the minority, it is possible that these instances represent the use of non-generic, context-specific strategies for recovering three-dimensional shape when motion and stereo are available, rather than true depth cue combination (Todd & Norman, 2003).

A second possibility is that numerous non-generic processing strategies may be employed that would allow for accurate responses to particular tasks, without the need to recover full Euclidean shape (Gårding, Porrill, Mayhew, & Frisby, 1995; Glennerster et al., 1996; Todd, 2004). For example, when matching the shape of two disparity-defined objects at different distances, all that is needed to match shape is knowledge of the ratio of the distances to the objects (Bradshaw et al., 2000; Glennerster et al., 1996), rather than the absolute distance information that is required for full scaling. With this information the ratio of the disparities can be adjusted, and shape can be accurately matched, while at no point in time does the observer need to recover metric shape information. Similarly it has been shown that to judge that a disparity-defined surface is fronto-parallel with respect to the observer, the visual system need only gauge that the horizontal size ratio of the surface is equal to the square of its vertical size ratio (Rogers & Bradshaw, 1995). Scaling with an absolute estimate of the viewing distance is again not necessary for accurate completion of this task.

Strategies such as these may be used more often than is widely appreciated (Todd, 2004) as in many situations it may be computationally more efficient for the brain to use less demanding processing strategies to effectively control behaviour. For example, tasks such as prehension, that have traditionally been viewed as requiring a metric representation of shape for their successful completion (Milner & Goodale, 1995; Servos, Goodale, & Jakobson, 1992), may rely on less demanding strategies such as the online nulling of the relative disparity between the object to be grasped and the hand, rather than an accurate specification of the metric structure of the object or scene (Bradshaw & Elliott, 2003; Hibbard & Bradshaw, 2003; Morgan, 1989). It is possible that the visual system may rarely need to recover full metric shape information in our everyday interactions in the world (Bradshaw et al., 2000).

A similar strategy can be identified for determining the rigidity or otherwise of stereoscopically viewed objects that are moving relative to the observer. For a given depth separation, binocular disparity is proportional to the inverse of the square of the viewing distance. Similarly, retinal image size is proportional to the inverse of the viewing distance. For an object moving directly towards an observer along the cyclopean line of sight, it can be shown that the rate of change of disparity, divided by the instantaneous disparity, will be twice the value of the rate of change of retinal image size, divided by the instantaneous retinal image size (see Appendix A). Calculation of this ratio would therefore provide a quick check as to whether an object maintains a constant shape as it moves through space.

The current study investigated the extent to which binocularly defined objects moving in depth exhibit shape constancy. Two possible results can be predicted. First, a perceptually constant cylinder might be one whose shape modulates as it moves through depth, countering the effects of incomplete distance scaling as shown by Johnston (1991) and others. Alternatively, a perceptually constant cylinder might be one that is physically constant, if observers are able either to combine disparity and motion information to accurately recover depth, or to use some other processing strategy, in order to ascertain the rigidity of the moving object.

The degree of shape constancy for simple, disparity-defined objects was determined in three tasks. In the first, a standard apparently circular cylinder (ACC) task, observers were required to adjust the shape of a hemi-cylinder so that it appeared to have a circular depth profile. Appropriate scaling of horizontal disparity to take account of object distance is required in order to accurately complete this task. In the second, a matching task, observers were asked to adjust the shape of an object presented at one distance, so as to match the shape of another object at a different distance. This is simpler than the ACC task, in that it does not require the scaling of disparity to take account of viewing distance (Glennerster et al., 1996). Finally, observers were asked to judge whether a cylinder moving in depth was increasing or decreasing in depth extent. Again, accurate performance in this task can be obtained without the need for disparity scaling. Taken together, the results of these three experiments will allow us to determine whether observers are able to achieve shape constancy by circumventing the need for disparity scaling, either by making simple comparisons across two objects, or by taking account of the changing information available in a single, moving object.

2. Method

2.1. Stimuli

The stimuli were random dot stereograms of elliptical hemi-cylinders; the surfaces of the cylinders were rendered with 100 red Gaussian blobs, 0.5 mm in diameter. Red was used as the shutter goggles have the minimum cross-talk at longer wavelengths. The blobs scaled appropriately with distance such that mean screen luminance increased with decreasing distance. Blob size and screen luminance could therefore be used as a cue to object distance, but not shape. The stimuli were programmed using an assumed interocular distance of 6.5 cm (Howard & Rogers, 2002), meaning that only subjects with this interocular distance would achieve veridical depth. In all three tasks we are interested in the pattern of responses across distance, more than the veridicality of shape judgements. The effect of a deviation from the programmed interocular distance would be to affect the veridicality of judgments, rather than the pattern of set shape over distance.

Each hemi-cylinder had a fixed height of 3 cm, and was presented at eye height with the principle axis aligned horizontally in the fronto-parallel plane. The ends of the cylinders were jittered with respect to the fixed length of 8 cm; this was achieved by displacing the horizontal position of the end dots of the cylinder by an amount determined by the sum of two sinusoids that had frequencies of 4 and 6 cycles over the circumference of the cylinder. The amplitude of the two sinusoids was 8 mm, and their phases varied randomly across trials. This, in conjunction with the relatively low dot-density used, ensured that the optical deformation of visible right-angle end cuts did not provide useful information about 3D shape (Todd & Norman, 2003).

2.2. Apparatus

Stimuli were viewed on a 21 in. monitor through Stereographics LCD shutter goggles synchronized to the 100 Hz refresh rate of the screen. The left and right eyes each received a new image every 20 ms. The monitor had a resolution of 800×600 pixels, and was viewed in a dark room at a fixed distance of 32.5 cm. Head movements were minimized with the use of a chin and headrest. Stimuli were presented at simulated distances of 40, 50, 60, and 70 cm. This resulted in unavoidable cue conflicts with accommodation, which would indicate the true distance of the screen, whereas binocular disparity, motion and vergence were consistent with the simulated distance and shape of the cylinder. Cue conflicts were, however, reduced by using a restricted range of simulated distances, and evidence has suggested that accommodation in isolation is a weak distance cue (MonWilliams & Tresilian, 1999). However, given the problems inherent in using simulated versus real objects (Buckley & Frisby, 1993; Durgin et al., 1995; Frisby, Buckley, & Horsman, 1995; Frisby, Buckley, & Duke, 1996; Girshick, Akeley, Watt, & Banks, 2004; Watt, Akeley, Girshick, & Banks, 2005) these conflicts must be kept in mind.

2.3. Participants

There were seven participants in total, two experienced psychophysical observers (the authors) and five observers who were naive to the purpose of the experiment, and had limited familiarity with psychophysical research. All had normal or corrected to normal vision, and good stereopsis.

