



The task-dependent use of binocular disparity and motion parallax information

Mark F. Bradshaw^{a,*}, Andrew D. Parton^a, Andrew Glennerster^b

^a *Department of Psychology, School of Human Sciences, University of Surrey, Guildford GU2 7XH, UK*

^b *University Laboratory of Physiology, University of Oxford, Parks Road, Oxford OX1 3PT, UK*

Received 19 October 1999; received in revised form 3 July 2000

Abstract

Binocular disparity and motion parallax are powerful cues to the relative depth between objects. However to recover absolute depth, either additional scaling parameters are required to calibrate the information provided by each cue, or it can be recovered through the combination of information from both cues (Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America*, 2, 343–349). However, not all tasks necessarily require a full specification of the absolute depth structure of a scene and so psychophysical performance may vary depending on the amount of information available, and the degree to which absolute depth structure is required. The experiments reported here used three different tasks that varied in the type of geometric information required in order for them to be completed successfully. These included a depth nulling task, a depth-matching task, and an absolute depth judgement (shape) task. Real world stimuli were viewed (i) monocularly with head movements, (ii) binocularly and static, or (iii) binocularly with head movements. No effect of viewing condition was found whereas there was a large effect of task. Performance was accurate on the matching and nulling tasks and much less accurate on the shape task. The fact that the same perceptual distortions were not evident in all tasks suggests that the visual system can switch strategy according to the demands of the particular task. No evidence was found to suggest that the visual system could exploit the simultaneous presence of disparity and motion parallax. © 2000 Elsevier Science Ltd. All rights reserved.

1. Introduction

A large number of recent studies have examined the extent to which binocular disparity and motion parallax information support veridical judgements of an object's depth, size and distance (see for example Todd & Bressan, 1990; Collett, Schwartz, & Sobell, 1991; Johnston, 1991; Todd & Norman, 1991; Rogers & Bradshaw, 1993; Liter, Braunstein, & Hoffman, 1994; Todd, Tittle, & Norman, 1995; Tittle, Todd, Perotti, & Norman, 1995; Rogers & Bradshaw, 1995a; Bradshaw, Glennerster, & Rogers, 1996; Glennerster, Rogers, & Bradshaw, 1996; Norman, Todd, Perotti, & Tittle, 1996). These experiments have largely concluded that depth constancy is considerably less than perfect, perceived shape is distorted and absolute distance is misestimated — despite the fact that sufficient information is available to support veridical perception.

The generality of this conclusion was questioned however by Bradshaw, Parton, and Eagle (1998) on a number of grounds including (i) the use of computer displays, (ii) the number of visual cues available, (iii) the nature of the perceptually guided task being performed, (iv) the size of the stimuli, and (v) the general experimental parameters used. The present experiments were designed in the context of points (i) to (iii) to determine the degree to which binocular disparity and/or motion parallax information (using carefully controlled, real world stimuli) can support the recovery of the visual information required to perform three separate experimental tasks. Using different tasks in depth constancy experiments is interesting as differences in performance can provide evidence about the nature of the representation involved in their completion (see for example Tittle et al., 1995; Glennerster et al., 1996). The nature of the physical world is usually described in terms of Euclidean geometric relations and it is often assumed that our perceptual representation of that world would embody the same correctly scaled descrip-

* Corresponding author. Fax: +44-1483-532813.

E-mail address: m.bradshaw@surrey.ac.uk (M.F. Bradshaw).

tive geometries. However the fact that such large and systematic departures from veridical perception and depth constancy have been reported challenges this notion.

The recovery of Euclidean geometry (which defines distance, angles and curves uniquely within an isotropic distance metric) from binocular disparity or motion parallax information poses a very difficult problem to the visual system. Not only must the retinal disparities and motions be measured accurately but, to be converted into metric depth information, knowledge of fixation distance and the inter-ocular separation are necessary in the case of disparity; and in the case of motion, knowledge of eye rotation and ego-motion is required (Koenderink and van Doorn, 1991). If these additional sources of information cannot be recovered reliably then perceived space will be a distorted version of physical space and failures in depth constancy will result. This may not be as serious for everyday purposes as it first seems, as many tasks can be completed successfully without deriving a full specification of metric Euclidean structure. For example, Glennerster et al. (1996) noted that to thread a needle only the signed disparities (or motions) of the thread and needle are required; and to compare the shapes of two surfaces, knowledge of the ratio of the depths of the features and the distance to the surfaces is sufficient. Therefore, it is possible that the visual system needs to reconstruct Euclidean structure only rarely.

The present experiments used three experimental tasks, based on the above logic, to determine the extent to which metric structure is recovered by the visual system. These included (i) a depth nulling task where one point had to be set equidistant to two reference points (see Fig. 1), (ii) a depth matching task where the depth of two stimuli presented at different distances had to be equated (see Fig. 3), and (iii) a shape judgement task where the depth of a stimulus had to be equated with its width (see Fig. 5). The information available to the observer was also manipulated as this may also influence (or limit) the nature of the representation/s involved in the control of a particular task. In different conditions stimuli were defined by binocular disparity alone, motion parallax alone or binocular disparity and motion parallax. The latter case is of interest as it has been demonstrated that Euclidean structure can be recovered through the combination of the cues without recourse to extra-retinal scaling factors (see below) whereas it has been argued that extra-retinal cues may be required in the single cue conditions.

