Perception, 2004, volume 33, pages 1155 – 1172

DOI:10.1068/p5253

Combining cues while avoiding perceptual conflicts

Maarten A Hogervorst¶
Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK; e-mail: hogervorst@tm.tno.nl

Eli Brenner
Department of Neuroscience, Erasmus MC, PO Box 1738, 3000 DR Rotterdam, The Netherlands; e-mail: e.brenner@erasmusmc.nl

Received 22 April 2003, in revised form 9 March 2004; published online 15 December 2004

Abstract. A common assumption in cue combination models is that small discrepancies between cues are due to the limited resolution of the individual cues. Whenever this assumption holds, information from the separate cues can best be combined to give a single, more accurate estimate of the property of interest. We examined whether information about the discrepancy itself is lost when this is done. In our experiments, subjects were required to combine cues to match certain properties while avoiding perceptual conflicts. In part 1, they combined expansion and change in disparity to estimate motion in depth; and in part 2, they combined perspective and binocular disparities to estimate slant. We compared the pattern in the way that subjects set the two cues with the patterns predicted by models of cue combination with and without a loss of information about the discrepancy. From this comparison we conclude that little information about the discrepancies between cues is lost when the cues are combined.

1 Introduction

The visual information that is available to judge a certain property can often be divided into different aspects, which are referred to as separate visual cues for that property. For instance, binocular disparity, linear perspective, and motion parallax can be considered to be cues for the perception of depth (eg Cutting and Vishton 1995). The division into separate aspects is based on differences in the underlying assumptions and in the required computations. It is therefore reasonable to assume that cues are initially processed largely independently.

When more than one cue contains information about a certain property, the information must somehow be combined. There are numerous suggestions for the way in which different cues are combined. The main difference between the various proposals lies in the kind of information that influences the combination. Kinds of information that have been considered include the reliability of the individual cues (Gillam 1968; Hogervorst and Eagle 1998; Landy et al 1995; Young et al 1993), the likelihood of the emerging interpretation (Bülthoff 1991; van Ee et al 2003; Geisler and Kersten 2002), recent experience (Atkins et al 2001; Jacobs and Fine 1999), and consistency between the cues (Brenner and van Damme 1999; Frisby et al 1995; Johnston et al 1993, 1994). In most proposals, the way in which the cues are combined is some form of weighted averaging, with a selection from the above-mentioned kinds of information determining the weights.

The way in which cues are combined is obviously only important if there are discrepancies between the values suggested by different cues. Under normal viewing conditions, such discrepancies can arise for two reasons: because of the limited accuracy of the estimated value from each cue, or because the assumption on which a cue is

¶Current address: Vision and Imaging, TNO Human Factors, PO Box 23, 3769 ZG Soesterberg, The Netherlands.

based is violated. If the discrepancy is large, then an assumption is likely to have been violated. But, if the discrepancy is small, it is reasonable to assume that it is caused by errors in estimating the individual values (Young et al 1993).

If a discrepancy is caused by inaccuracy, then a weighted average will provide an improved estimate of the value of the property of interest. The discrepancy itself is deemed to be uninteresting, so there is no point retaining information about it. If the discrepancy is ignored, then a somewhat smaller value obtained from one cue will compensate for a somewhat larger value obtained from another cue to yield a percept that cannot be distinguished from the original. The simplest case is when there are only two cues. We will use x and y to indicate the values that the two cues suggest for the property of interest (p). We use w to indicate the weight given to x (any value between zero and one). Since there are only two cues, the weight given to y will be $1 - w$. Thus:

$$p = wx + (1 - w)y \,. \tag{1}$$

According to equation (1) a given value of the percept (p) can arise from many combinations of values of the cues x and y. This is illustrated for a weight, w, of 0.5 by the diagonal line in figure 1a. For different weights the slope of the line will be different.

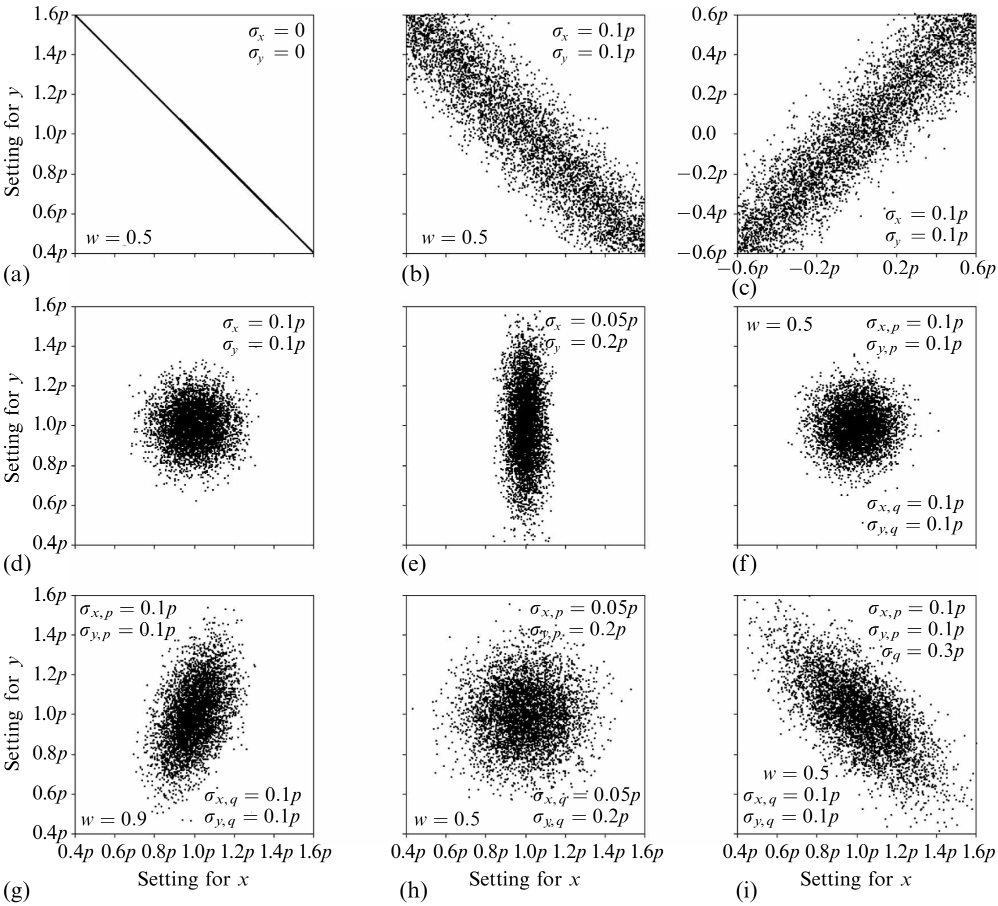


Figure 1. Simulations of settings for cues x and y ($n = 5000$) following equations (1) [panel (a)], (2) [panel (b)], (3) [panel (c)], (4) and (5) [panels (d) and (e)], and (7) and (8) [panels (f), (g), (h), and (i)]. w is the weight given to cue x; $\sigma_{x,p}$ is the standard deviation of the errors in estimating cue x for judgment p (expressed as a fraction of the magnitude of p).

Of course, we do not really expect the points to lie on a straight line (as in figure 1a), because the whole reason for averaging is that there are errors in the estimates of the individual cues. Thus a more realistic version of equation (1) is

$$p = w(x + \delta x_{p,i}) + (1 - w)(y + \delta y_{p,i}), \tag{2}$$

where $\delta x_{p,i}$ denotes the error in judging the value of p from cue x on occasion i, and $\delta y_{p,i}$ denotes the error in judging the value of p from cue y on occasion i. Figure 1b shows the outcome of a simulation of 5000 matches to a fixed reference value p, on the assumption that the errors in x and y are both normally distributed with a standard deviation of 10% of the value of p (and no bias), and that the two cues are given equal weights ($w = 0.5$).