2.4. General procedure

Observers were allocated to complete either the ACC or matching task first at random. All, however, completed the motion task last, since data from the ACC task of each observer were needed to generate some of the stimuli in the motion task. To standardize stimulus presentation across the three tasks cylinders were presented along two diagonal paths, at a distance of 40, 50, 60 or 70 cm. These distances were chosen in relation to the experimental literature (and considerations of cue conflict) to cover a sufficient range to expect an effect of distance on perceived shape (van Damme & Brenner, 1997). At the 40 and 70 cm distances the cylinders were laterally separated from the mid-point of the screen by 8 cm (either to the left or to the right), resulting in two diagonal depth profiles with an angle of 28° relative to the line of sight (Fig. 1).

2.5. The apparently circular cylinder task

The experimental task was to set the depth of the hemi-cylinder to be equal to half its height, i.e., to set the cylinder so that it appeared circular (Johnston, 1991). Shape adjustments were made on a computer keyboard, with one key increasing and another decreasing the cylinder's depth. The cylinder's depth could be adjusted between 0 cm (a flat fronto-parallel surface) and 9 cm (six times the depth needed for a circular cylinder), this range had proved sufficient in a pilot study, and was reported to be sufficient here. A key press triggered the presentation of the first stimulus which had an initial depth taken at random from the full range. There was no time limit to adjustment. Once satisfied with their settings observers

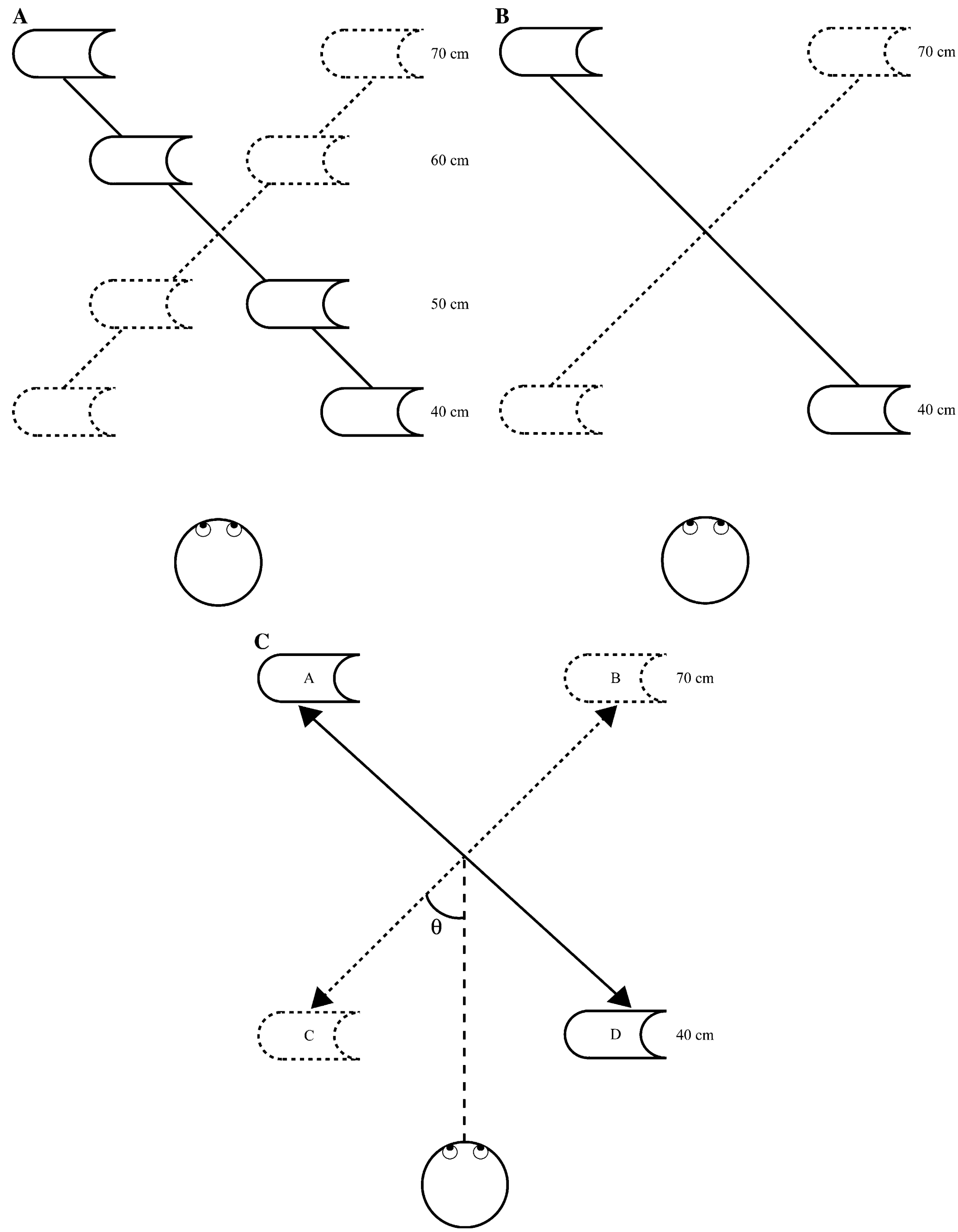

Fig. 1. Schematic diagram of the stimuli for the three tasks seen from above, (A) the ACC task, (B) the matching task and (C) the motion task. Stimulus presentation was standardized by presenting the stimuli along two diagonal depth profiles in each task. These depth profiles (notionally labelled 'left' solid line/arrow, and 'right' dashed line/arrow) cut an angle of $\theta = 28°$ with respect to the observers' line of sight. For further details see the text.

pressed a key once, which removed the cylinder from the screen, and again to trigger the presentation of the next stimulus. Fifteen settings were made at each distance, each presented at random, resulting in a total of 60 settings. The diagonal depth profile (left or right, see Fig. 1A) on which the cylinder appeared on a given trial was also determined at random.

2.6. The matching task

The matching task required observers to match the shape of two cylinders (the reference and the match) that appeared at the maximum and minimum distances used in the ACC task (70 and 40 cm). This task is similar to the ACC task but does not require the recovery of metric shape information (Glennerster et al., 1996). Rather, accurate settings may be achieved if the ratio of the distances to the two objects can be determined. An initial key press triggered the simultaneous presentation of two cylinders, each at eye level, laterally separated from the midpoint of the screen by ±8 cm. The cylinder to be adjusted (the match) was indicated by a monocularly presented dot positioned 12.5° above it; the dot was monocular so as not to provide any additional depth information. The initial depth of the match cylinder was taken at random from the full adjustment range, which was the same as in the ACC task. The reference cylinder always specified a physically circular cylinder. Adjustments to the match cylinder were made using the computer keyboard. When satisfied with their setting, observers pressed a key once to remove the stimuli, and again to trigger the presentation of the next trial. The match was presented 15 times at each distance in a random order, with the match and reference alternating position (left or right, see Fig. 1B) randomly over trials.

