Available online at www.sciencedirect.com

SCIENCE DIRECT®

PERGAMON

Vision Research 43 (2003) 2451–2468

Vision Research

www.elsevier.com/locate/visres

Weighted linear cue combination with possibly correlated error

İpek Oruç [a,*], Laurence T. Maloney [a,b], Michael S. Landy [a,b]

[a] Department of Psychology, New York University, 6 Washington Place, 8th Floor, New York, NY 10003, USA

[b] Center for Neural Science, New York University, 6 Washington Place, 8th Floor, New York, NY 10003, USA

Received 10 October 2002

Abstract

We test hypotheses concerning human cue combination in a slant estimation task. Observers repeatedly adjusted the slant of a plane to 75°. Feedback was provided after each setting and the observers trained extensively until their setting error stabilized. The slant of the plane was defined by either linear perspective alone (a grid of lines) or texture gradient alone (diamond-shaped texture elements) or the two cues together. We chose a High and Low variance version of each cue type and measured setting variability in four single-cue conditions (Low, High for each cue) and in the four possible combined-cue conditions (Low–Low, Low–High, etc.).

We compared performance in the combined-cue conditions to predictions based on single-cue performance. The results were consistent with a linear combination of estimates from cues. Six out of eight observers did better with combined cues than with either cue alone. For three observers, performance was consistent with optimal combination of uncorrelated cues. Three other observers' results were also consistent with optimal combination, but with the assumption that internal cue estimates were correlated. The remaining two observers were consistent with sub-optimal cue combination.

Keywords: Cue combination; Slant perception; Pictorial depth cues; Correlated cues

1. Introduction

Many visual tasks involve estimation of object properties such as size, shape, depth, orientation, and location. For each of these properties, there is typically more than one method to derive an estimate of the property from available visual information. We refer to these distinct methods as *cues*. In estimating depth, for example, human observers can use any of several different depth cues (see, e.g., Kaufman, 1974). Multiple visual cues are available for the localization of simple luminance-defined features (Watt, Morgan, & Ward, 1983) and of texture-defined edges (Landy & Kojima, 2001). In Fig. 1, we present a concrete example where an estimate of the slant of the plaza with respect to the observer's line of sight could be based on either a texture gradient cue or a linear perspective cue.

When there are multiple cues available, it is of interest to determine whether observers combine information from the cues and what rule of combination they use. If we denote the estimates derived from each of n cues by random variables $S_1, S_2, \ldots, S_n$, then the problem of cue combination is to reduce these n estimates to one estimate, S. The most common cue combination model in the literature is a *weighted linear combination*

$$S_w = \sum_{i=1}^{n} w_i S_i, \tag{1}$$

where $w_1, w_2, \ldots, w_n$ are non-negative weights, constrained to sum to 1. We assume that the expected value of each cue in isolation is the true value of the property, s, i.e. the estimate available from each cue is unbiased. Because the weights are constrained to sum to 1, the combined estimate is also unbiased. The variance of the estimator S_w is determined by the choice of weights and the variances of the individual cues, $\mathrm{Var}(S_i) = \sigma_i^2 > 0$. If the cues are uncorrelated, the choice of weights that minimizes the variance of the combined estimate is

$$w_i = \sigma_i^{-2} \Big/ \sum_{j=1}^{n} \sigma_j^{-2}, \tag{2}$$

a result due to Cochran (1937). If we define the *reliability* of a cue to be the reciprocal of its variance,

* Corresponding author. Tel.: +1-212-998-7851; fax: +1-212-995-4349.

E-mail address: ipek@cns.nyu.edu (İ. Oruç).

0042-6989/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0042-6989(03)00435-8

Fig. 1. *Piazza San Marco, Venice*. Two pictorial cues are available to define the slant and tilt of the plaza with respect to the line-of-sight of the camera, linear perspective and the texture gradient formed by humans, pigeons, other miscellaneous objects, and their shadows.

$r_i = \sigma_i^{-2}$, then the combined estimate has minimum variance precisely when the weight assigned to each cue is proportional to its reliability. This combination rule is a form of weak fusion, in the terminology of Clark and Yuille (1990) (see Landy, Maloney, Johnston, & Young, 1995). With the optimal weights in Eq. (2), the reliability of the combined estimate S_w is

$$r = \sum_{i=1}^{n} r_i, \tag{3}$$

and it is evident that the reliability of the combined estimate is as great or greater than the reliability of any single cue.