Of course, cues are not always given equal weights. A common way to determine the weights given to cues is by introducing conflicts between the values specified by two cues, and asking people to make a single judgment on the basis of the combined stimulus (eg Brenner et al 1996; Bülthoff and Mallot 1988; Cornilleau-Pérès and Droulez 1993; Norman and Todd 1995; Regan and Beverley 1979; Rogers and Collett 1989; Turner et al 1997; Young et al 1993). The kind of data that one obtains is often similar to the data shown in figure 1b, but not necessarily with a slope of -1. The weights assigned to the cues can be obtained by fitting a line to the data: $w = slope/(slope - 1)$.

Obtaining data like those shown in figure 1b does not necessarily mean that information about the discrepancy is lost. People are expected to make a single judgment irrespective of whether they detect the discrepancy or not. The present study was designed to determine whether people lose all information about the discrepancy when the discrepancy is small (and is therefore likely to be caused by inaccuracy rather than by the violation of an assumption that is used to determine one of the values).

If we ask people to set two cues (x and y) to match a certain judgment (p) for which both cues provide an estimate, they could set x and y as in figure 1b. However, we can also specifically ask them to avoid discrepancies between the cues. This can be done indirectly. If one knows how a discrepancy will be interpreted perceptually (see the experiments for examples), one can ask subjects to avoid that percept. Now, consider the case that the second percept (q) is based on the discrepancy between the same estimates of the same cues:

$$q = (x + \delta x_{p,i}) - (y + \delta y_{p,i}) . \tag{3}$$

This relationship is illustrated in figure 1c, where q has a value of zero (no perceived discrepancy between the cues) and $\delta x_{p,i}$ and $\delta y_{p,i}$ are the same values as they were in figure 1b [and equation (2)]. Asking subjects to find a setting that matches a given value of p, while also ensuring that the percept that arises from the discrepancy between the cues is avoided ($q = 0$), will result in settings that fit both equations (2) and (3). In that case, the equations can be combined:

$$x = p + (1 - w)q - \delta x_{p,i} , \tag{4}$$

$$y = p - wq - \delta y_{p,i} . \tag{5}$$

Thus, if q is judged on the basis of the discrepancy between the two estimates of x and y that were averaged to judge p, then the variability in the settings of cue x will only depend on the accuracy of judgments of x ($\delta x_{p,i}$) and the variability in the settings of cue y will only depend on the accuracy of judgments of y ($\delta y_{p,i}$). This is illustrated in figures 1d and 1e for normally distributed errors in x and y. In figure 1d both have a standard deviation of 10% of the value of p, while in figure 1e $\delta x_{p,i}$ has a standard deviation of 5% of p, and $\delta y_{p,i}$ has a standard deviation of 20% of p.

Note that the variability in the settings of x and y directly follows the magnitudes of the errors in these cues.

Equations (4) and (5) are valid only if the errors in x and y are identical for the judgments of p and q, as they will be if the discrepancy itself is used to judge the value of q. If q is judged independently on the basis of similar (but not identical) information, then the error term will be different from that in equation (2). Thus, instead of equation (3), we get

$$q = (x + \delta x_{q,i}) - (y + \delta y_{q,i}) \,, \tag{6}$$

which is equivalent to equation (3), but the errors are no longer identical to those for judgment p, and are therefore indicated by different subscripts. When equations (2) and (6) are combined, we get

$$x = p - w\delta x_{p,i} - (1-w)\delta y_{p,i} + (1-w)(q - \delta x_{q,i} + \delta y_{q,i}) \,, \tag{7}$$

$$y = p - w\delta x_{p,i} - (1-w)\delta y_{p,i} - w(q - \delta x_{q,i} + \delta y_{q,i}) \,. \tag{8}$$

If the judgments of p and q are independent, we therefore expect the variability in x to no longer depend only on $\delta x_{p,i}$ and that in y to no longer depend only on $\delta y_{p,i}$. Figure 1f shows the predicted settings for $w = 0.5$ and a standard deviation of 10% of p for all four independent (normally distributed) errors. In this case there is no correlation between the settings. In fact, the pattern is very similar to that in figure 1d. However, if w is large (figure 1g; $w = 0.9$) or small, or if $\delta x_{p,i}$ and $\delta y_{p,i}$ are much larger than $\delta x_{q,i}$ and $\delta y_{q,i}$ (not shown), then there will be a positive correlation between the set values of x and y. If the errors in determining q are much larger than those in determining p, then there will be a negative correlation (also not shown, but an equivalent situation is shown in figure 1i, as will be explained in the next paragraph). Finally, if the errors in y are larger than the errors in x (figure 1h), this will not directly be evident in the settings (compare this figure with figure 1e).

It is not critical that judgment q is determined on the basis of the difference between x and y (set to zero), rather than on their ratio (set to one). It is also irrelevant whether the errors in q are related to judging the same measures, x and y, or are completely independent. The latter can be understood by considering what would happen if we introduced an additional term for variability in q in equations (7) and (8). Of course, if we add a lot of variability in q, we will get a negative correlation. This is illustrated in figure 1i [where $w = 0.5$ and an additional normally distributed error term δq_i with a standard deviation of 30% of p has been added to q in equations (7) and (8)].

It is evident from comparing equations (4) and (5) with equations (7) and (8) (and from figure 1i) that a large variability in the judgment of q (or p) that is not related to the cues (x and y) would mask all the differences that we just described. It is therefore critical that such additional sources of error (such as fluctuations in memorised references) are avoided. One way to achieve this is by asking people to match two parts of a presentation with no time limit for doing so. We use this method to investigate the way in which cues are combined for two different spatial judgments. In both cases, we combine binocular cues, which are more likely to be inaccurate than to be based on the wrong assumptions, with monocular cues for which the underlying assumption is known (and could be violated). We will look for correlations between the settings for the two cues (as an indication that the errors are independent) and evaluate the pattern of errors in relation to the above-mentioned equations.

The first spatial judgment that we will examine is that of the speed of motion in depth provided by a combination of a change in retinal image size (looming) and a change in binocular disparity. We know that, when changing image size and changing binocular disparity are in conflict, people perceive an intermediate velocity of motion in depth (Brenner et al 1996; Regan and Beverley 1979). We also know that the assumption underlying the use of changing image size to judge motion in depth is that the (3-D) size of the object does not change. Any conflict that is detected should therefore result in a perceived change in object size. The question is whether apparent changes in size arise directly from detecting the cue conflict, or whether they are the result of independently judging the size on the basis of related information, such as the instantaneous binocular disparity and retinal image size.

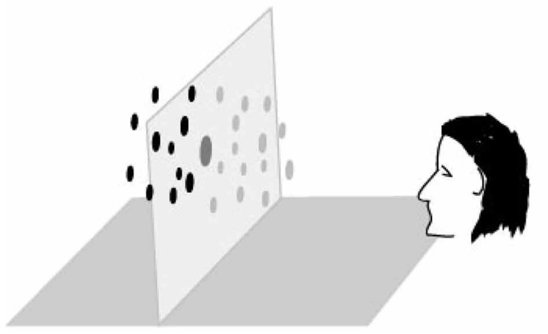
The second spatial judgment that we will investigate is that of surface slant from a combination of binocular disparity and perspective (eg Allison and Howard 2000; Buckley and Frisby 1993; Gillam 1968). Various aspects of perspective are based on slightly different information (eg texture being isotropic, lines being parallel). We limit ourselves to one aspect of perspective: foreshortening. The cue that we will use is the fact that the retinal image of a circle is elliptical if the circle is slanted relative to the line of sight. Any conflict that is detected between the slant derived from the elliptical shape and that derived from the binocular disparities will be interpreted as a deviation in the shape of the object. The object will look elliptical rather than like a slanted circle. We will therefore instruct subjects to match a reference slant while at the same time making sure that the target looks circular.