It is important here to consider the type of information that is sufficient to perform each of the experimental tasks and in which experimental conditions it is available. When only two views (generated by binocular viewing or motion parallax) are available, signed disparities (or relative motions) can be recovered, which

are sufficient to support a range of activities such as the *nulling task* (see Garding, Porrill, Mayhew, & Frisby, 1995; Glennerster et al., 1996). The recovery of disparity, or motion parallax signals, provide only the ratios of depths of points on a surface. This is known a 'bas relief structure', and requires no calibration of the disparity or motion vectors. With further information, other tasks become possible. For example, in the matching task (described above), subjects need to do more than simply recover the bas relief structure. Given a reference surface at one distance, they must choose the correct depth setting using a matching surface presented at a different viewing distance. To do this correctly requires only a judgement of the ratio of viewing distances: a subject who estimated the two surfaces to be at 50 cm and 1 m would make the same setting as a subject who estimated them to be at 1 and 2 m, although their estimates of the 'shapes' of the stimuli (if they were asked this) would be quite different. Finally, if an estimate of the viewing distance, based on vergence angle or equivalent, is incorporated in the calculation, then the bas relief ambiguity can be resolved and the structure determined uniquely (Euclidean structure). This is sufficient to support veridical performance in the shape task, which unlike the matching task, requires a comparison of a distance in the fronto-parallel plane (width) with a distance along the z -axis (depth). The hierarchical stratification of tasks described here is similar to that described by Todd and Bressan (1990), Koenderink and van Doorn (1991) where successive levels of the hierarchy are characterised by more tightly constrained solutions, culminating in a unique, Euclidean description of surface shape.

When three or more views are available the situation is somewhat different. Ullman (1979) showed that given four non-coplanar points viewed in orthographic projection, three frames provide sufficient information to recover 3-D shape up to an isotropic size scaling. In theory, therefore, it is possible to recover metric structure (up to an isotropic scaling factor) directly when an additional third frame is available in a motion sequence (see also Koenderink & van Doorn, 1991). This scheme is distinct from the 2-view case as the information added by the third frame is sufficient to determine a unique set of solutions. However, although possible theoretically, it has been suggested in practice that motion information is not sufficient, in isolation, to determine 3-D shape without bias (Braunstein, Litter, & Tittle, 1993; Litter et al., 1994; Norman & Todd, 1993). Indeed, it has been shown that initial image measurements (e.g. of image acceleration, upon which the computation of Euclidean structure depends) are extremely sensitive to noise (Eagle & Blake, 1995). This can lead to systematic biases in perceived 3D structure (Hoger-vorst & Eagle, 1998).

So far we have discussed the conditions where disparity or parallax information is available in isolation. The third experimental viewing condition provides the cues simultaneously (i.e. at least two binocular pairs of views are available). This condition was included to assess possible interactions between the two cues in the recovery of useful information about the world. It has been shown that, in principle, disparity and parallax could be combined to recover veridical information about depth and distance without recourse to additional scaling parameters (Richards, 1985). This scheme takes advantage of the possibility, referred to above, that motion information can be used to determine the shape of a projected object (up to an isotropic size scaling), which in turn, can be used to distinguish between the family of surfaces (parameterised by vergence angle or viewing distance) that is consistent with a particular pattern of binocular disparities. Through this process absolute size and distance could, in theory, also be determined and so provide a further avenue for the recovery of full metric structure. Some evidence has been found in support of this model (Richards & Lieberman, 1985; Johnston, Cumming, & Landy, 1994; Bradshaw et al., 1998) although other evidence is less supportive of the model suggesting that the cues may remain relatively independent in the recovery of depth (e.g. Bradshaw, Frisby, & Mayhew, 1987; Tittle et al., 1995; Brenner & van Damme, 1997; Brenner & Landy, 1999). The stimuli we used here differed in several respects from those used in previous experiments. We consider the implications of these differences in the Section 5.

The three experimental tasks were investigated in turn in the three experiments reported below. Although each task could, in principle, be completed successfully on the basis of a single internal representation or model of the world (e.g. Marr & Nishihara, 1978) that maintains Euclidean spatial relations, the degree to which such a representation is computed should be revealed by any difference in performance on the different tasks. Glennerster et al. (1996), for example, found depth constancy was 75% in their shape task whereas it was close to 100% in their depth-matching task. This led them to conclude that different strategies must be available to the observer when faced with different tasks (see also Koenderink & van Doorn, 1991). The same comparison of tasks has not been carried out for stimuli defined by motion parallax or in the situation where both disparity and parallax information is available.

The critical comparisons are between the results for the matching and shape tasks. If the visual system constructed a single Euclidean representation, irrespective of task, then any distortion present in the results of the shape task should also be evident in the results of the matching task. We examine this comparison for

binocular disparity and motion parallax cues when they are presented in isolation or together. Performance in the nulling task is predicted to be non-biased in each viewing condition.