In this paper, we are interested in studying cue combination as an ideal observer problem without the strong assumptions that cues are Gaussian random variables or that they are uncorrelated. In Appendix A, we derive the optimal choice of weights for linear cue combination when the internal estimates corresponding to the cues are correlated. We present the result for the case where there are only two cues in the next section. We report an experiment where observers were asked to estimate surface slant using two pictorial cues: linear perspective and texture gradient. The purpose of our experiment was to compare human performance in single-cue and combined-cue conditions, and to test whether human cue combination is optimal for the conditions of our experiment. The pattern of optimal behavior is different if cues are correlated, and it depends on the value of correlation. The design of the experiment allows us to test whether the two cues are correlated, whether human cue combination is optimal, and whether a linear combination rule is an appropriate model for human cue combination.

As will become clear, not all our observers' behavior was consistent with optimal cue combination. We would also like to test whether these sub-optimal observers benefited from combining cues in the sense that performance with combined cues exceeded performance with either cue alone.

The weighted linear rule with weights proportional to cue reliability satisfies other criteria of statistical optimality. We did not have to assume that the uncorrelated cues are Gaussian random variables to derive Eq. (2) but, if they are, there is no unbiased non-linear rule that has lower variance. In the Gaussian case, Eq. (1) is also the maximum likelihood estimate of s, and, if we introduce a Gaussian prior on the unknown parameter, it is of the same form as the maximum a posterior (MAP) estimate but with an additional term S_0 (the mean of the prior distribution) introduced into the weighted sum. All of these results are derived in Appendices A and B.

1.1. Previous experimental work

There is considerable evidence suggesting that human observers do use a weighted linear combination rule when multiple cues are available, and that the weights depend on cue reliability. Young, Landy, and Maloney (1993) found that texture and stereo cues to depth were combined using a weighted average and that the weights changed in the appropriate directions when noise was added to each cue. Johnston, Cumming, and Landy (1994) found weighted averaging behavior for combinations of motion and stereo cues to depth. Ghahramani, Wolpert, and Jordan (1997) looked at the combination of auditory and visual information for localization. Their results were consistent with weights based on reliability, although visual reliability was so much higher than auditory in their experiment that it was impossible to assess whether weights were optimal.

Van Beers, Sittig, and Denier van der Gon (1998, 1999) studied combinations of visual and proprioceptive cues to location within a plane. They indirectly estimated the reliabilities of vision and proprioception alone (Van Beers et al., 1998). They found that vision was more reliable for width than depth discriminations, and that proprioception was more reliable tangential to the forearm than radially along it. Because these two discrimination ellipses are typically not aligned, a maximum likelihood estimator that assumes Gaussian individual cue estimates predicts that, in cue conflict conditions, estimates should lie on a curve (i.e., not a straight line) between the visually and proprioceptively defined locations. They found weak evidence for this qualitative prediction.

These studies did not separately measure the reliability of single cues. Instead, they varied the depth that each cue signaled and analyzed the effect of changing a cue on the observer's response across many trials, without testing whether observers use correct weights (Eq. (2)) in combining cues. Moreover, while the studies just described are all consistent with the claim that observers used a weighted linear rule to combine cues, they are also consistent with an alternative "cue-switching" strategy. The cue-switching observer uses only one cue on each trial, but varies the proportion of times each cue is used as a function of its reliability. When observers' judgments are averaged across trials, this cue-switching strategy would resemble the results just cited in many respects. However, the performance of a cue-switching observer with combined cues could never exceed his or her performance with the more reliable of the available cues in isolation. Further, a cue-switching strategy would lead to flattened discrimination psychometric functions as cue conflicts increase, which was not found in these studies (see Landy & Kojima, 2001, for further discussion).

Landy and Kojima (2001) studied the localization of a texture-defined edge cued by changes in two textural features (e.g., local orientation and scale). Performance was measured both with edges defined by pairs of cues and edges defined by each cue alone. The single-cue data provide an estimate of the reliability of the individual cues which, in turn, provides an estimate of the optimal cue weights to use for the two-cue stimuli. In making this prediction, Landy and Kojima assumed that estimates of localization from the two cues could be treated as uncorrelated random variables. An ideal cue combination model was fit to the entire data set, and it predicted many aspects of the data quite well.

Ernst and Banks (2002) used a similar strategy for the case of visual and haptic estimation of the size of an object. Single-cue studies were used to estimate the reliability of the two modalities alone. The cue weights and the variability in the two-cue conditions were well-predicted by the individual cue reliabilities and the ideal observer model. This successful prediction held even as the visual reliability was manipulated by the addition of visual noise to the stimuli. The amount of noise was varied within a single block of trials. The data were consistent with a dynamic choice of cue weights, on a trial-by-trial basis, where the choice of visual weight depended on the visual noise presented on any given trial.