2.7. The motion task

In the motion task observers saw a cylinder translating diagonally toward or away from them along the diagonal depth profiles (Fig. 1C). Two directions of motion (toward and away) were used for a number of reasons. If observers are unable to combine motion and stereo information, or use an alternative processing strategy to accurately access the 3D shape constancy of objects moving in depth, three direct predictions can be made from the results of Johnston (1991). First, objects moving toward the observer should need to contract in depth extent to be perceived as perceptually constant in shape. Second, objects moving away from the observer should need to expand in depth extent to be perceived as perceptually constant in shape. Third, if observers accurately track the moving objects and predominantly base their decisions on scaled object disparity at each point along the motion path the contraction needed with movement toward, and the expansion needed with movement away, should be isotropic.

The direction of motion, toward (start point 70 cm) or away (start point 40 cm), varied randomly as did the diagonal depth profile along which the cylinders moved (Fig. 1C). There were therefore four possible combinations of start point and direction of motion, (1) the cylinder started at 'A' and moved toward the observer to 'D', (2) the cylinder started at 'D' and moved away from the observer to 'A', (3) the cylinder started at 'B' and moved toward the observer to 'C', or (4) the cylinder started at 'C' and moved away from the observer to 'B' (Fig. 1C). During their movement, cylinders remained either physically circular in shape, expanded or contracted in depth extent along the line of sight, or modulated their depth extent in relation to individual observers' settings in the ACC task. These latter stimuli were generated by performing a linear regression of set shape onto distance for each observer's results in the ACC task. At each point along the motion trajectory these cylinders assumed the shape of an apparently circular cylinder for that observer at that distance. If perceived shape varies linearly with distance, and objects moving in depth exhibit the same degree of shape constancy as static objects, then it is predicted that these stimuli would appear constant in shape, regardless of their direction of motion. There were five levels of expansion or contraction; these were 0.25, 0.5, 1, 2 and 4. These values were chosen to incorporate the range of depth settings seen in a typical ACC task. When the cylinder's depth extent changed during translation, it did so smoothly over the course of the whole movement, with the mid-motion depth always being equal to a circular cylinder. The exception was the ACC modulating cylinders, as although their depth extent changed smoothly over the course of the movement, the range of shapes presented did not necessarily include one that was physically circular. This is because these cylinders' depths directly depended on each observer's mean ACC settings at each of the four distances.

There were therefore six conditions in the motion task; four levels of expansion or contraction, the ACC modulated cylinders, and the physically constant cylinders. Each was presented moving toward and away 20 times (i.e., 10 times along each depth profile), this resulted in 240 trials overall. The trials were presented in a random order, and split into two blocks of 120 to avoid observer fatigue. Observers triggered the first trial with a key press, a cylinder then appeared randomly at either the far or near distance, either to the left or to the right (Fig. 1C). This remained stationary for 1 s before translating in depth at a speed of 17.8 cm s^{-1}, either toward or away from the observer. At the end of the movement trajectory the cylinder disappeared. The observer's task was to decide if the cylinder they saw moving toward or away from them was expanding or contracting in depth extent. They registered their decision with a key press, which also triggered presentation of the next stimulus. Prior to starting the task the terms 'expanding' and 'contracting' in depth extent were explained clearly to observers using a cardboard model of the stimuli. This ensured that observers did not confuse the cylinders' changing depth extent with any perceived change in size during the movement.

3. Results

3.1. The ACC task

The mean set shape (depth to width ratio) was calculated as a mean of the set shape at each of the two positions (one on each diagonal depth profile) for a given distance (Fig. 1). The data indicate that across observers there was only a weak effect of distance on the set shape of an ACC, with marginally more depth being set in the ACC at far compared to near distances (Fig. 2). A Friedman test on the mean set shape of an ACC at each distance across observers showed there to be a significant effect of distance on observer's settings, $\chi^2 = 10.71$, df = 3, $p < 0.01$ (Monte Carlo sig, 99% confidence, 100,000 iterations), the magnitude of this effect is, however, moderate. Given the known individual differences found between observers in shape judgment tasks such as these (Champion, Simmons, & Mamassian, 2004; Glennerster et al., 1996; Oruç, Maloney, & Landy, 2003; Todd & Norman, 2003) individual observers' data were investigated. The individual ACC data show there to be large differences in the set shape of an ACC between some observers (Fig. 3). In particular, only observers D and PS appear to show a clear effect of distance on ACC settings, which is likely therefore to be driving the overall significant result. This is supported by the fact that when these observers are removed the group effect of

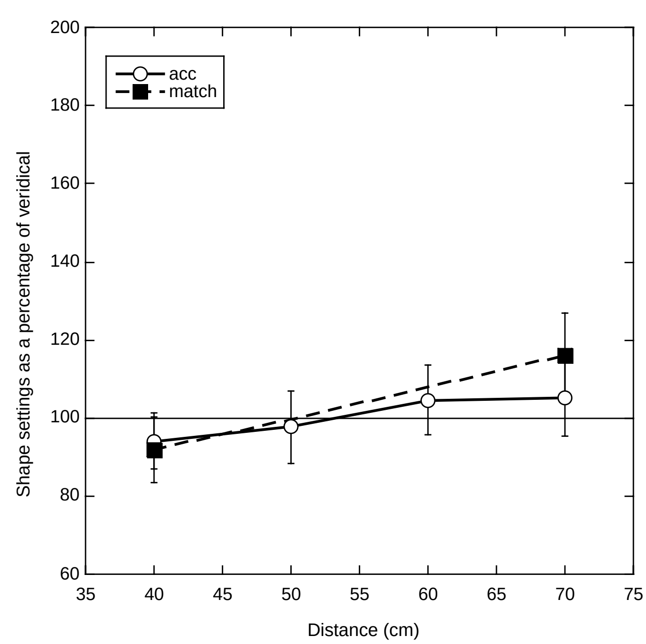

Fig. 2. The overall ACC and matching results (error bars show 95% confidence intervals). The horizontal line indicates the depth needed for a physically circular cylinder. The scale is the same as Figs. 3 and 4 for comparison.