The use of 'real-world' stimuli in the present experiments may be important as it has often been argued that the use of computer displays may have contributed to the poor performance in depth constancy experiments due to the range of potential artefacts associated with such stimuli. For example, Frisby, Buckley, and Duke (1996) reported that Weber fractions were almost three times better in a 3D length discrimination task when they replicated the simulated stimuli of Todd and Bressan (1990) using real objects (knarled sticks) and natural viewing (see also Durgin, Proffitt, Olson, & Reinke, 1995; Bradshaw et al., 1998). However, in real world situations with well-illuminated environments, it is often not possible to maintain adequate control over the range of cues available and so it is difficult to attribute performance to any particular depth cue, or group of depth cues. Improvement in performance, therefore, such as that illustrated above, can not be attributed unambiguously to difficulties with simulated stimuli per se. It is interesting to note in this regard that performance in Frisby et al.'s monocular viewing conditions remained surprisingly good (although no recognised depth cue was present): this would have been unlikely in the experiments of Todd and Bressan.

To counter such difficulties, rigorous control over the viewing conditions and the contribution of extraneous cues was maintained in the present experiments. This was particularly important if our second goal of assessing the interaction of disparity and parallax information was to be undertaken in a meaningful way. Therefore the 'real-world' stimuli developed here remained rather artificial (small lights in a dark room) in construction and presentation (c.f. Bradshaw et al., 1998) although they were free of sub-optimal spatial or temporal sampling, accommodation conflicts and lack of natural viewing geometry, often associated with CRT displays. Accommodation conflicts, for example, would have been significant if our stimuli had been presented on a single CRT screen, since they spanned the range from 1.5 to 3 m. To assess the efficacy of our experimental controls, a fourth viewing condition was included for each task: static monocular viewing. For our binocular and motion parallax experiments to be meaningful, the performance in this control condition should be poor relative to the other three viewing conditions.

Each experiment was carried out at 1.5 and 3.0 m to determine whether performance varied with viewing distances and so that estimates of depth constancy could be obtained.

2. Experiment 1 — the nulling task

The nulling task used in this experiment was based on the Howard–Dolman stereo acuity test (Howard, 1919). Observers were required to move a LED along the centre line so that it appeared equidistant or co-linear with two fixed LEDs positioned on either side of it (see Fig. 1). Theoretically, to complete this task only signed relative motion or disparity is required (i.e. a full recovery of depth is not necessary). Note that the difference between the fronto-parallel plane and the Veith–Muller circle is negligible at the viewing distances and stimulus extents used here. Although threshold performance is well established for both disparity-defined surfaces (Howard, 1919; Julesz, 1971; Bradshaw & Rogers, 1999) and parallax-defined surfaces (Graham, Baker, Hecht, & Lloyd, 1948; Rogers & Graham, 1979), no study has yet explicitly compared performance using physical stimuli and the same observers.

2.1. Method

2.1.1. Observers

Four experienced psychophysical observers with normal or corrected-to-normal acuity participated as unpaid volunteers.

2.1.2. Apparatus

Three bright yellow LEDs were presented in the dark with any additional sources of light, and surface reflections, removed. Two of the LEDs were fixed in position 3.8° either side of the centre line. Their midpoint lay either 150 or 300 cm along the centre line from the observer. The third LED could be moved at a constant speed (0.1 m/s) backwards or forwards along the centre line and was under observer control. Great care was taken in positioning the lights to prevent the observers being able to employ a strategy using 2-D information

to complete the task. The LEDs were arranged to lie in the same horizontal plane as the observers' eyes to prevent changes in the vertical position with changes in distance and their brightness was randomised between trials. The moving light flashed at a rate of 5 Hz to eliminate the sense of perceived motion.

A headrest was used that was either fixed, or free to move from side-to-side. For the static conditions the headrest was fixed such that the centre line of the stimuli was aligned with either the dominant eye (monocular viewing) or midway between the eyes (binocular viewing). In the moving conditions the observers were required to move their head 6.5-cm either side of the centre line (twice the interocular distance). They moved their heads at a rate of 1 Hz and were paced by a metronome.

Observers wore standard ear-defenders (UltraMuff II, Racal Safety Ltd.) throughout the trials and conversation between experimenter and observer was avoided. The experiment was run in total darkness but between each trial the room lights were switched on to enable the experimenter to record the results and set-up the next trial.

The adjustable LED started at a random position in front or behind the targets. The experiments were performed under the following four viewing conditions (i) monocular static, (ii) binocular static, (iii) monocular with head motion and (iv) binocular with head motion. Ten settings were made within each condition at both viewing distances (150 and 300 cm). The experiments were blocked by condition and viewing distance, which were ordered randomly.

2.2. Results

The results from individual subjects were similar so performance, averaged over observers is presented in Fig. 2. As predicted, performance in the disparity and/or parallax conditions was reliable and showed no bias. Throughout the paper the motion parallax settings were converted into units of 'equivalent disparity' for comparison and further computation. Equivalent disparity is defined as the relative motion created as the head moves through the inter-ocular distance — 6.5 cm (see Rogers & Graham, 1982 for further details). Settings expressed as min arc disparity, or equivalent disparity, are presented in Fig. 2b.