2 Part 1: Motion in depth from changing retinal size and disparity

A 3-D (stereoscopic) scene was simulated in which a central disk moved straight towards the observer. The change in binocular disparity and the change in angular size of the moving disk were manipulated independently. There were three kinds of experiments. One in which the subject manipulated both cues, and two in which the subjects manipulated only one of the two cues. In the free-cues experiment the subject's task was to set both cues to create a scene in which the object moved at a constant speed in depth without changing its size. In the fixed-random-cue experiments one of the cues was fixed at some random value and the subject set the other cue. Subjects either had to create a scene in which the object moved at a constant speed, or one in which the object did not change in size. The 'constant-speed' task is expected to involve weighted averaging, because there are two cues for the same attribute (motion in depth). The 'constant-size' task could be performed by eliminating the discrepancy between the motion cues, or by independently judging the size of the target on the basis of the (retinal) image size and binocular information. Finally, in the fixed-veridical-cue experiments the subject also set only one of the two cues, but the other cue was fixed to the correct value. The task was to create a scene in which the object moved at a constant speed in depth without changing its size (as in the free-cues experiment).

2.1 *Methods*

2.1.1 *Apparatus.* Images were presented on a computer screen by a Silicon Graphics Onyx Reality Engine (120 Hz; horizontal size: 39.2 cm, 815 pixels; vertical size: 29.3 cm, 611 pixels; spatial resolution refined with anti-aliasing techniques). Subjects sat with their head in a chin-rest at 80 cm from the screen. The images were viewed through liquid-crystal shutter spectacles that were synchronised with the refresh rate of the monitor. Alternate images were presented to the left and right eye, so that each eye received a new image every 16.7 ms (60 Hz). These images were calculated with the individual subject's interocular distance taken into account. Only the red gun of the monitor was used because the shutter spectacles work best for red images (transmitting more than 50 times as much light when open than when shut).



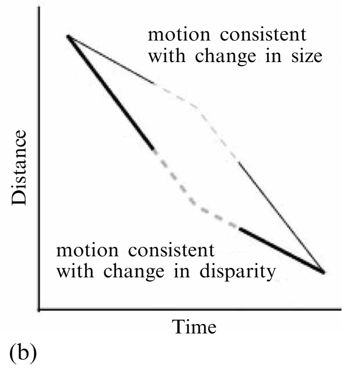


Figure 2. (a) Schematic representation of the simulated scene. The large grey circle is the target and the small ones are the static disks. The vertical plane represents the screen. (b) Subjects split the total change in binocular disparity and/or angular extent between two 750 ms intervals of motion at a constant velocity. The task was to find a combination of velocities for the two cues for which the target appeared to approach at the same speed in both intervals. The intervals were separated by a 500 ms period during which the target was invisible (dashed parts of the lines).

2.1.2 *Stimuli.* The 3-D scene that was simulated consisted of a central moving target disk surrounded by 25 static disks. Figure 2a gives an impression of this simulated scene. The target was at eye height and moved straight toward the midpoint between the subject's eyes. Static disks surrounded the target's path in order to ensure a good percept of motion in depth from changing binocular disparity (Regan et al 1986). All static disks had a diameter of 2 cm and were oriented parallel to the computer screen with their centres 9 cm from the central target's path (in the simulated space, not on the screen). The static disks formed a radial pattern surrounding the target, with one disk every 14.4°. The simulated distance of each static disk along the target's path was selected at random from between the initial and final simulated distance of the central disk on that trial (ie covering the same range as the motion of the central disk).

Two different target sizes (1 cm and 4 cm diameter) were used, because a larger target size was expected to result in more accurate judgments of motion from changing size, and therefore in more weight being given to changing size relative to that given to changing disparity. The total change in image size and change in disparity of the target disk was fixed for each trial. What subjects could change was the extent to which the change took place during the first or second part of the movement. Within each part, the change in both size and disparity was consistent with a constant-velocity approach, but this velocity was not necessarily the same for the two cues or for the two parts, as explained below.

On each presentation, the target disk appeared at a certain simulated distance (16, 20, or 24 cm) behind the screen, and moved towards the subject so that after 1.5 s it was at the same simulated distance in front of the screen. Thus there were three average simulated velocities: 21.33, 26.67, and 32 cm s^{-1}. However, the target disk did not necessarily move at this velocity throughout the presentation. It moved at a constant velocity that was either fixed to a certain value or set by the subject for the first half of the presentation (the first 750 ms). It then moved at the (constant) velocity that would get it to the final position exactly in time during the second half of the presentation. This applied both to the velocity specified by the change in image size and to the velocity specified by the change in disparity, but the distribution of the motion between the first and second halves of the presentation could differ for the two cues (see figure 2b). Thus, in all the tasks subjects changed the amount of motion that took place during the first half of the presentation, either for one cue only or for both of the cues independently. The motion in the second half was automatically adjusted so that the total change in image size and in binocular disparity was not affected by the subject's setting.

In order to make sure that subjects were comparing velocities rather than detecting changes in velocity, we made the target disk disappear around the time that it changed velocity (although a pilot study showed that this does not make a difference). Thus, subjects only saw the moving target for 500 ms of the 750 ms of motion at the first velocity. After that the target was invisible for 500 ms before reappearing to move on for the last 500 ms at the second velocity. Note that the target disk continued to 'move' when it was not visible. The surrounding disks remained visible when the target disk disappeared, as well as during the 500 ms between presentations. The full range of values was available for each setting. Thus, subjects could make all of the change take place during the first interval, all of the change take place during the last interval, or anything in between these extremes. The initial setting was a random value from within this range.

2.1.3 *Subjects.* Three subjects participated in the experiments. One was one of the authors. The other two were not aware of the objectives of the study. All have normal (corrected) vision.

2.1.4 *Procedure.* The subject was seated in a dark room at a distance of 80 cm from the computer screen with his/her head on a chin-rest, wearing the shutter spectacles. Each stimulus sequence was repeated over and over again until the subject was satisfied with the setting (as indicated by a mouse-button click). There was no time limit. When subjects had to adjust both cues, they adjusted the speed defined by binocular disparities by moving the computer mouse 'vertically' and the speed defined by the change in size by moving the computer mouse 'horizontally'. When they were only adjusting one cue, they adjusted the cue by moving the mouse horizontally (for speed judgments) or vertically (for size judgments), irrespective of which cue was being varied on that trial.

In the free-cues experiment, the subject adjusted the speeds indicated by both cues so that the disk appeared to approach at a constant speed without changing in size. In the fixed-random-cue experiments, the subject either adjusted the change in disparity or the change in size. In one type of session, the subject was asked to make the speed appear to be constant, regardless of whether a change in size was perceived. In another type of session, the subject was asked to make the size appear to remain constant, regardless of whether a change in speed was perceived. Whenever the subject judged the task to be impossible, he/she could indicate this and proceed with the next trial without finishing the setting. This occurred only in the fixed-random-cue experiment. In that experiment about 10% of the trials in which the speed had to be set to appear constant and about 1% of the trials in which the size had to be set to appear to remain constant were considered to be impossible. In the fixed-veridical-cue experiment (in which the speed indicated by one cue was fixed to the correct value) the task was always to set the speed to appear constant.