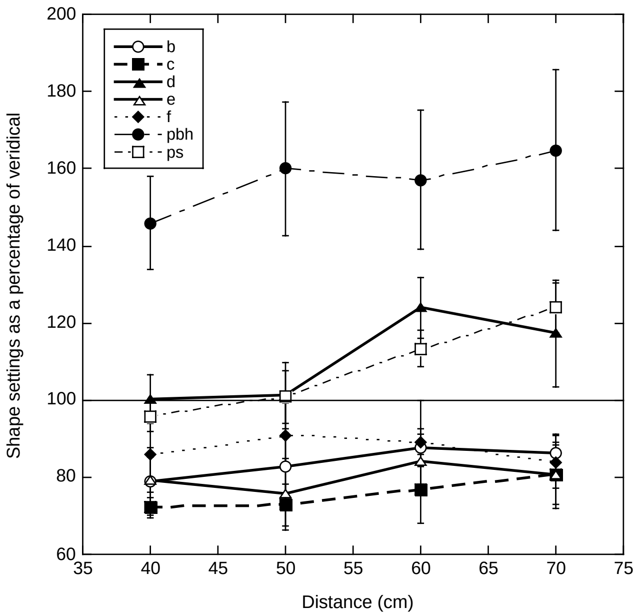

Fig. 3. Individual observer's ACC data (error bars show 95% confidence intervals). The horizontal line indicates the depth needed for a physically circular cylinder.

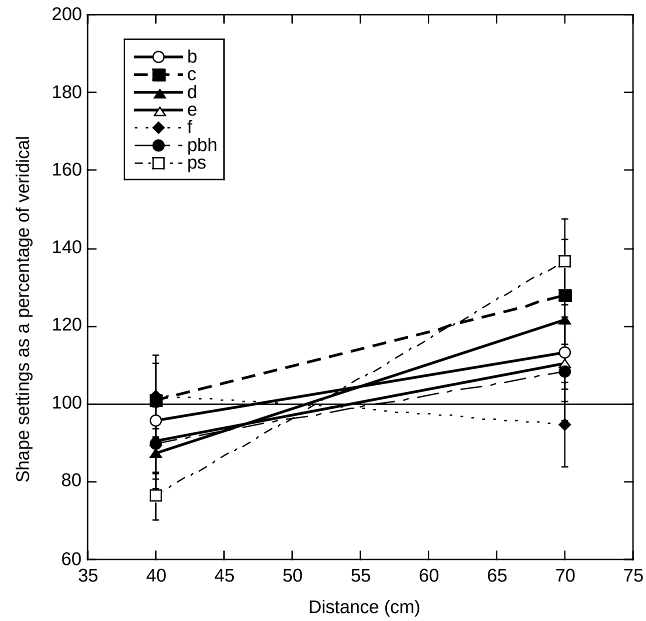

Fig. 4. Individual observer's matching data (error bars show 95% confidence intervals). The horizontal line on the graph represents the depth needed for a physically circular cylinder. Symbols are the same as the ACC data in Fig. 3 for comparison.

distance on ACC settings becomes non-significant $\chi^2 = 5.88$, df = 3, $p = 0.123$ (Monte Carlo sig, 99% confidence, 100,000 iterations). This does, however, need to be tempered by the loss of statistical power due to removing two observers.

3.2. The matching task

The matching task data showed a clearer effect of distance than the ACC task (Fig. 2). Overall, when matching the shape of a cylinder across distance observers set significantly more depth in far compared to near cylinders (Wilcoxon $z = 2.20$, $p = 0.03$). This overall result was reflected by individual observers with the exception of observers F and PBH (Fig. 4). The clearer distance effect in the matching task was not, however, accompanied by reduced variability in observer's settings, as there was no significant difference between the standard deviation of shape settings in the matching and ACC tasks at 40 cm, Wilcoxon $z = 1.18$, $p = 0.30$, or 70 cm, Wilcoxon $z = 0.17$, $p = 0.94$ (these being the two shared distances between tasks). This contrasts with the subjective impressions of the majority of observers, who found this task much easier to complete. The matching task data are thus consistent with observers scaling the disparities they set in the match cylinder with an incorrect estimate of viewing distance in a similar manner to the standard ACC task (Johnston, 1991). It also suggests that in this instance observers did not use the alternative processing strategy to set the relative shapes of the two objects, which does not require distance scaling (Bradshaw et al., 2000; Glennerster et al., 1996).

3.3. The motion task

Psychometric functions were fitted to the results using Probit analysis, and 95% confidence limits were calculated using a Bootstrap technique (Foster & Bischoff, 1991). The 50% point of this fitted function represents the amount by which a cylinder needed to expand or contract when moving toward or away from the observer in order that the observer was equally likely to judge it to be expanding or contracting, i.e., to be perceived as physically constant in shape with movement in depth. Representative psychometric functions from observer E are shown in Fig. 5, a clear separation of the toward and away curves can be seen. In order to have the appearance of a physically constant three-dimensional shape, cylinders moving toward the observer needed to contract in depth extent (PSE < 1), whereas cylinders moving away from the observer needed to expand in depth extent (PSE > 1). The mean PSE with movement toward the observer was 0.89, and away from the observer 1.35 (Fig. 6), this difference was significant (Wilcoxon $z = 2.37$, $p = 0.02$).

Fig. 6 shows that the significant effect of direction of motion on observers' responses was consistent across all observers bar two. Observer F showed no clear effect with both PSEs approximately equal to one, whereas the PSEs of observer PBH were both approximately the same but greater than one indicating that for this observer cylinders needed to expand in depth extent whichever direction they were moving. Looking at the physically constant cylinders in isolation supports the overall PSE result, 66% of physically constant cylinders moving toward the observer were seen as expanding in depth extent, compared to only 35%

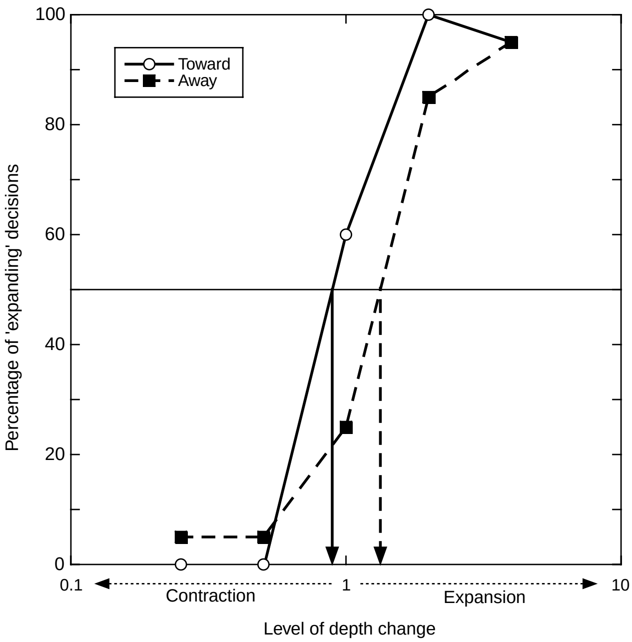

Fig. 5. Representative psychometric functions from observer E. The point of subjective equality (PSE) in the motion task measures the level of expansion or contraction needed for half the observers decisions to be 'expanding' and half 'contracting', i.e., the level of expansion or contraction in depth needed for a 'perceptually constant cylinder'. This is represented on the graph where the 'toward' and 'away' functions intersect the horizontal 50% line, and is indicated by the solid 'toward' and dashed 'away' arrows, respectively. For this observer cylinders needed to expand when moving away (PSE > 1) and contract when moving toward (PSE < 1) to be perceived as physically constant in shape.