One-way *t*-tests revealed that none of these conditions differed significantly from veridical (i.e. zero error) and an ANOVA revealed that there were no significant differences between them. Settings in the parallax only condition were generally higher than those made when disparity was present. This is consistent with previously reported findings from studies using CRT displays (Rogers & Graham, 1982; Bradshaw & Rogers, 1996) although the magnitude of the differ-

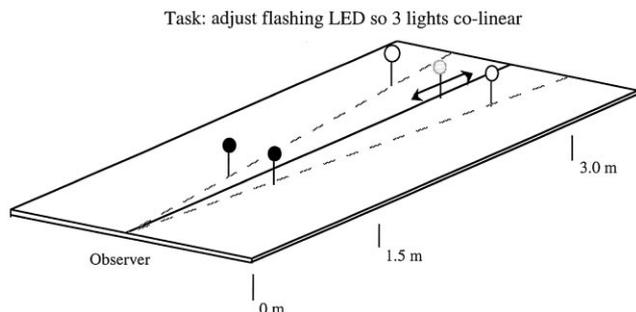


Fig. 1. Illustrates the procedure in the nulling task. In the illustration the task, of aligning the central LED to be equi-distant with the flanking LEDs, is being performed at 3.0 m. (○) LEDs switched on, (●) LEDs switched off (i.e. used for the experiment at 1.5 m viewing distance) and the dotted circle represents a flashing LED. The dashed lines projecting from the observer denote a visual angle of $\pm 3.8^\circ$.

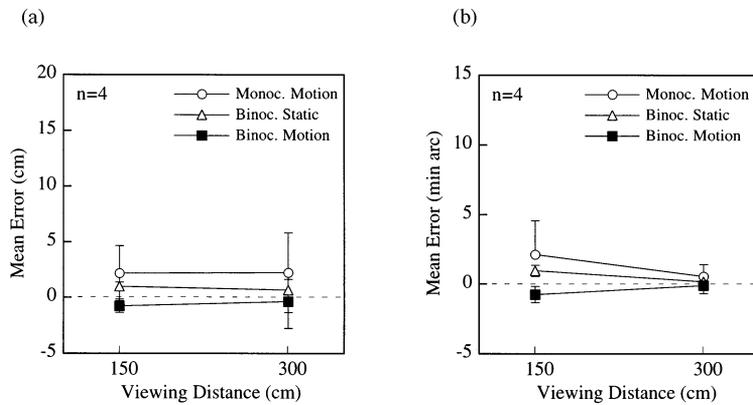


Fig. 2. Plots the mean ($n = 4$) depth settings, in cm (a) and min arc (b), as a function of viewing distance. The error bars indicate average within subject error. (○) Monocular motion, (△) static binocular viewing and (■) binocular motion. The dashed horizontal line indicates veridical performance.

ence was greater than that previously reported. When both cues were present, thresholds appeared to be determined by the most sensitive component (i.e. disparity). This would be predicted by many cue combinations models (e.g. Landy, Maloney, Johnston, & Young, 1995).

S.D. of 10 and 31 cm (at 150 and 300 cm respectively) were found in the static monocular condition, which were much greater than those in the other three conditions (on average 4.5 and 9.7 cm). Mean error in this condition (13.5 and 23 cm at 150 and 300 cm respectively) was also significantly worse when compared to the other three conditions ($F_{3,9} = 9.066$; $P < 0.005$). These results show that in the main experimental conditions, subjects really did use the binocular disparity and motion parallax to perform the tasks.

3. Experiment 2 — the matching task

Here, observers were required to adjust the base-to-apex distance of one triangle, defined by three LEDs, in order to match the base-to-apex distance of a reference triangle also defined by three LEDs but located at a different viewing distance (see Fig. 3).

Signed disparities that may have been used to complete the nulling task are insufficient for this task. However, in principle, a full Euclidean representation is not required either but rather the task could be completed (i.e. the relative disparities or motions could be matched) if the ratio of the viewing distances to the two displays was recovered accurately (see Glennerster et al., 1996) and used to scale the depth information. The angular sizes of the base LEDs would be sufficient to specify their relative distances — the absolute distance to either is unnecessary.

3.1. Method

The participants, experimental apparatus and stimuli were the same as described above except for the following details. A total of six LEDs were used to create a ‘fixed-size’, reference triangle and an ‘adjustable’ triangle (see Fig. 3). The two triangles were visible throughout each trial where the base-to-apex distance of the adjustable had to be set to match the base-to-apex distance of the reference triangle. The LEDs at the base of the reference were positioned 212 cm from the observer and at ± 20 cm ($\pm 5.4^\circ$) to the centre line. The base-to-apex distance was set to be either 20 or 60 cm depending on the condition. The LEDs of the reference were positioned 2 cm above the horizontal plane, which contained the observer’s eyes and the LEDs of the adjustable ‘triangle’. The separation of the LEDs, which formed the base of the adjustable triangle was fixed at 40 cm and was positioned at either 150 cm ($\pm 7.6^\circ$) or 300 cm ($\pm 3.8^\circ$). The central LED (defining its apex) could be moved along the centre line by the observer (as in experiment 1). The correct setting at 300 cm for the smallest reference triangle corresponded to about 5 arc min of the relative displacement between

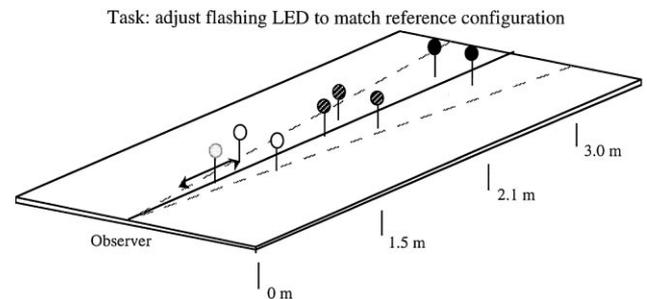


Fig. 3. Illustrates the procedure in the matching task. In the illustration the task, of setting the adjustable LED to match the base-to-apex distance of the reference triangle (hatched circles), is being performed at 150 cm. (●) represent LEDs that are switched off.