In each session, each of the three average velocities was shown 20 times in random order. In the free-cues experiment each subject participated in two such sessions. In the fixed-random-cue experiments each subject participated in four sessions. In two of these, the subject was asked to set the speed to remain constant, and in the other two to set the size to remain constant. The cue that was to be adjusted (change in disparity or change in size) was determined at random for each trial (note that the subjects did not need to know which cue they were manipulating to perform the task). In the fixed-veridical-cue experiments, each subject participated in two sessions. Again, the cue that had to be adjusted was determined at random for each trial. This whole sequence of 8 sessions was run twice, once for a small target (1 cm diameter) and once for a large target (4 cm diameter). Each subject therefore participated in a total of 16 sessions.

2.1.5 *Analysis.* The main experiments are those in which both changing size and changing disparity had to be set. In order to make sure that subjects compare the two intervals within each trial, rather than matching the setting on successive trials, we used three different average velocities. Consequently, we cannot simply plot velocities, because the settings of the two cues for a higher average velocity are obviously both higher, which would give a large positive correlation. In order not to have to plot data for each average velocity separately, we expressed all the settings as percentages of the average set velocity (an attractive feature of doing so is that, after scaling, the standard deviations in the settings were independent of the average velocity). Several examples are shown as percentages of the correct velocity (figures 3a – 3c), but we reasoned that using percentages of the subject's average set velocity for that cue and average velocity (on that session) is more appropriate, because doing so eliminates any possibility that systematic errors influence our analysis. Removing systematic error in this manner also allows us to determine correlation coefficients across subjects, sessions, and average velocities.

The fixed-veridical-cue experiments were included in order to determine the accuracy with which subjects can set a single cue. The fixed-random-cue experiments were included in order to determine the weights given to the two cues when visibly in conflict. We did so by determining the slope of a line fit to the value set by the subject (dependent variable) as a function of the fixed random value (independent variable) for each cue that was set within each session by each subject. For each task (set constant velocity

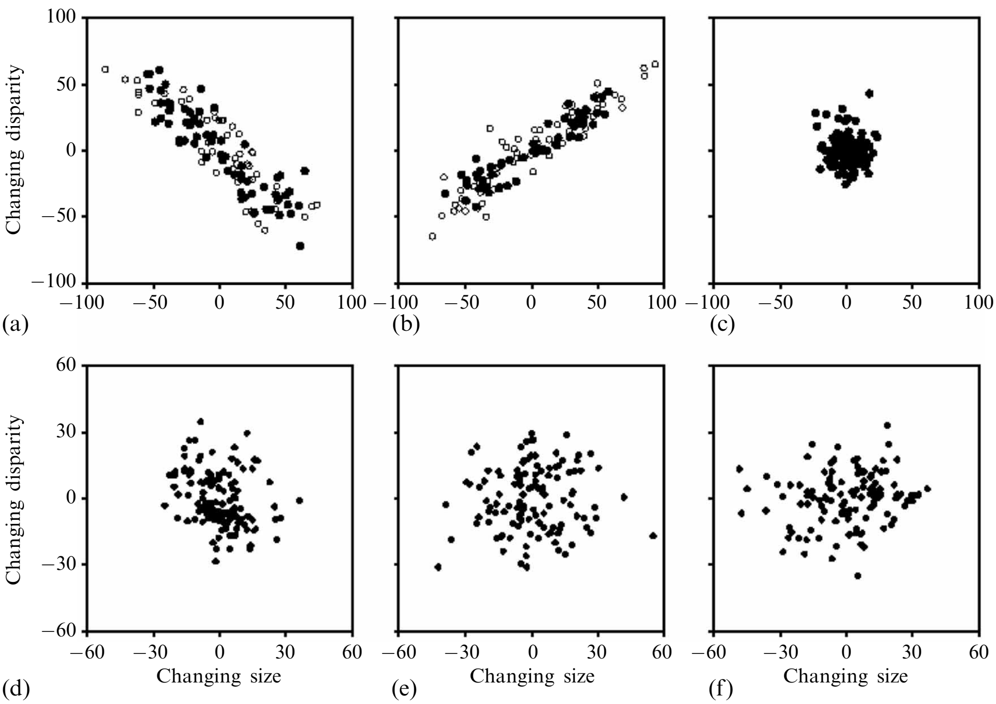


Figure 3. Panels (a) – (c): One naïve subject's performance for the large target in three of the five tasks. (a) Setting a constant speed with one cue fixed to a random value. Solid symbols are for settings of changing disparity while changing size is fixed. Open symbols are for settings of changing size with changing disparity fixed. (b) Setting a constant size with one cue fixed to a random value. (c) Setting both cues so that both speed and size remain constant. The values are percentage deviations from the correct value, ranging from −100% (target static in first interval) to 100% (all motion in first interval). Panels (d) – (f): The three subjects' performance when setting both cues for the small target. The values are percentage deviations from the average value that the subject set under these conditions (note that the scale is reduced to ±60% for better visibility).

or constant size) and target size (small or large), the slopes were then averaged (after transforming half of them so that all slopes describe the amount of changing disparity per unit of changing size, in accordance with the way we plot our data).

Cue averaging for speed judgments is expected to give a negative slope, in accordance with weighted averaging. The precise value of the slope will depend on the weights given to the two cues. Thus the two target sizes are expected to give different slopes in the speed judgment task, because increasing the size of the target will increase the reliability of the changing-size cue so that the weight given to that cue will increase (Young et al 1993). Combining the cues to obtain a constant target size is expected to lead to a slope of 1, because the change in image size must approximately match the change in binocularly defined distance for the target not to change its apparent size.

2.2 *Results and discussion*

Figures 3a – 3c show the results of one naïve subject for the large target in two sets of experiments. The settings are shown as deviations from the correct values. Figure 3a shows the data for the fixed-random-cue experiments in which the subject was required to set a constant speed. The solid symbols are for settings in which the change in disparity was adjusted and the change in size was fixed (to a random value). The open symbols are for settings in which the change in size was adjusted and the change in disparity was fixed. The settings are consistent with equation (2). Figure 3b shows the data for the fixed-random-cue experiments in which the subject was required to set a constant size. The settings are consistent with equation (3) or (6). The slope appears to be slightly smaller than 1, which could be due to systematic errors in judging binocularly defined distance (Brenner and van Damme 1999). Figure 3c shows the settings for the free-cues experiments. There is slightly more variability in changing disparity than in changing size, and there is no evident correlation between the cues. For both cues the average value is close to veridical.

Figures 3d, 3e, and 3f show the three subjects' settings for the small target in the free-cues experiments. The settings are shown as deviations from the average values set by the subject on that session for that average target velocity. The differences in accuracy between the subjects are clearly quite small.

Figure 4 shows a summary of all three subjects' data in all the experiments. The left panel shows the data for the large target, and the right panel shows the data for the small target. The data points are for the free-cues experiments. They are the percentage deviations from the average value set for that stimulus on that session. The two diagonal lines are the average slopes for the two tasks in the fixed-random-cue experiments (set constant velocity and set constant size). The shaded histograms at the ends of these diagonals show the deviations from the diagonal lines in the fixed-random-cue experiments. The shaded histograms along the vertical and horizontal directions show the deviations from the average settings when one of the cues was fixed to the right value (fixed-veridical-cue experiments). The outline histograms show the corresponding deviations in the free-cues experiments (corresponding to the shown data points).

2.2.1 *Correlations.* When all subjects' data were taken together, there was a very modest positive correlation between the settings for the large target (0.11; $p = 0.03$), and no correlation for the small target (0.00; $p = 0.99$; see figure 4). When analysed separately, one subject had a significant negative correlation (-0.30; $p < 0.01$) for the small target (see figure 3d), and another had a significant positive correlation (0.33; $p < 0.01$) for the large target. The other four correlations were not significant. Thus the errors in the settings for the two cues are not always completely independent, but there is no consistency across subjects. The lack of a clear correlation between the settings does not prove that no information about the discrepancy is lost (see section 1), but it is consistent with that idea.