Using a discrimination paradigm rather than an estimation task, Hillis, Ernst, Banks, and Landy (2002) found evidence for cue combination using an optimal choice of weights. This occurred both for combinations of haptic and visual cues to size (as in Ernst & Banks, 2002) as well as for combinations of stereo and texture cues to surface slant. A similar study by Knill and Saunders (2002) confirmed that observers use optimal weights for combining stereo and texture cues to slant.

Jacobs (1999) examined cue combination for depth estimates based on texture and motion. The combined-cue stimuli were rendered elliptical cylinders defined by a texture cue (circles placed randomly on the surface of the cylinder) and a motion cue (smooth movement of the texture elements within the surface). Single-cue stimuli consisted either of static texture or dots moving within the surface of the cylinder. Observers indicated perceived depth by adjusting an ellipse until it appeared equal in shape to the perceived cross-section of the cylinder. The variability for combined-cue stimuli was somewhat lower than for single-cue stimuli (although not always). A scatterplot of depth predicted by optimal weighted cue combination (using the single-cue results) versus perceived combined-cue depth showed high correlation, although much of that was due to the wide variation in portrayed depth. Better scatterplots resulted when the prediction also included a prior bias for circularity but, all in all, this study provides scant evidence of optimal weighting by observers.

1.2. Cue combination with correlated cues

An assumption shared by all of the above-mentioned studies is that cues are uncorrelated. This assumption is plausible when cues are drawn from distinct modalities as in Ernst and Banks (2002). Visual cues, however, are all based on the same retinal image, and certainly share at least one noise source, the Poisson statistics of the retinal quantum catch. They may share sources of neural noise as well. We think of cues as neatly modularized in the stimulus and in the neural processing that follows. But, binocular disparity and motion are often both relevant for the responses of neurons (e.g., in cortical area MT) even though we think of stereo and motion as separate depth cues. Similarly, V1 neurons are often tuned for both orientation and spatial frequency even though they were treated as separate "cues" by Landy and Kojima (2001). Of course, even when cues are based on the same retinal information, they may still be uncorrelated particularly if the visual system reorganizes itself so as to de-correlate cues (Barlow, 1989; Barlow & Földiák, 1989).

2. Combining correlated cues

In this section, we present basic results for weighted linear cue combination for two cues. In Appendix A, we derive these results for n cues. Here, we can set $w_1 = w$, and then $w_2 = 1 - w$. When the two cues, S_1 and S_2, are uncorrelated, the reliability of a weighted linear combination of the two cues is

$$r_w = [w^2 r_1^{-1} + (1-w)^2 r_2^{-1}]^{-1}. \quad (4)$$

We plot r_w versus w in Fig. 2A. Note, first of all, that $r_w = r_1$ when $w = 1$, and that $r_w = r_2$ when $w = 0$. In these two cases, the observer has discarded one of the cues and the reliability is that of the other cue. The smooth curve reaches a maximum of $r_1 + r_2$ when the weights are chosen to minimize variance and, accordingly, maximize reliability. Notice that, even if the observer fails to choose exactly the optimal weights, the reliability of the combined-cue estimate is higher than the reliability of either cue alone over a wide range of weights. These results indicate that the observer can benefit from cue combination even if his or her knowledge of the reliabilities of the available cues is imprecise. There are several papers in the statistical literature discussing conditions under which Gaussian cue combination based on imprecise estimates of cue reliabilities is beneficial (Cohen & Sackrowitz, 1974; Graybill & Deal, 1959; Zacks, 1966).

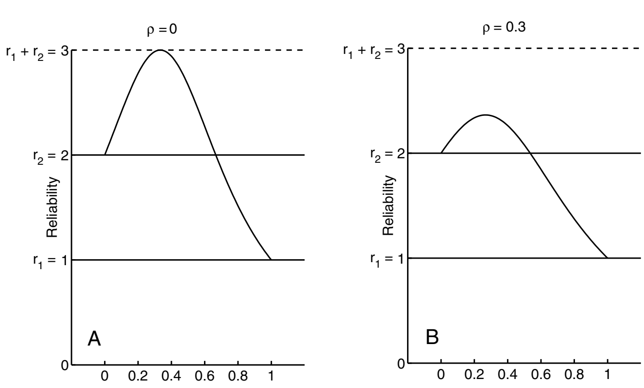

Fig. 2. *Reliability of the weighted linear combinations of cues.* The reliability r_w of the weighted linear combination $wS_1 + (1-w)S_2$ is plotted versus w for hypothetical cue S_1 with reliability $r_1 = 1$ and hypothetical cue S_2 with reliability $r_2 = 2$. Reliability is the reciprocal of variance. (A) The cues are uncorrelated ($\rho = 0$). The maximum value of r_w is the sum $r_1 + r_2 = 3$. It is marked on the plot. (B) The correlation between the cues is $\rho = 0.3$.