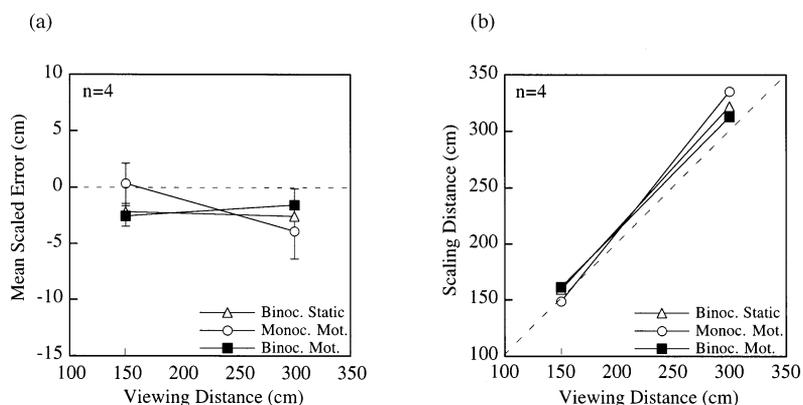


Fig. 4. (a) Plots the mean ($n = 4$) scaled error settings in cm for the small and large triangles combined as a function of viewing distance. The error bars indicate average within subject error. (b) Plots the 'estimated' scaling distance at which the angular extent of the depth settings would have been veridical as a function of the viewing distance. The dashed line indicates veridical performance (a) and perfect depth constancy (b). The different symbols represent the individual viewing conditions as shown in the legend.

the base and apex LEDs, which was well above the magnitude of the settings established in experiment 1 (except in the static monocular condition).

3.2. Results

The results from the 20 and 60 cm reference triangles were combined as their departures from veridical when scaled relative to the 20 cm triangle (i.e. signed error) were found not to differ significantly ($F_{1,3} = 5.579$ $P > 0.05$). Performance in each condition, expressed as mean scaled error of the veridical base-to-apex distance, are presented in Fig. 4a.

Again performance was poorest in the static monocular condition. S.D. were 43 and 120% of veridical (at 150 and 300 cm respectively) compared to 18 and 25% in the other three conditions. Mean error in this condition was 16 and 35%.

When disparity and/or parallax was present, mean settings indicated a slight tendency to underestimate the size of the base-to-apex distance, which is consistent with previous data using disparity-defined surfaces and viewing distances greater than 1 m. However, performance showed no significant bias ($P > 0.05$) in any of these conditions. The mean settings, averaged across the disparity and parallax conditions, departed from veridical by only 7% at 150 cm and 12% at 300 cm.

To illustrate the extent of depth constancy in the experiment we follow convention and calculate the 'estimated scaling distances' as described previously by Glennerster et al. (1996). This also allows an explicit comparison with the results from the shape task reported in experiment 3 below. The estimated scaling distance is defined as the distance at which the disparity, or equivalent disparity, chosen by the observer in setting the adjustable triangle would produce the physically correct base-to-apex distance of the reference tri-

angle. Although we argue that knowledge of the absolute distances to either triangle is not necessary to make a correct match (i.e. only the ratio of viewing distances is required), for this calculation we assume that the reference is seen veridically at 212 cm. The converted settings are plotted in Fig. 4b.

The dashed line (slope of 1) in Fig. 4b depicts perfect constancy. The data show only a small departure from this veridical line with the slopes of the plot lines being 1.24, for monocular motion, and ≈ 1.03 in the binocular conditions. The results are similar to those reported by Glennerster et al. (1996) who used a similar task but only disparity defined, simulated stimuli.

4. Experiment 3 — the shape task (depth = width)

This task was based on the 'apparently circular cylinder' task developed by Johnston (1991) but here the observer was required to adjust the base-to-apex distance of a triangle, defined by three LEDs, to match the separation between the base LEDs (i.e. to set $b = h$ in Fig. 5).

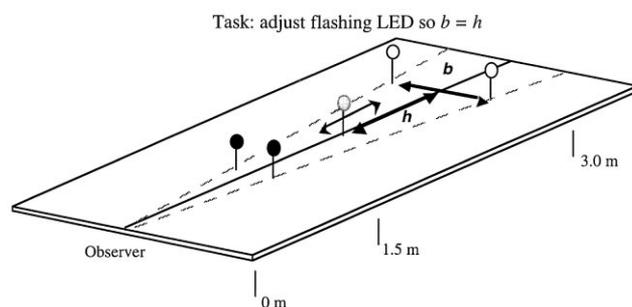


Fig. 5. Illustrates the procedure for the shape task. In the illustration the task, of setting the base-to-apex distance to be equal to the separation of the base LEDs (i.e. $b = h$) is being performed at 150 cm. (○) Lit LEDs, (●) unlit LEDs and the dotted circle represents an adjustable flashing LED.