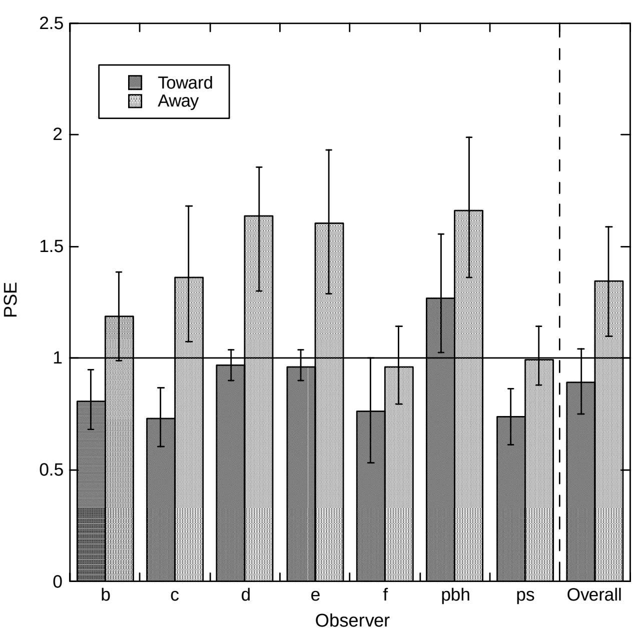

Fig. 6. Individual observer's PSEs with movement toward and away in depth (error bars show 95% confidence intervals determined from the psychometric function fit). A PSE < 1 indicates a cylinder needed to contract in depth extent to be perceived as constant in shape, whereas a PSE > 1 indicates that a cylinder needed to expand in depth extent to be perceived as constant in shape.

of cylinders moving away from the observer. This difference was statistically significant, Wilcoxon $z = 2.38$, $p = 0.02$.

Given that only two out of seven observers showed a significant effect of distance in the ACC task, and that this distance effect was weak, it was not meaningful to further explore the extent to which the ACC modulating cylinders were seen as remaining perceptually constant in depth extent. Interestingly, although there were large individual differences in the level of expansion and contraction needed with movement away from and towards the observer (Fig. 6), there was no corresponding significant difference between PSEs,[1] Wilcoxon $z = 0.68$, $p = 0.58$. This suggests that overall the level of expansion needed with movement away and contraction needed with movement toward was isotropic. It also suggests that asymmetries that have been observed between convergence and divergence responses (Hung, Zhu, & Ciuffreda, 1997) did not affect observers' perceptions of shape.

4. Discussion

4.1. Comparing the tasks

The ACC task produced only a weak effect of distance in a limited number of observers. This is surprising as, although the distance range was restricted so as to minimize cue conflicts, a similar distance range has produced clear effects in previous studies (e.g., van Damme & Brenner, 1997), and was sufficient to produce clear distance effects in both the matching and motion tasks presented here. It may be that the use of a handheld reference (van Damme & Brenner, 1997) acts to reduce the variability of shape settings in ACC type tasks resulting in a clearer effect of distance. Alternatively, the effect of distance on settings may have been less clear here as cylinders were presented along two diagonal depth paths. Along these paths disparity and vergence thresholds may vary with eccentricity, which will be expected to systematically affect the perception of shape from stereopsis (Scarfe & Hibbard, 2004). In contrast, the matching task showed a robust distance effect, with observers tending to underestimate the far and overestimate the near distances, consistent with incorrect distance scaling. Observers also failed to use the alternative processing strategy by which to match the shape of the two cylinders (Bradshaw et al., 2000; Glennerster et al., 1996).

The failure of observers to use this strategy may be due to a number of factors. To set the ratio of disparities accurately, so as to match the shapes of the cylinders, some estimate of the ratio of the viewing distances must be obtained. This might come, for example, from the ratio of the heights of the stimuli, if it is assumed that they are

[1] PSEs for movement towards were compared with 1/PSEs for movement away so as to compare the degree of expansion/compression appropriately.

the same size (Glennerster et al., 1996). In the Glennerster et al. (1996) study, the match and the reference were viewed in dim lighting conditions on separate monitors placed on a black and white textured linoleum surface. From this information, the ratio of distances could more readily be estimated from the angle subtended by the monitors, or other distance cues such as the texture gradient.

The situation in the matching task presented here was different as the stimuli were viewed in darkness. This may have made it much less straightforward for the visual system to assume size constancy, and estimate the ratio of stimulus distances in order to match the cylinder's shape. The matching strategy, as with the ACC task itself, will be affected by the available distance information (Glennerster, Rogers, & Bradshaw, 1998), and in this instance the distance information available may not have been sufficient. Contrary to this, Bradshaw et al. (2000) have demonstrated that observers can accurately match shape when viewing very sparse stimuli (triangles defined by three isolated LEDs) in darkness. The absence of additional cues to size and distance cannot therefore fully explain why observers did not accurately match shape here. It does, however, seem that stimuli are not always matched in instances where the strategy could be used (Todd & Norman, 2003), which together with the present data points to the fact that the strategies used may be more context dependent than was originally thought. One possibility is that the pattern of eye movements required to fixate the two objects may affect matching performance. Although evidence suggests that observers can make relatively good distance judgements across changes in vergence (Brenner & van Damme, 1998; although see, Brenner & Landy, 1999), this may not be the case for changes in version. The Bradshaw et al. (2000) study required observers to make relatively small (<1°) vertical version movements to fixate the stimuli whereas the present study and others (Todd & Norman, 2003) have required much larger horizontal version movements (≈18° in the present study). This difference may have affected whether the matching strategy was used.

Interestingly, even though observers failed to use the matching strategy this did not result in all observers making similar shape settings in the ACC and matching tasks.

The results were mixed; some observers did make similar settings in the two tasks (e.g., Observers D and F), whereas others did not (e.g., Observers C and PBH). If observers perceived the reference correctly during the matching task, accurate performance in both the ACC and the matching task would result in the same settings being made, however, this is unlikely to be the case. Observers were clearly unable to use the matching strategy here, so both the match and the reference will have been subject to incorrect distance scaling. This would have acted to magnify the effects of distance miscaling in the matching task compared to ACC task, and could explain why a robust distance affect was found in the matching task, but only a weak effect in the ACC task.