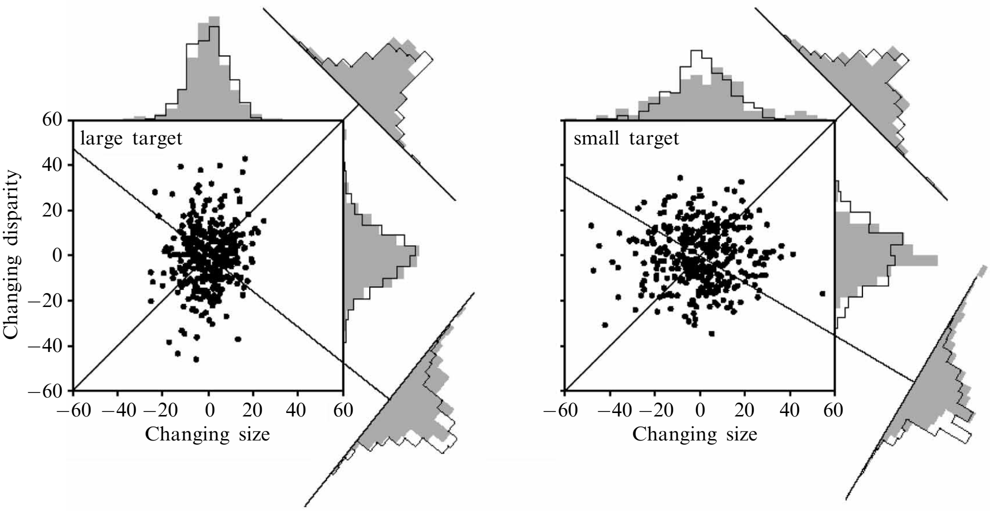


Figure 4. Summary of the three subjects' settings in all the experiments. The axes show the deviations from the mean value set by that subject for that particular total displacement ($n = 3$) during the first interval (as percentages of the mean value). The points show the data when the subjects set both cues independently. The outlined histograms show the distribution of these points in various directions. The shaded histogram along the horizontal axis shows the distribution of settings for changing size when the changing disparity was fixed at the correct value. The shaded histogram along the vertical axis shows the distribution of settings for changing disparity when the changing size was fixed at the correct value. The slopes of the diagonal lines with negative slope show the average slopes of lines fit to the settings that were made when one cue was fixed at a random value and subjects had to set the other to achieve a constant perceived velocity. The shaded histograms at the end of these axes show the distribution of deviations from these lines in this condition. The diagonal lines with positive slope show how we expect the points to cluster when one cue is fixed at a random value and subjects have to set the other to achieve a constant perceived size. Again, the shaded histograms at the ends show the deviations from these lines.

2.2.2 *Weights and slopes.* The average slope of the fit lines for the fixed-random-cue experiments in which the task was to set the size to remain constant was not significantly different from 1. For the large target the average slope was 1.07 ± 0.35 (standard deviation). For the small target the average slope was 0.89 ± 0.16. We had expected these slopes to be 1 because the cues have to be equated rather than averaged for this setting, so there are no weights involved [see equation (3)].

The average slope of the fit lines for the fixed-random-cue experiments in which the task was to set the speed to remain constant was -0.78 ± 0.16 for the large target and -0.54 ± 0.07 for the small target. These slopes are indicated by the descending lines in figure 4. The slopes correspond with weights for the changing-size cue (w) of 0.44 and 0.35 for the large and small target, respectively [$w = slope/(slope - 1)$, as explained in section 1]. As expected, the slope is less steep for the smaller target, because less weight is given to the (less reliable) estimate of velocity from changing size when the target is smaller, so a larger change in size is needed to compensate for a given change in disparity.

2.2.3 *Variability.* Comparing the shaded and outline (horizontal and vertical) histograms shows that the variability when two cues have to be set (free-cues experiments) does not differ from that when only one cue has to be set and the other is veridical (fixed-veridical-cue experiments). Thus having to set two cues did not make subjects' settings less reliable. However, comparing the widths of the histograms for the two cues suggests that the weights given to the cues are not optimal. The weights that we derived from the speed judgments in the fixed-random-cue experiments (0.44 and 0.35; see previous section) indicate that changing size is always given less weight than changing disparity. However, the horizontal and vertical histograms in figure 3 show that, for the large target, changing size is in fact set more reliably (narrower distribution), so it should have been given more weight ($w > 0.5$; slope < -1). Thus, the cues do not appear to be combined in a statistically optimal manner. Perhaps changing size is given less weight than one would predict from the variability, because changing (angular) size is only a reliable cue for motion in depth if the physical size of the object does not change. Thus, perhaps the likelihood that the assumptions for using the cues have been violated is also considered.

If a discrepancy between the cues is interpreted as a violation of the assumptions underlying the use of one of the cues, we could expect the weight given to that cue to decrease when the cue conflict becomes more conspicuous. This does not appear to happen. In the free-cues experiments our subjects set the target to appear to be a rigid object approaching at a constant speed. In the fixed-random-cue experiments the unavoidable cue conflict often resulted in very clear changes in apparent size. Nevertheless, when we look at our data in terms of deviations from a weighted average, there is no difference between the two experiments. This can be seen by comparing the histograms at the end of the descending lines in figure 4. The deviations from the fitted lines in the fixed-random-cue experiments (shaded histograms) are similar to the deviations in the same direction in the free-cues experiments (corresponding outline histograms; note that in the former case the points are spread along the whole line as in figure 3a).

If the weights had been adjusted on the basis of the magnitude of the conflict in the fixed-random-cue experiments, then the relationship between the two cues would not be a straight line, so the deviations from the fitted lines would have been larger in this condition. Thus, the weight given to the changing-size cue does not appear to decrease as the cue conflict becomes more conspicuous. Similarly, if subjects had switched between two (or more) percepts when the cues were in conflict (van Ee et al 2002, 2003), the deviations would have been larger in the fixed-random-cue experiments. Probably such switches did not occur in our experiment because we did not have as extreme conflicts. However, it is possible that the percept did switch on occasional trials, but that our subjects responded to this by indicating that they could not make a satisfactory setting on that trial. It is rather surprising that, even when the violation of the assumption underlying the changing-size cue is clearly detected (ie when the object clearly appears to change size), changing size is still considered to be a reliable source of information about motion. We will return to this issue in more detail in section 4. Here we will just mention that it suggests that conflicts between the cues are not detected explicitly, but that perceived size and motion are derived from the same input and therefore have related errors.

This brings us back to the equations that we introduced in section 1. Knowing the weights given to the cues and the variability in the settings, we can now try to model our data. For the case in which the same errors contribute to both judgments [no information about the discrepancy lost; equations (4) and (5)] this is very straightforward. We can take the weights determined from the fixed-random-cue experiments and the errors for the individual cues from the fixed-veridical-cue experiments and use these values to run simulations (like those shown in figures 1d and 1e) for the free-cues experiments. Figures 5a and 5b show the simulations for a large and small target, respectively. The distributions are similar to the data shown in figure 4.