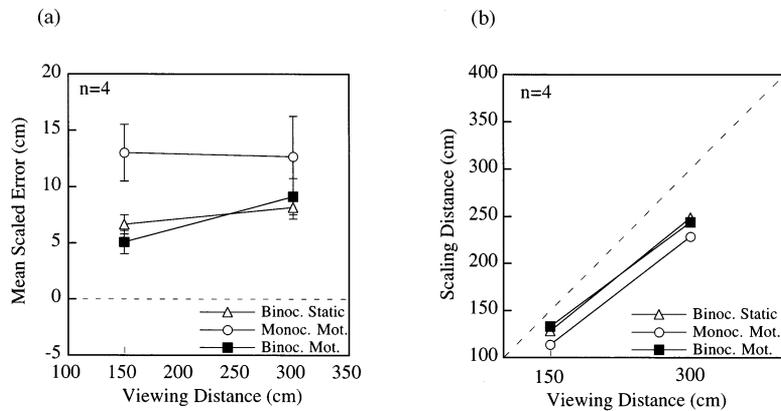


Fig. 6. (a) Plots the mean ($n=4$) scaled error settings relative to the 20 cm triangle, in cm, as a function of viewing distance. The error bars indicate average within subject error. (b) Plots the 'estimated' scaling distance at which the angular extent of the depth settings would have been veridical as a function of the viewing distance. The dashed line indicates veridical performance (a) and perfect depth constancy (b). The different symbols represent the individual viewing conditions as shown in the legend.

This task was chosen because, on the basis of binocular disparity information, a representation that specifies Euclidean structure is necessary for it to be completed successfully. To recover this information from two views (e.g. for static binocular viewing) the relative disparities must be scaled by an estimate of the viewing distance. In the case of motion parallax two different routes for the computation of Euclidean structure are possible. Using only two frames, a method analogous to the binocular case could be used. Using three or more frames, methods based on structure-from-motion (SFM) algorithms could be used, as described in the Section 1, to recover structure up to an isotropic size scaling. The combination of disparity and parallax information may be critical for this task as it affords the possibility of recovering metric structure without recourse to extra-retinal scaling factors (Richards, 1985). Therefore, performance in the disparity and parallax condition may be different from that in the individual cue conditions.

4.1. Method

Except were mentioned, the participants, experimental apparatus and stimuli were the same as described above. The stimuli consisted of three LEDs. The position of one LED, could be adjusted by the observer to move along the centre line while the others were fixed at either 150 or 300 cm, $\pm 3.8^\circ$ either side of the centre line (as in experiment 1). The base separation was therefore 20 cm at 150 cm and 40 cm at 300 cm. Note that, unlike the last experiment, the angular size of the separation between the base LEDs is not a cue to distance. The task was to adjust the base-to-apex distance to match the physical separation between the base LEDs. If viewing distance is taken into account completely then the setting must be twice as large in cm, but

half as large in terms of angular units at the further distance. On each trial, the initial location of the LED was set at random with an equal number of trials starting either side of the veridical position.

4.2. Results

The mean performance in each viewing condition, averaged over observers, is shown in Fig. 6a as mean scaled error (relative to the 20 cm triangle). Systematic biases in the settings are evident within all depth-cue conditions and one-way t -tests revealed them to be significantly different ($P < 0.05$) from veridical in all cases. This performance is markedly different from that found in experiment 1 or 2 where no significant biases were found. Therefore it appears that the shape task is performed less well than the other tasks despite the fact that information about viewing distance is similar in all cases.

In order to quantify the degree to which viewing distance was taken into account in the observers' depth settings 'effective scaling distances' were again computed in a similar manner to that discussed above. The effective scaling distance is the distance at which the observers' setting in units of disparity or equivalent disparity, would correspond to a veridical response (i.e. a depth interval of 20 or 40 cm at 150 or 300 cm respectively). The transformed data is presented in Fig. 6b with the diagonal hatched line indicating a perfect correspondence between the estimated and actual viewing distance. It is evident that the data in each condition indicate a high degree of depth scaling with slopes of ≈ 0.76 in the binocular conditions and 0.8 in the monocular parallax conditions. The fact the depth constancy estimates are less than 1.0 suggests that intervals in depth are increasingly compressed with viewing distance (see Foley, 1980; Johnston, 1991) in each condition.

Overall, the results of the experiment show neither the near perfect (veridical) performance reported by Durgin et al. (1995) and Frisby et al. (1996) nor the extremely poor levels of depth constancy ($\approx 30\%$) reported by Johnston (1991). The magnitude of the depth constancy estimates in the disparity condition are close to those reported by Glennerster et al. (1996).

5. General discussion

The present set of experiments were designed to establish the ability of the visual system to perform a range of 3D judgements when stimuli were defined by binocular disparity, motion parallax or both cues together. Although there was little effect of the type of depth cue available within each experiment, there was a strong effect of task between the experiments. This difference occurred despite the fact that each experiment was performed (as far as possible) under identical conditions.