Consider the case where observers are presented with a physically circular 'reference' cylinder at a far distance, and have to match its shape in an adjustable 'match' cylinder at a near distance. First, the reference will be perceived as squashed in depth relative to its physically circular shape. The observer's task is then to match this *perceived* shape in the near cylinder. However, any settings made in this near cylinder will also be misperceived, such that depth is overestimated. The set *physical* shape of the match will therefore be further squashed relative to the *perceived* shape of the (*physically circular*) reference. The converse applies when the 'reference' is near and the 'match' is far. Therefore, when the matching strategy cannot be used, the matching task will produce greater measured miscaling compared to the ACC task because in doing the matching task observers are miscaling twice.

Different observers may also have processed the two tasks differently, or adopted different heuristic strategies in their completion. As Todd (2004) has suggested, the strategies that observers adopt to mediate their responses in the face of ambiguous information do not necessarily remain constant over different response tasks. Individual differences are a consistent feature of many shape judgment tasks, and to this extent the present data are no exception. These differences are an important and consistent feature of this area of research, and thus need to be explained in order to fully understand the processing that underlies observers performance (Glennerster et al., 1996; Hibbard, Bradshaw, Langley, & Rogers, 2002; Scarfe & Hibbard, 2004; Todd, 2004; Todd & Norman, 2003).

4.2. Stereo-motion combination

Despite the availability of sufficient information from the combination of disparity and motion cues (Richards, 1985), observers failed to accurately identify objects that exhibited physical shape constancy as they moved in depth. To be perceived as remaining constant in shape, cylinders needed to expand in depth when moving away, and contract in depth when moving towards the observer. A perceptually constant cylinder was thus one that modulated its shape in a way that countered the effects of scaling disparity with incorrect estimates of viewing distance (Johnston, 1991). Overall the expansion and contraction appeared to be isotropic suggesting that observers accurately tracked the moving objects, and based their decisions on (mis)scaled object disparity at each point in the course of the movement. These results are further evidence that motion and stereo are rarely combined to veridically estimate shape (see Landy & Brenner, 2001; Todd & Norman, 2003 for reviews).

Observers were also unable to compare the changing disparity of a moving cylinder with its changing image size to determine if it was remaining rigid in shape. There have been reports in the literature that the visual system may be poor at utilizing information regarding the rate of change of disparity (Harris & Watamaniuk, 1995). However, in the current study the dot lifetimes were not limited, as in the Harris and Watamaniuk (1995) study, the stimuli therefore contained consistent interocular velocity differences that could have been detected by mechanisms tuned directly to this property (Brooks & Stone, 2004), and used in the same way to assess whether an object was expanding or contracting in depth.

Retinal size, as with disparity, needs scaling by an estimate of viewing distance. If observers were scaling size incorrectly, as with disparity, the increasing angle subtended by a cylinder (of constant size) as it moved toward the observer could be interpreted as a change in object size. Thus physically constant cylinders moving toward the observer could be perceived as expanding in size, and cylinders moving away contracting in size. If observers confused this with the depth of the object, it could provide an alternative explanation for the motion results. This is unlikely for a number of reasons. Observers were all well briefed on the task before commencing; this included a demonstration of the depth and width of a cardboard model of the cylinder stimuli, as well as an explanation of the task. Also all cylinders across all tasks were the same size and scaled with distance appropriately, but a clear separation in the psychometric functions was found. Therefore, although size is subject to distance mis-scaling too, this would have been consistent across all the stimuli, and could not account for the differential pattern of results found with different levels of cylinder expansion or contraction.

Overall the motion and matching tasks show a greater level of consistency with one another than with the ACC task. It can be seen by comparing Figs. 4 and 6 that those individuals showing an effect of distance in the motion task also show an effect in the matching task. This is surprising as the ACC and motion tasks should be more consistent given that observers are likely to have been miscaling twice in the matching task. The majority of observers commented on how much "easier" the matching and motion tasks were compared to the ACC task, which is consistent with previous research suggesting that the ACC is an intuitively difficult task for observers to perform (Todd & Norman, 2003). Judging the shape of cylinders may be difficult because fusion cannot occur where the disparity gradient exceeds a critical value of ≈ 1 (Burt & Julesz, 1980), toward the horizontal edges of the stimuli. It may be that observers find this task difficult but have increased confidence in making relative shape judgements in the matching and motion tasks, this could affect either the way observers combine the available cues, or the response strategies and heuristics they adopt (Todd, 2004), and may account for the comparative subjective difficulty of the ACC task.

4.3. Comparing real and virtual stimuli

Although distortions of distance, shape and size have also been observed using real-world stimuli and in the natural environment (Bradshaw et al., 2000; Cuijpers et al., 2000; Koenderink et al., 2000, 2002; Loomis & Philbeck, 1999, 2002; Wagner, 1985), in most cases it is not possible to make direct comparisons. In the natural environment the observer is generally much less constrained and there are many more visual cues available to combine. The reliability and utilization of these visual cues will differ greatly between computer-generated stimuli and viewing in the natural environment, particularly as studies in the natural environment generally use a much greater distance range. Some real-world studies have used a methodology that is more comparable to those with computer-generated stimuli, or have directly compared the two, but the results are mixed.

Bradshaw et al. (2000) demonstrated similar perceptual distortions using sparse real-world stimuli, but others have found stark differences when comparing real and virtual objects, such that shape is perceived more veridically when viewing real objects (Buckley & Frisby, 1993; Durgin et al., 1995; Frisby et al., 1995, 1996). At present it is unclear whether these differences represent deficiencies inherent in using computer-generated stimuli, such as the role of accommodative cues to the screen (Girshick et al., 2004; Watt et al., 2005), or are the consequence of a robust visual system exploiting non-generic cues that are present in the full cue environment (Todd & Norman, 2003). Without identical visual information it will remain difficult to compare studies using real and simulated stimuli, without making assumptions about the role of conflicting or uncontrolled cues.

4.4. Summary

In summary, the results provide a striking demonstration of the perceptual consequences of the failure to obtain shape constancy. When sufficient information for the recovery of Euclidean shape is unavailable, physically constant disparity-defined objects are perceived to expand in depth when moving towards the observer and to contract in depth when moving away. This is consistent with recent research demonstrating that a physically constant world is not necessarily a perceptually stable one (Tcheang, Gilson, & Glennerster, 2005). It is therefore justifiable to ask why, given well-documented biases in perception (Todd & Norman, 2003), even in the natural environment (Wagner, 1985), we perceive the world to be a relatively stable place. Much will depend on the type and quality of the visual cues available (Knill & Richards, 1996), but the visual system may also embody a strong assumption toward world stability, such that stereo and motion information is ignored when it is not consistent with this, or is less reliable (Glennerster, Gilson, Tcheang, & Parker, 2003). The fact that observers ignore stereo and motion information that indicate a lack of stability when this is in conflict with other cues is consistent with the current results, which demonstrate a failure to use stereo and motion to determine the rigidity or otherwise of objects moving in depth. It appears that the stability of the world is assumed rather than determined from geometrical information, but that this assumption has little consequence for our everyday interactions.