If the two cues S_1 and S_2 are correlated with correlation ρ, the optimal choice of weights in Eq. (1) is given by the formula

$$w = \frac{r_1'}{r_1' + r_2'}, \quad (5)$$

where $r_i' = r_i - \rho\sqrt{r_1 r_2}$, $i = 1, 2$ is the *corrected reliability* of the cue, discounted for correlation. If the correlation is zero, then Eq. (5) reduces to Eq. (2).

The reliability of the weighted linear estimate in the correlated case is

$$r_w = [w^2 r_1^{-1} + (1-w)^2 r_2^{-1} + 2\rho w(1-w)(r_1 r_2)^{-1/2}]^{-1}. \quad (6)$$

We plot r_w versus w in Fig. 2B for the same reliabilities r_1 and r_2 as in Fig. 2A, but now with $\rho = 0.3$. Note, first of all, that the maximum possible reliability is less than in Fig. 2A even though the reliabilities of the cues are unchanged. The effect of correlation is always to reduce the maximum possible reliability when weights are constrained to be non-negative and to sum to 1. [1]

If we substitute the optimal values for the weights, the resulting optimal reliability is

$$r = \frac{r_1 + r_2 - 2\rho\sqrt{r_1 r_2}}{1 - \rho^2}. \quad (7)$$

When $\rho = 0$, Eq. (6) reduces to Eq. (4), and Eq. (7) reduces to Eq. (3).

3. Experimental implications

In Fig. 3, we summarize the possible outcomes of an experiment in which we measure the reliability of two cues, r_1 and r_2, and the reliability of the two cues combined, r_c. Assume, for convenience, that $r_1 < r_2$. The schematic in Fig. 3 is a caricature of Fig. 2A. Solid horizontal lines mark r_1 and r_2. Consider the possible values that r_c can take on if the two cues are combined

[1] It is possible to find reliabilities r_1, r_2 and a value of correlation ρ such that the maximum possible reliability of a weighted linear combination of two cues with these reliabilities and correlation exceeds that achievable with two uncorrelated cues with the same reliabilities. In this case, one of the weights must be negative. We discuss this case in Appendix A.

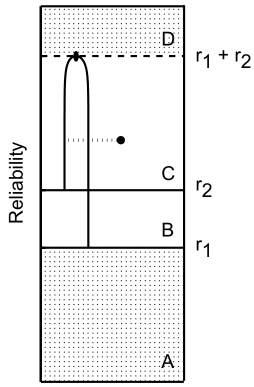

Fig. 3. *A reliability plot.* The plot summarizes the outcome of two single-cue conditions and the combined-cue condition. Two dark horizontal lines mark r_1 and r_2, the reliabilities of the two single cues. A dotted line marks $r_1 + r_2$, the maximum reliability achievable by a weighted linear combination of two uncorrelated single cues. A black dot marks the observer's reliability in the combined-cue condition. The regions of the plot are labeled A, B, C, and D. If the black dot falls into regions A or D, then the observer's performance is inconsistent with a weighted linear combination of cues with non-negative weights. If it falls in C, then the observer's reliability with two cues is greater than it would be with either cue alone, evidence that cue combination is occurring. The curved line is a schematic of the curve in Fig. 2A showing how the reliability of a weighted linear combination rule varies with weight.

by a weighted linear combination with non-negative weights summing to 1. The minimum possible value for r_c is r_1 and this occurs when the weights applied to the two cues are 1 and 0. The maximum possible value, if the cues are uncorrelated, is $r_1 + r_2$ and we have marked this maximum with a dashed horizontal line. The reliability of a weighted cue combination with non-negative weights summing to 1 must fall between r_1 and $r_1 + r_2$.

Consequently, if the observed value of r_c falls outside this region, we can reject the hypothesis that the cues are combined by a weighted linear combination rule with non-negative weights. The rejection region is shaded in the diagram. If r_c falls in the lower region (Region A), then the combined-cue estimate is worse than either cue alone which is not possible with a weighted linear combination rule with non-negative weights. The cues are interfering with one another when both are present. If r_c falls in the upper rejection region (Region D), then the outcome is too reliable to be the result of a weighted linear combination of the two cues. If the distribution of the cues is Gaussian, then this cannot happen. But with other distributions there could be non-linear rules where this is possible (Appendix B).

If r_c falls into Region B, the combined-cue reliability falls between the reliabilities of the cues in isolation. This outcome is consistent with a weighted linear combination but with a choice of weights that is distinctly sub-optimal. The observer would be better off ignoring the less reliable cue.