Observers performed well on the nulling task. Although this task could have been solved on the basis of a full specification of metric depth, simply knowing the sign of the disparity or the motion signal is sufficient for successful performance. The results from this task demonstrated that accurate and reliable settings were possible within our experimental paradigm. Observers also performed well on the depth-matching task. In contrast, in the depth-to-width or shape task, settings differed significantly from veridical. It is the pattern of results between the latter two tasks that leads us to suggest that the nature of the representation used in their completion may be different. In both situations, the disparity and motion information generated by veridical settings was well above threshold, but the subsequent processing required to support each task is not necessarily the same. In principle, as described in the Section 1, the use of the ratio of the viewing distances to calibrate and set the disparities and motions is sufficient to support veridical performance in the matching task. That is, the recovery of Euclidean structure (although also sufficient) is not necessary. In contrast, in order to complete the shape task veridically, information about Euclidean structure is required. The results are consistent with the suggestion that the visual system may switch perceptual strategy on the basis of the available information and demands of a particular task (see Rogers and Bradshaw, 1995b; Tittle et al., 1995; Glennerster et al., 1996). If subjects used a single strategy in the matching and shape tasks (e.g. the one that works for matching — setting the depth according to the square of the estimated distance ratio) then, strictly, they should not be able to do the shape task (poor sensitivity should be evident with no prediction for bias).

However due to the nature of our results the alternative view that a single representation of space supports behaviour in all tasks cannot be rejected outright. If such a representation did underpin performance and it was subject to a systematic distortion (based on, for example, a depth compression) then it may support good performance on our first two tasks but only relatively poor performance on the third. This alternative hypothesis seems unlikely, however, when other data are taken into account. For example, Glennerster et al. (1996) used a greater range of viewing distances, which included some that were considerably closer than those used here and found that at near distances the depth was 'stretched' in their shape task, whereas for far distances it was compressed. Despite this, the results for their depth matching task remained close to veridical at all viewing distances (see also Foley, 1980; Johnston, 1991; Tittle et al., 1995). A single representation cannot account for these results. Taken together with our results, these findings are inconsistent with the view that a single (e.g. Euclidean) representation of depth for every point in space is constructed by the visual system on the basis of the disparity and/or parallax information and used to support all tasks.

The simultaneous presence of both disparity and parallax, which in principle, has been shown to be sufficient to provide metrical information (Richards, 1985) without recourse to extra-retinal scaling information, did not appear to affect performance in any of the three tasks (see also Tittle et al., 1995; Brenner & van Damme, 1997; Brenner & Landy, 1999). This was also the case for the shape task, which might have been thought to show a particular benefit if the visual system was able to exploit this strategy. Performance in this condition did not differ significantly from that in the single-cue binocular condition in any of the three tasks.

One possible reason for the similarity in the results for different cue conditions is that only relatively small rotation angles were used in the present experiment (see Johnston, Cumming, & Landy, 1994). The algorithm Johnston et al. proposed for combining disparity and motion information relies on the detection of changes in disparity. For the particular configuration of LEDs used in the present experiment, and the small rotation angle created by a 13-cm translation at 1.5 or 3 m, the change in relative disparity is small, and much smaller than in the conditions reported by Johnston et al. A similar argument holds for the single-cue motion stimuli. Eagle and Blake (1995) and Hogervorst and Eagle (1998) examined the effect of human detection thresholds for higher order motion parameters (e.g. change in displacement divided by mean displacement, $\Delta d/dm$) and the effects of sensitivity to these parameters on SFM algorithms. The differences in the motion of elements between frames generated by a large range of structures is very small when the rotation angle is

small and so the potential effect of noise in the measurements is significant (Eagle & Blake, 1995). The pattern of biases that have been observed in shape judgements (e.g. LITER et al., 1994) have been explained in terms of an optimal decision rule, given the presence of noise in the measurements of stimulus motion (Hogervorst & Eagle, 1998). These biases are predicted to be most severe for relatively deep structures undergoing small rotations (Hogervorst & Eagle, 1998), which is typical of the stimuli used in the present experiments.

Bradshaw et al. (1998) did however find some evidence to suggest that the two cues were combined to support accurate size *and* depth judgements using similar stimuli and viewing conditions. The difference between their results and those reported here is not clear but may reflect the type of strategy adopted by observers in performing the respective tasks. Certainly, greater emphasis may have been placed on size and depth as observers had to set both dimensions in their study.

Real world stimuli were used in the present experiments because of recent suggestions that depth constancy may be underestimated when simulated stimuli are used (e.g. Durgin et al., 1995; Frisby et al., 1996) due to the artefacts associated with such stimuli. The estimates of depth constancy here, however, appear commensurate with studies that used simulated stimuli and the fairly large departures from veridical performance on the shape task also complements results from experiments using simulated stimuli (e.g. Todd & Bresnan, 1990; Norman & Todd, 1993; LITER et al., 1994; Tittle et al., 1995; Glennerster et al., 1996; Norman et al., 1996). The conditions in our experiments, however, were rather different from those encountered when viewing everyday scenes given they comprised several bright LEDs presented in complete darkness. These viewing conditions were designed to preclude the use of scene-based pictorial cues to recover viewing distance. This was important if we were to create the conditions in which possible interactions between disparity and parallax cues for the recovery of absolute depth could be established. Well lit, structured environments appear to support 100% constancy (e.g. Durgin et al., 1995) although what elements in such scenes determines performance remain to be identified.