Appendix A

Consider a point P at eye height at a depth Z and horizontal position X (Fig. 7).

The visual directions of this point in the left eye and right eye are given by

$$\tan\theta_{\mathrm{PL}} = \frac{D\sin\alpha + I/2}{D\cos\alpha} \tag{1}$$

and

$$\tan\theta_{\mathrm{PR}} = \frac{D\sin\alpha - I/2}{D\cos\alpha}, \tag{2}$$

respectively.

Assuming that $D^2 \gg I^2$, the optic array disparity of this point, $\gamma_{\mathrm{P}} = \theta_{\mathrm{PR}} - \theta_{\mathrm{PL}}$, is given by

$$\tan\gamma_{\mathrm{P}} \approx \frac{-I\cos\gamma}{D}. \tag{3}$$

Now consider a second point Q, at a position $(X, Z + d)$, assuming that (i) $D \gg I$ and (ii) $D \gg d$, its optic array disparity is given by

$$\tan\gamma_{\mathrm{Q}} \approx \frac{-ID\cos\alpha + Id}{D^2}. \tag{4}$$

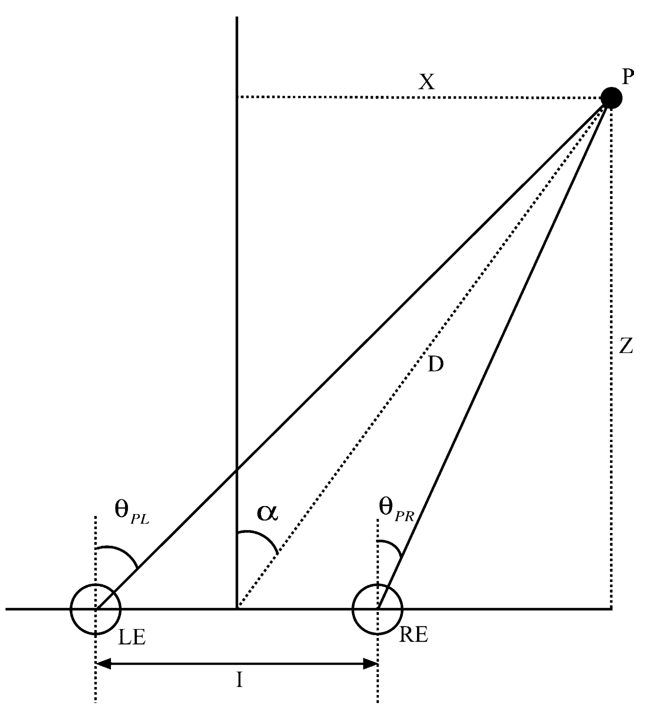

Fig. 7. The geometry of a point P at a depth Z, and horizontal position X from the cyclopean eye. The visual direction of this point in the left eye (LE), right eye (RE) and cyclopean eye is given by θ_{PL}, θ_{PR} and α, respectively. The inter-ocular distance is denoted by I, and the distance of P from the cyclopean eye by D.

With fixation on point P, the disparity of point Q, γ, is given by

$$\tan\gamma = \frac{-ID^2\cos\alpha + IDd + ID^2\cos\alpha}{D^3 + DI^2\cos^2\alpha - I^2 d\cos\alpha}. \tag{5}$$

Again, assuming (i) $D \gg I$ and (ii) $D \gg d$

$$\tan\gamma \approx \frac{Id}{D^2}. \tag{6}$$

Now consider a point W, at position $(X + w, Z)$

Then

$$\tan\theta_{\mathrm{WL}} = \frac{D\sin\alpha + I/2 + w}{D\cos\alpha}. \tag{7}$$

If $\phi = \theta_{\mathrm{WL}} - \theta_{\mathrm{PL}}$, and assuming $w \ll D$, then

$$\tan\phi \approx \frac{w\cos\alpha}{D}. \tag{8}$$

Let

$$S = \tan\gamma = \frac{Id}{D^2}. \tag{9}$$

If we assume that the current distance and direction of the point from the observer are given by D and α, that the (constant) radial and angular velocity are given by $\dot{D}$ and $\dot{\alpha}$, and that the location of the point at time $t = 0$ was D_0 and α_0, then

$$S = \frac{Id}{(D_0 + \dot{D}t)^2} \tag{10}$$

and

$$\frac{\dot{S}}{S} = -\frac{2\dot{D}}{D}. \tag{11}$$

Let

$$T = \tan\phi = \frac{w\cos\alpha}{D} = \frac{w\cos(\alpha_0 + \dot{\alpha}t)}{D_0 + \dot{D}t}. \tag{12}$$

Then

$$\frac{\dot{T}}{T} = -\dot{\alpha}\tan\alpha - \frac{\dot{D}}{D}. \tag{13}$$

From this it is clear that if either $\dot{\alpha}$ or $\alpha = 0$

$$\left(\frac{\dot{S}/S}{\dot{T}/T}\right) = 2. \tag{14}$$

In other words, the ratio of the change in disparity divided by the disparity itself, to the change in retinal size divided by the retinal size itself, equals 2 for an object moving rigidly in depth at a constant speed along the cyclopean line of sight.

To test the stability of (14) and its applicability in the current experimental situation, simulations were carried out. Objects moving along the cyclopean line of sight where α and $\dot{\alpha} = 0$ represent a restricted situation. For Eq. (14) to be a useful strategy the ratio should be equal to 2 for a range of α and $\dot{\alpha}$. Simulations were carried out using the parameters of the experimental setup to access (14) for a range of depth path angles (θ in Fig. 1). Over these different path angles both α and $\dot{\alpha}$ vary. The simulations showed that Eq. (14) gives a good approximation of a rigidly moving object even without the additional information of α and $\dot{\alpha}$. Over a range of depth path angles (0° to 40°), error in (14) reached a maximum of 10.1% at 40°, and decreased rapidly with decreasing path angle. For the path angle of the experiment, error in (14) was 5.28%. This suggests that Eq. (14) represents a useful strategy for observers to adopt. It should be noted that, even if observers did not adopt (14), α and $\dot{\alpha}$ should be readily available from the optic array, thus allowing the strategy to be applied more generally.

References

Bradshaw, M. F., & Elliott, K. M. (2003). The role of binocular information in the 'on-line' control of prehension. *Spatial Vision, 16*, 295–309.