For the case in which the two judgments have independent errors, such simulations are less straightforward, because we have four error terms to estimate. Moreover, a closer examination of equations (7) and (8) reveals that we have no hope of obtaining a good fit for the large-target data, because the variability in x (changing size) can only be

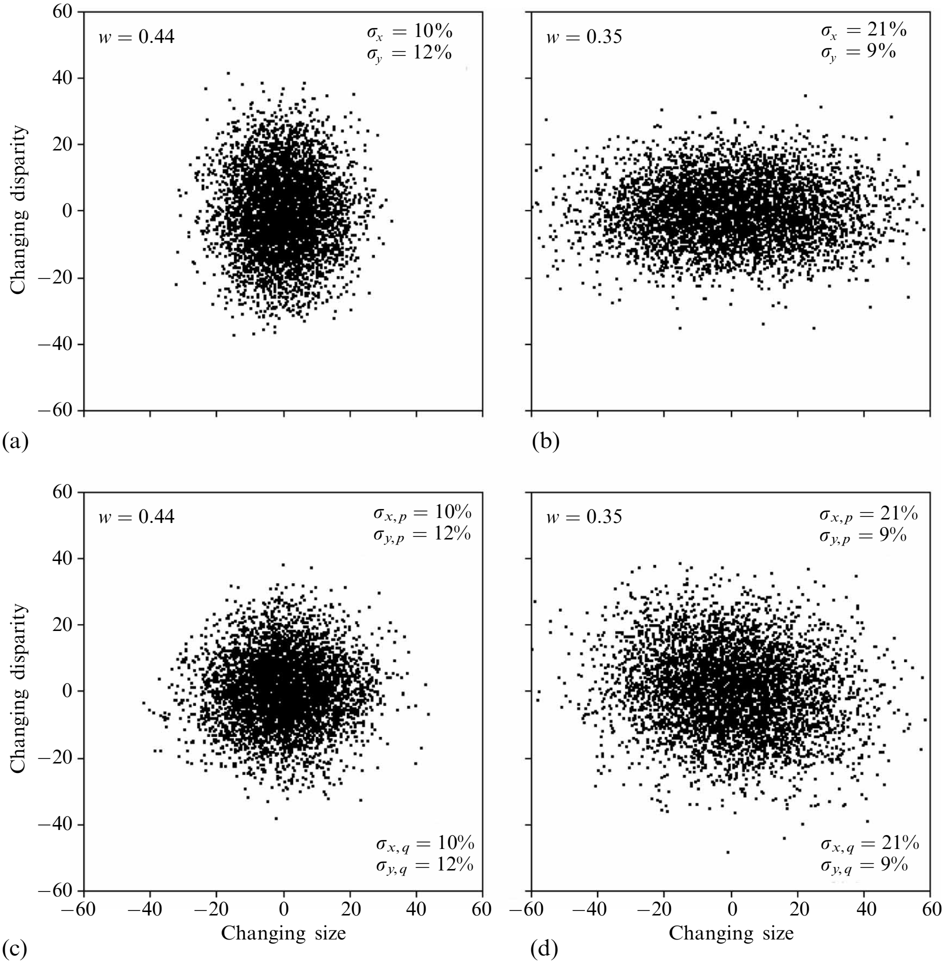


Figure 5. Predictions for the settings when both cues were set by the subject. Panels (a) and (b) show predictions based on equations (4) and (5), where size is judged from the discrepancy between motion cues. Panels (c) and (d) show predictions based on equations (7) and (8), where size is judged independently of motion. The value of w is derived from the fixed-random-cue experiments (as described in section 1). The standard deviations in judging x and y are taken from the fixed-veridical-cue experiments. Panels (a) and (c) are for the large target. Panels (b) and (d) are for the small target.

smaller than that in y (changing disparity) if the weight given to cue x is larger than 0.5 ($1 - w$ must be smaller than w in the last term of the two equations).

Figures 5c and 5d show the outcomes of simulations with the same values for the errors as in figures 5a and 5b, and with the assumption that the errors in judging each cue have the same magnitude for the two judgments. As was to be expected on the basis of the reasoning above, these distributions are less similar to those of the real data (figure 4) than those in figures 5a and 5b. We could have avoided the negative correlation in figure 5d by assuming that the errors in judging the cues are smaller for size judgments (q in the equations) than for motion judgments (p). However, if so, it would be even more evident that changing the accuracy of either cue (eg by changing target size) influences the settings of both cues in the same way [because only variability related to q can influence x differently than y; see equations (7) and (8)]. Thus, the pattern of results that we found is clearly less easily explained by independent errors for both judgments [equations (7) and (8)].

3 Part 2: Surface slant from binocular disparity and from foreshortening

In our second experiment, the task was to align two slanted surfaces while making them both appear circular. The upper surface was farther away than the lower one (see figure 5). The cues that we used were binocular disparity and perspective (foreshortening). These cues are known to interact to produce the perceived slant (eg Allison and Howard 2000; Buckley and Frisby 1993; Gillam 1968). Again, we compared the settings for each surface by a single subject on repeated trials under identical conditions (the same pair of distances). And, again, settings for each cue (and now also each surface) were expressed as deviations from the average setting in order to be able to average across subjects and pairs of distances. Moreover, doing so makes our analysis insensitive to systematic errors that could arise from inappropriate values of cues that we did not control, such as accommodation that specifies a fixed distance.

We also compared the settings for the two surfaces within each trial. We reasoned that, if all the estimates are independent, then the variability in the difference between the settings for the two targets (for each cue) will be about 40% larger than that for the individual settings. In contrast, if variability in estimating the reference contributes substantially to the errors, then the variability in the difference between the settings can even be smaller than that for the settings themselves.

3.1 *Methods*

3.1.1 *Apparatus.* For two of the subjects the stimuli were generated with a Silicon Graphics O2 computer on a 29.8×23.9 cm monitor (1280 by 1024 pixels; 75 Hz). Standard (4 bits) subpixel interpolation techniques were used to increase the spatial resolution. For the third subject the stimuli were generated with the equipment used for the motion stimuli, but at the pixel resolution and frame rate that are mentioned here. Stereoscopic images were created by displaying the view for the right eye in green and the view for the left eye in red, and viewing the images with appropriately coloured filters in front of the eyes. To avoid visible crosstalk between the images, the surfaces were presented on a yellow background. The eye separations of the individual subjects were taken into account when computing the images.

3.1.2 *Stimuli.* Each stimulus was a computer-generated simulation of a scene consisting of two vertically separated ellipses (as shown in figure 6a). The centre of the upper ellipse was further away than that of the lower ellipse. In order to align the two ellipses, the subjects had to match the slant of each ellipse to the slant of an imaginary plane through the two distances. Five slants of the imaginary plane were used: 35°, 40°, 45°, 50°, and 55°. Each ellipse could be rotated around a frontal horizontal axis through its centre.

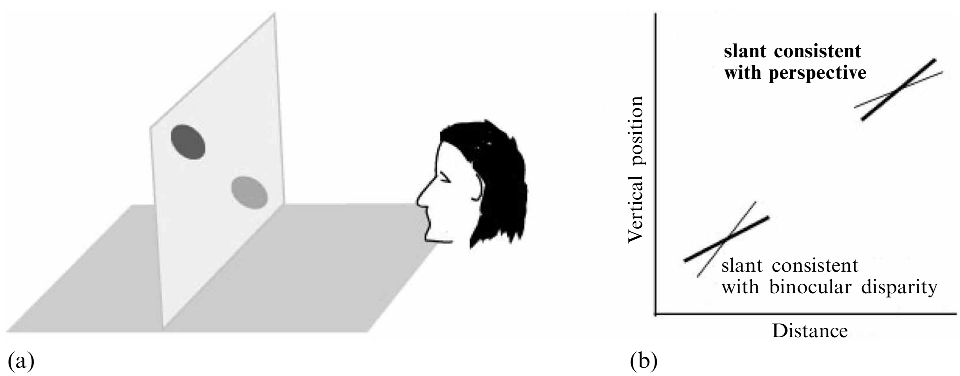


Figure 6. (a) Schematic representation of the simulated scene. (b) Subjects adjusted both the stereoscopic slant and the height-to-width ratio (perspective slant) of each surface until the two surfaces appeared to be aligned.

The width of the ellipses was 2.8 cm in the simulated scene. The centres of the ellipses were separated vertically by 5 deg.