In conclusion, the results of the current experiments provide evidence that the visual system uses different strategies dependent on the nature of the task being performed. In addition, systematic distortions of depth, similar to those reported for simulated stimuli, can occur for real stimuli under viewing situations when extraneous visual cues are removed. Finally, under the present viewing conditions and tasks there was little evidence of motion and disparity information being integrated for the accurate recovery of depth or shape information.

Acknowledgements

We would like to acknowledge the extensive discussions with Richard A. Eagle in the preparation of the experiments reported in this paper. The work was supported by the Wellcome Trust and the MRC Interdisciplinary Research Centre in Cognitive Neuroscience (Oxford).

References

- Bradshaw, M. F., & Rogers, B. J. (1996). The interaction of binocular disparity and motion parallax in the computation of depth. *Vision Research*, *36*, 3457–3468.
- Bradshaw, M. F., & Rogers, B. J. (1999). Sensitivity to stereoscopic corrugations as a function of corrugation frequency. *Vision Research*, *39*, 3049–3056.
- Bradshaw, M. F., Frisby, J. P., & Mayhew, J. E. W. (1987). The recovery of structure from motion: no evidence for a special link with the convergent disparity mechanism. *Perception*, *16*, 361–367.
- Bradshaw, M. F., Glennerster, A., & Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research*, *36*, 1255–1265.
- 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*, 1317–1333.
- Braunstein, M. L., LITER, J. C., & Tittle, J. S. (1993). Recovering three-dimensional shape from perspective translations and orthographic rotations. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 598–614.
- Brenner, E., & Landy, M. S. (1999). Interaction between the perceived shape of two objects. *Vision Research*, *39*, 3834–3848.
- Brenner, E., & van Damme, W. (1997). Interaction between perceived shape and distance. *Investigative Ophthalmology and Visual Science*, *38*, S1010.
- Collett, T. S., Schwartz, U., & Sobell, E. (1991). The Interaction of oculomotor cues and stimulus size in stereoscopic depth constancy. *Perception*, *20*, 733–754.
- Durgin, F. H., Proffitt, D. R., Olson, J. T., & Reinke, K. S. (1995). Comparing depth from binocular disparity to depth from motion. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 679–699.
- Eagle, R. A., & Blake, A. (1995). Two-dimensional constraints on three-dimensional structure from motion tasks. *Vision Research*, *35*, 2927–2941.
- Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, *87*, 411–434.
- Frisby, J. P., Buckley, D., & Duke, P. A. (1996). Evidence for the good recovery of lengths of real objects seen with natural stereo viewing. *Perception*, *25*, 129–154.
- Garding, J., Porrill, J., Mayhew, J. E. W., & Frisby, J. (1995). Binocular stereopsis, vertical disparities and relief transformations. *Vision Research*, *35*, 703–722.
- Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research*, *36*, 3441–3456.
- Graham, C. H., Baker, K. E., Hecht, M., & Lloyd, V. V. (1948). Factors influencing the thresholds for monocular movement parallax. *Journal of Experimental Psychology*, *38*, 205–223.
- Hogervorst, M. A., & Eagle, R. A. (1998). Biases in three-dimensional structure-from-motion arise from noise in the early visual system. *Proceedings of the Royal Society*, *265*, 1587–1593.

- Howard, H. J. (1919). A test for the judgement of distance. *American Journal of Psychology*, 2, 656–675.
- 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 stereo and motion cues. *Vision Research*, 34, 2259–2275.
- Julesz, B. (1971). *Foundations of cyclopean perception*. Chicago, IL: University of Chicago Press.
- Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, 8, 377–385.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modelling of depth cue combination: in defence of weak fusion. *Vision Research*, 35, 389–412.
- Liter, J. C., Braunstein, M. L., & Hoffman, D. D. (1994). Inferring structure from motion in two-view and multiview displays. *Perception*, 22, 1441–1465.
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organisation of three-dimensional shapes. *Proceedings of the Royal Society London B*, 200, 269–294.
- Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception and Psychophysics*, 53, 279–291.
- 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.
- Richards, W., & Lieberman, H. (1985). Correlation between stereo ability and the recovery of structure from motion. *American Journal of Optometry and Physiological Optics*, 62, 111–118.
- Richards, W. (1985). Structure from stereo and motion. *Journal of the Optical Society of America*, 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. (1995a). Disparity scaling and the perception of fronto-parallel surfaces. *Perception*, 24, 155–180.
- Rogers, B. J., & Bradshaw, M. F. (1995b). Are different estimates of 'd' used in disparity scaling. *Investigative Ophthalmology and Visual Science*, 36, 230.
- Rogers, B. J., & Graham, M. E. (1979). Motion parallax as an independent cue for depth perception. *Perception*, 8, 125–134.
- Rogers, B. J., & Graham, M. E. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research*, 22, 216–270.
- 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., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception and Psychophysics*, 48, 419–430.
- Todd, J. T., & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception and Psychophysics*, 50, 509–523.
- 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.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.