Bradshaw, M. F., Parton, A. D., & Eagle, R. A. (1998). The interaction of binocular disparity and motion parallax in determining perceived depth and perceived size. *Perception, 27*, 1317–1331.

Bradshaw, M. F., Parton, A. D., & Glennerster, A. (2000). The task-dependant use of binocular disparity and motion parallax information. *Vision Research, 40*, 3725–3734.

Brenner, E., & Landy, M. S. (1999). Interaction between the perceived shape of two objects. *Vision Research, 39*, 3834–3848.

Brenner, E., & van Damme, W. J. M. (1998). Judging distance from ocular convergence. *Vision Research, 38*, 493–498.

Brooks, K. R., & Stone, L. S. (2004). Stereomotion speed perception: Contributions from both changing disparity and interocular velocity difference over a range of relative disparities. *Journal of Vision, 4*, 1061–1079.

Buckley, D., & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research, 33*, 919–933.

Burt, P., & Julesz, B. (1980). Modifications of the classical notion of Panum's fusional area. *Perception, 9*, 671–682.

Champion, R. A., Simmons, D. R., & Mamassian, P. (2004). The influence of object size and surface shape on shape constancy from stereo. *Perception, 33*, 237–247.

Cuijpers, R. H., Kappers, A. M. L., & Koenderink, J. J. (2000). Large systematic deviations in visual parallelism. *Perception, 29*, 1467–1482.

Durgin, F. H., Proffitt, D. R., Olson, T. J., & Reinke, K. S. (1995). Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 679–699.

Foster, D. H., & Bischoff, W. F. (1991). Thresholds from psychometric functions: Superiority of bootstrap to incremental and probit variance estimators. *Psychological Bulletin, 109*, 151–159.

Frisby, J. P., Buckley, D., & Horsman, J. M. (1995). Interaction of stereo, texture, and outline cues during the pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception, 24*, 181–198.

Frisby, J. P., Buckley, D., & Duke, P. A. (1996). Evidence for good recovery of lengths of real objects seen with natural stereo viewing. *Perception, 25*, 129–154.

Gårding, J., Porrill, J., Mayhew, J. E. W., & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Research, 35*, 703–722.

Girshick, A. R., Akeley, K., Watt, S. J., & Banks, M. S. (2004). Improved stereoscopic performance with consistent vergence and accommodative cues in a novel 3-D display. *Perception, 33*, 42.

Glennerster, A., Gilson, S. J., Tcheang, L., & Parker, A. J. (2003). Perception of size in a 'dynamic Ames room'. *Journal of Vision, 3*, 490a.

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research, 36*, 3441–3456.

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1998). Cues to viewing distance for stereoscopic depth constancy. *Perception, 27*, 1357–1365.

Harris, J., & Watamaniuk, S. N. J. (1995). Speed discrimination of motion-in-depth using binocular cues. *Vision Research, 35*, 885–896.

Hershenson, M. H. (1999). *Visual space perception: A primer*. London: MIT Press.

Hibbard, P. B., Bradshaw, M. F., Langley, K., & Rogers, B. J. (2002). The stereoscopic anisotropy: Individual differences and underlying mechanisms. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 469–476.

Hibbard, P. B., & Bradshaw, M. F. (2003). Reaching for virtual objects: Binocular disparity and the control of prehension. *Experimental Brain Research, 148*, 196–201.

Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth: Depth perception*. Toronto: Toronto University Press.

Hung, G. K., Zhu, H., & Ciuffreda, K. J. (1997). Convergence and divergence exhibit different response characteristics to symmetrical stimuli. *Vision Research, 37*, 1197–1205.

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

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research, 34*, 2259–2275.

Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.

Koenderink, J. J., van Doom, A. J., & Lappin, J. S. (2000). Direct measurement of the curvature of visual space. *Perception, 29*, 69–79.

Koenderink, J. J., van Doom, A. J., Kappers, A. M. L., & Todd, J. T. (2002). Pappus in optical space. *Perception & Psychophysics, 64*, 380–391.

Landy, M. S., & Brenner, E. (2001). Motion–disparity interaction and the scaling of stereoscopic disparity. In L. R. Harris & M. R. M. Jenkin (Eds.), *Vision and attention*. New York: Springer Vertag.

Loomis, J. M., & Philbeck, J. W. (1999). Is the anisotropy of perceived 3-D shape invariant across scale. *Perception & Psychophysics, 61*, 397–402.

Loomis, J. M., Philbeck, J. W., & Zahorik, P. (2002). Disassociation between location and shape in visual space. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 1202–1212.

Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.

MonWilliams, M., & Tresilian, J. R. (1999). Some recent studies on the extra retinal contribution to distance perception. *Perception, 28*, 167–181.

Morgan, M. J. (1989). Vision of solid objects. *Nature, 339*, 101–103.

Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 173–186.

Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research, 43*, 2451–2468.

Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America A, 2*, 343–349.

Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature, 361*, 253–255.

Rogers, B. J., & Bradshaw, M. F. (1995). Disparity scaling and the perception of frontoparallel surfaces. *Perception, 24*, 155–179.

Rogers, B. J., & Collett, T. S. (1989). The appearance of surfaces specified by motion parallax and binocular disparity. *The Quarterly Journal of Experimental Psychology, 41A*, 697–717.

Scarfe, P., & Hibbard, P. B. (2004). Noise in horizontal-disparity and vergence signals predicts systematic distortions in the estimation of shape. *Perception, 33*, 95.

Servos, P., Goodale, M. A., & Jakobson, L. S. (1992). The role of binocular vision in prehension: A kinematic analysis. *Vision Research, 32*, 1513–1521.

Tcheang, L., Gilson, S. J., & Glennerster, A. (2005). Systematic distortions of perceptual stability investigated using immersive virtual reality. *Vision Research, 45*, 2177–2189.

Tittle, J. S., & Braunstein, M. L. (1993). Recovery of 3-D shape from binocular disparity and structure from motion. *Perception & Psychophysics, 54*, 157–169.

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 663–678.

Todd, J. T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences, 8*, 115–121.

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*, 31–47.

Todd, J. T., Tittle, J. S., & Norman, J. F. (1995). Distortions of three-dimensional space in the perceptual analysis of motion and stereo. *Perception, 24*, 75–86.

van Damme, W., & Brenner, E. (1997). The distance used for scaling disparities is the same as the one used for scaling retinal size. *Vision Research, 37*, 757–764.

Wagner, M. (1985). The metric of visual space. *Perception & Psychophysics, 38*, 483–495.

Watt, S. J., Akeley, K., Girshick, A. R., & Banks, M. S. (2005). Achieving near-correct focus cues in a 3-D display using multiple image planes. *Proceedings of SPIE Human: Vision and Electronic Imaging*, 5666–5673.