3.1.3 *Procedure.* The subject was seated in a darkened room at a distance of 35 cm from the screen, with his or her head fixed on a head-rest. The subject could use the mouse buttons to vary the stereoscopic slant and the vertical elongation of each of the two ellipses (figure 6b). The task was to align the two ellipses while making each appear to be circular. The stereoscopic slant is defined on the basis of the binocular disparities. The slant defined by perspective is defined by the angular height-to-width ratio of the ellipse. At the start of each trial the stereoscopic and perspective defined slants were chosen at random from between 30° and 60°. The subject adjusted the surfaces until she or he was satisfied (there was no time limit). Within each session each of the five slants of the imaginary plane (35°, 40°, 45°, 50°, and 55°) was presented 10 times, in random order. Each subject participated in three sessions.

3.1.4 *Subjects.* Three subjects took part in this experiment: the two authors and one naïve subject. All had normal (corrected) vision.

3.1.5 *Analysis.* The settings were analysed in two ways. The first was similar to the analysis of the free-cues velocity experiments. For each trial, ellipse, and cue we calculated the deviation from the average setting made by that subject, for that ellipse and cue, during that session, for that reference slant. We plotted these deviations and calculated the correlation between the deviations for the two cues. The second way of analysing the settings started by calculating the difference between the slants specified by a given cue for the two ellipses within a trial. This was done for each cue, and the differences for one cue were compared with those for the other cue.

3.2 *Results and discussion*

Figure 7a shows the difference (in degrees) between the individual and the average slant settings. The figure includes the data for both ellipses, all five simulated slants, and all three subjects. Taking all the data together in this manner gives a significant

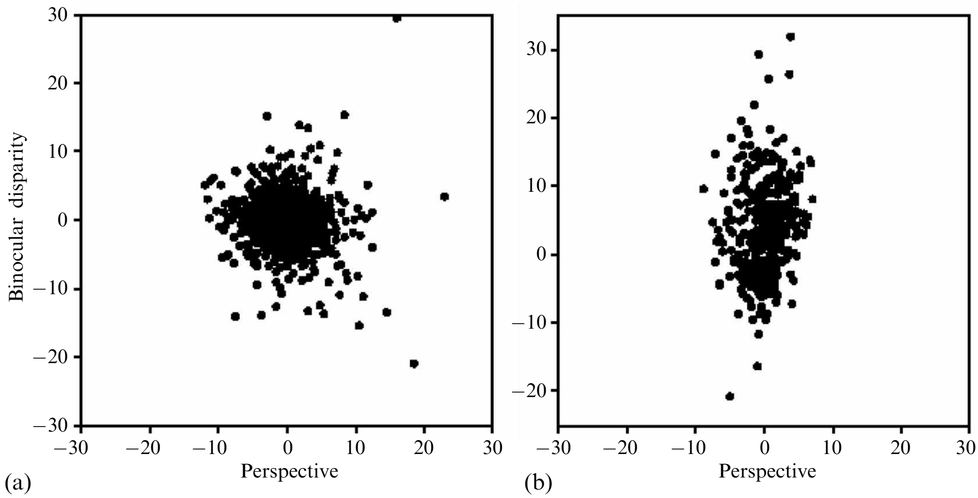


Figure 7. Summary of the settings in the slant experiment. The values are differences between set slants (in degrees) when considering the depth indicated either by the binocular disparities or by the distortion of the image by perspective. (a) The deviations from the mean value set by each subject ($n = 3$) for each slope ($n = 5$) and target (two per presentation). (b) The difference between the values set for the far and near target within each presentation. [Since the average set slant from binocular disparity was slightly larger for the far target, the vertical axis in panel (b) has been shifted by 5°.]

correlation (-0.07; $p = 0.04$), but again the correlation is weak and is not consistent across subjects. The individual correlation coefficients are -0.10 ($p = 0.07$), -0.09 ($p = 0.12$), and 0.17 ($p < 0.01$).

In the introduction, we explained that finding correlations between settings for the two cues can only be considered as evidence that information about the discrepancy is lost if we are sure that the correlation is not caused by fluctuations in the estimated value of the reference. Our method of analysis ensures that systematic biases (perhaps caused by misperceptions of space—Cuijpers et al 2002) cannot introduce correlations, because we only consider the deviations from the average values for a given position and orientation in space. However, if the perceived depth varies from trial to trial (see Brenner and van Damme 1999), estimates of the reference slant will also vary. If the reference slant varies, we should see a positive correlation between the settings for the two ellipses for each cue. We determined these correlations for each reference slant for each subject. The average of the fifteen correlations was 0.84 for the perspective cue and 0.32 for the binocular cue. Thus it is almost certain that the estimates of the reference slant varied across trials. However, if variability in estimating the reference slant had completely dominated the variability in the settings, we would have seen a strong positive correlation in figure 7a, because in our models such variability is equivalent to introducing a large error term for p in equations (4) and (5), or (7) and (8).

The direct influence of (both real and perceived) variability in the reference can be eliminated by taking the difference between the slant setting of the upper and lower ellipse for each of the cues, and comparing these differences. Figure 7b shows this comparison for all the settings. If all the judgments were independent, and variability in the reference negligible, we would have expected the points to form a similar pattern to that in figure 7a, but with a larger range (in all directions). If variability in estimating the reference slant had completely dominated the variability in the settings, we would have expected the range of errors to decrease for both cues (because a major source of variability is removed). What we see is inconsistent with both these predictions: the range increases for binocular disparity and decreases for perspective. The explanation is probably that the variability in the judged slant of the reference is caused by variability in estimating the distances of the ellipses. If the distance is misjudged, then subjects will not only misjudge the reference slant, but they will also misjudge the difference between the relationships between binocular disparity and slant for the two surfaces. Variability that arises from such errors is not removed by comparing differences between the slant setting of the upper and lower ellipse.

There was a significant correlation between the differences between the settings for the two cues (0.19; $p < 0.01$), which was consistent across subjects, but was quite small and was not significant for the individual subjects. The correlation coefficients for the individual subjects are 0.11 ($p = 0.18$), 0.14 ($p = 0.11$), and 0.04 ($p = 0.64$). It is difficult to evaluate the small positive overall correlation because a small negative correlation was found for the settings themselves (figure 7a), and because we have no independent estimate for the weight given to each cue when judging slant. Moreover, the apparent correlation between subjects' settings within one scene (leading to lower variability for the perspective cue in figure 7b than in figure 7a) complicates the interpretation. Considering that the correlations are hardly detectable we conclude that the results of the second experiment also provide no evidence for a loss of information about the discrepancy between the cues when they are combined.

4 General discussion

It is evident from figures 1f and 1h that not finding a correlation between the settings for two cues does not prove that the two judgments involved are based on the same estimates [as in equations (4) and (5)]. Thus, our failure to find clear correlations between the settings for the two cues (such as those shown in figures 1g and 1i) does not prove that there is no loss of information about the discrepancy. However, also considering the pattern of settings makes this claim more convincing, because an intersection of judgments based on different estimates [equations (7) and (8)] predicts that the settings will either be radially symmetric (for near-equal weights) or else they will be correlated (at least for normally distributed variability). Neither is the case in figures 4 and 7b.

Are these findings inconsistent with previous results? In most cue-conflict studies subjects were not asked to report perceived discrepancies between the cues. Even in the few studies in which they were encouraged to do so (eg Hillis et al 2002), the conditions were not specifically chosen to make such conflicts easy to detect. Thus, subjects may have perceived the consequences of a conflict but not been able to use this effectively for performing the task. For instance, in Hillis et al (2002) the subjects may have perceived a slightly different texture than was simulated when there was a cue conflict, but may not have been able to use this to improve their performance because the texture was different on each presentation anyway. We specifically chose conflicts that were easy to detect, and made sure that subjects made settings that avoided perceptual conflicts.

In the first part of this study we saw that subjects even gave considerable weight to changing size as a cue for motion in depth when they saw the target change size (fixed-random-cue experiments). This can be considered as a loss of information about the discrepancy, because, surely, changing size should be given little weight when the assumption underlying the use of changing size to judge motion in depth is so clearly violated. Our subjects could see the target change size, and nevertheless used a cue that specifically relies on the object not changing size to judge its motion. Similarly, data by van Ee et al (2002, 2003) on the combination of monocular and binocular slant cues suggest that even subjects who have explicitly indicated that they see a simulation of a slanted trapezoid do not completely ignore the slant specified by perspective when judging the slant of the trapezoid. Such findings are consistent with previous suggestions that different attributes (such as position, shape, motion, and size) are determined largely independently, even when they rely on the same input. In particular, it is consistent with the claim that no attempt is made to achieve consistency between the attributes (Brenner and van Damme 1999). The latter claim is also consistent with our earlier suggestion that the reason that changing size (as a cue for perceived motion in depth) is given less weight than its resolution predicts is that it is considered to be less reliable owing to the possibility that objects change size.

Our study suggests that information about discrepancies between cues is not lost when the cues are combined; not even when the discrepancies are small. In the first part of our study, subjects had to match velocities. In the second, they had to match a slant. In both cases we would have undoubtedly found a negative correlation between the cues if the subjects had not tried to set a second aspect (size or shape) as well. For each percept, subjects combine various cues. In each case they presumably do so in a more or less optimal manner (van Beers et al 1996, 1999; Eagle and Hogervorst 1999; Ernst and Banks 2002; Hogervorst and Eagle 2000; Knill and Saunders 2003). The optimal manner does not necessarily have to involve precisely the same computations for both judgments (as we assumed in our equations). However, if the computations were completely different, then the errors in the use of the cues would have been independent for the two judgments. This seems not to be the case. Nevertheless, we cannot conclude that the computations are identical (ie that detected discrepancies between cues for one judgment are actually used to make other judgments), because our analysis is only sensitive to aspects of visual processing that introduce errors. What we can conclude is that judgments of motion in depth and of changing size (and estimates of slant and shape) share the stage of processing at which the errors arise.

Acknowledgments. This work was funded by the BBSRC (grant 43/SO9621 awarded to Maarten Hogervorst) and by the European Commission Research Training Network "Perception for Recognition and Action" (HPRN-CT-2002-00226).

References

Allison R S, Howard I P, 2000 "Temporal dependencies in resolving monocular and binocular cue conflict in slant perception" *Vision Research* **40** 1869 – 1886

Atkins J E, Fiser J, Jacobs R A, 2001 "Experience-dependent visual cue integration based on consistencies between visual and haptic percepts" *Vision Research* **41** 449 – 461

Beers R J van, Sittig A C, Denier van der Gon J J, 1996 "How humans combine simultaneous proprioceptive and visual position information" *Experimental Brain Research* **111** 253 – 261

Beers R J van, Sittig A C, Denier van der Gon J J, 1999 "Integration of proprioceptive and visual position-information: an experimentally supported model" *Journal of Neurophysiology* **81** 1355 – 1364

Brenner E, Berg A V van den, Damme W J van, 1996 "Perceived motion in depth" *Vision Research* **36** 699 – 706

Brenner E, Damme W J M van, 1999 "Perceived distance, shape and size" *Vision Research* **39** 975 – 986

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

Bülthoff H H, 1991 "Shape from X: psychophysics and computation", in *Computational Models of Visual Processing* Eds M S Landy, J A Movshon (Cambridge, MA: MIT Press) pp 305 – 330

Bülthoff H H, Mallot H A, 1988 "Integration of depth modules: stereo and shading" *Journal of the Optical Society of America A* **5** 1749 – 1758

Cornilleau-Pérès V, Droulez J, 1993 "Stereo-motion cooperation and the use of motion disparity in the visual perception of 3-D structure" *Perception and Psychophysics* **54** 223 – 239

Cuijpers R H, Kappers A M L, Koenderink J J, 2002 "Visual perception of collinearity" *Perception & Psychophysics* **64** 392 – 404

Cutting J E, Vishton P M, 1995 "Perceiving layout and knowing distances: the integration, relative potency, and contextual use of different information about depth", in *Perception of Space and Motion* Eds W Epstein, S Rogers (Tampa, FL: Academic Press) pp 69 – 117

Eagle R A, Hogervorst M A, 1999 "The role of perspective information in the recovery of 3D structure-from-motion" *Vision Research* **39** 1713 – 1722

Ee R van, Adams W J, Mamassian P, 2003 "Bayesian modeling of cue interaction: bistability in stereoscopic slant perception" *Journal of the Optical Society of America A* **20** 1398 – 1406

Ee R van, Dam L C J van, Erkelens C J, 2002 "Bi-stability in perceived slant when binocular disparity and monocular perspective specify different slants" *Journal of Vision* **2** 597 – 607

Ernst M O, Banks M S, 2002 "Humans integrate visual and haptic information in a statistically optimal fashion" *Nature* **415** 429 – 433

Frisby J P, Buckley D, Wishart K A, Porrill J, Gårding J, Mayhew J E W, 1995 "Interaction of stereo and texture cues in the perception of three-dimensional steps" *Vision Research* **35** 1463 – 1472

Geisler W S, Kersten D, 2002 "Illusions, perception and Bayes" *Nature Neuroscience* **5** 508 – 510

Gillam B J, 1968 "Perception of slant when perspective and stereopsis conflict: experiments with aniseikonic lenses" *Journal of Experimental Psychology* **78** 299 – 305

Hillis J M, Ernst M O, Banks M S, Landy M S, 2002 "Combining sensory information: Mandatory fusion within, but not between, senses" *Science* **298** 1627 – 1630

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 of London, Series B* **265** 1587 – 1593

Hogervorst M A, Eagle R A, 2000 "The role of perspective effects and accelerations in perceived three-dimensional structure-from-motion" *Journal of Experimental Psychology: Human Perception and Performance* **26** 934 – 955

Jacobs R A, Fine I, 1999 "Experience-dependent integration of texture and motion cues to depth" *Vision Research* **39** 4062 – 4075

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

Johnston E B, Cumming B G, Parker A J, 1993 "Integration of depth modules: stereopsis and texture" *Vision Research* **33** 813 – 826

Knill D C, Saunders J A, 2003 "Do humans optimally integrate stereo and texture information for judgments of surface slant?" *Vision Research* **43** 2539 – 2558

Landy M S, Maloney L T, Johnston E B, Young M, 1995 "Measurement and modeling of depth cue combination: in defense of weak fusion" *Vision Research* **35** 389–412
Norman J F, Todd J T, 1995 "The perception of 3-D structure from contradictory optical patterns" *Perception & Psychophysics* **57** 826–834
Regan D, Beverley K I, 1979 "Binocular and monocular stimuli for motion-in-depth: changing-disparity and changing-size feed the same motion-in-depth stage" *Vision Research* **19** 1331–1342
Regan D, Erkelens C J, Collewijn H, 1986 "Necessary conditions for the perception of motion in depth" *Investigative Ophthalmology & Visual Science* **27** 584–597
Rogers B J, Collett T S, 1989 "The appearance of surfaces specified by motion parallax and binocular disparity" *Quarterly Journal of Experimental Psychology A* **41** 697–717
Turner J, Braunstein M L, Andersen G J, 1997 "Relationship between binocular disparity and motion parallax in surface detection" *Perception & Psychophysics* **59** 370–380
Young M, Landy M S, Maloney L T, 1993 "A perturbation analysis of depth perception from combinations of texture and motion cues" *Vision Research* **33** 2685–2696

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION
VOLUME 33 2004

www.perceptionweb.com