Suppose now that r_c falls into the interior of the region labeled C: $r_2 < r_c < r_1 + r_2$. The reliability of the combined cues is greater than that of either cue alone, demonstrating that within-trial cue combination has occurred. We can, in particular, reject the cue-switching hypothesis (that the observer is carrying out the task by using one cue on each trial but changing cues from trial to trial). However, the observer's performance is still sub-optimal compared to the prediction of the weighted linear combination rule for uncorrelated cues. It is possible that the cues are uncorrelated but that the observer has simply picked weights that are sub-optimal. Even though the weights are sub-optimal, we note that the observer is still benefiting from combining cues. But a second possibility is that the observer is behaving optimally but that the cues are correlated. Finally, the cues could be correlated and the observer could also choose sub-optimal weights, in which case r_c could be either in Region B or C.

The design of the experiment we have described (measure the reliability of two cues in isolation and in combination) does not allow us to distinguish the presence of correlation between the cues from a sub-optimal choice of weights in a linear combination rule. To do that, we need a more complex experimental design where we use two levels of reliability for each cue and measure performance in the four possible single-cue conditions and the four possible combined-cue conditions. We describe how we do this as part of the methods section that follows.

4. Experiment

4.1. Introduction

We next describe an experiment in cue combination designed to detect correlation between the cues. In previous work, cue independence has generally been assumed, and results have by and large been consistent with cue independence either because the cues were logically independent as they came from different modalities (Ernst & Banks, 2002; Gepshtein & Banks, 2003), or two relatively independent visual cues were used such as texture, motion and/or stereo disparity (Hillis et al., 2002; Johnston et al., 1994; Knill & Saunders, 2002; Landy et al., 1995; Young et al., 1993). When the cues were less clearly "separable", such as the orientation and spatial frequency cues to a texture border used by Landy and Kojima (2001), there were indications of cue correlation (although not pointed out by the authors). In that paper, the reliability of the combined cues was not always better than that of the individual cues. In the experiment we describe now, we intentionally have chosen two pictorial cues to depth (texture and linear perspective) that arguably are not truly separate cues and thus might be expected to show evidence of cue correlation.

In this experiment, the observer was trained to set an inclined plane to a criterion slant (75°). The slant of the plane was signaled by either a texture gradient, a linear perspective grid, or both cues superimposed. Fig. 1 is an example of an everyday scene containing both cues. The design of the experiment allowed us to test the hypotheses that the rule of combination is a weighted linear combination with non-negative weights, the choice of weights is optimal, and the internal cue estimates are uncorrelated. In cases where the choice of weights is sub-optimal we will test whether the combination was beneficial.

4.2. Methods

4.2.1. Apparatus

The stimuli were displayed on a SONY Trinitron 21 in. monitor (Model GDM-G500) using a computer equipped with a VSG 2/3 frame buffer from Cambridge Research Systems. The screen of the monitor is close to physically flat (less than 1 mm of deviation along any horizontal line). It was covered with a black cardboard mask with a rectangular aperture revealing only the center region of the screen where stimuli were presented. The observers viewed the stimuli in an otherwise dark room. To further minimize distractions and extraneous visual cues, the observers viewed the stimulus region through a hood made out of black cardboard. The observers were seated 109 cm away from the screen, resting their heads on a chin rest positioned so that viewing was monocular (using the right eye), from a specified viewing point. At the given viewing distance, the aperture subtended approximately 19.9° × 14.4°.

4.2.2. Stimuli

The stimuli were virtual rectangular planes that, from the observer's viewpoint, appeared to be behind the plane of the monitor surface. The virtual plane could be set to any specified slant between 70° and 80°. The virtual plane could contain either or both of the two following kinds of patterns: (1) a grid of horizontal and vertical lines (the *linear perspective cue*) and (2) a randomly distributed collection of diamonds (the *texture gradient cue*). It might be possible for observers to memorize the shape of a specific texture element, or certain location on the grid and use it as an artifactual cue to slant. To avoid this, for each trial we randomized the position of all the texture elements, and the phase of the grid with respect to the plane. The locations of the diamonds were computed as follows: the number of diamonds on any given trial was randomly chosen from a Poisson distribution with one of two possible density values (one for the "high variance" and one for the "low variance" condition). The location of each diamond was then chosen from a uniform distribution over the plane's surface subject to the constraint that no two diamonds overlap. The contents of the plane at any slant were mapped onto the display region via perspective projection appropriate to the observer's viewpoint. An example of a projected stimulus is shown in Fig. 4.

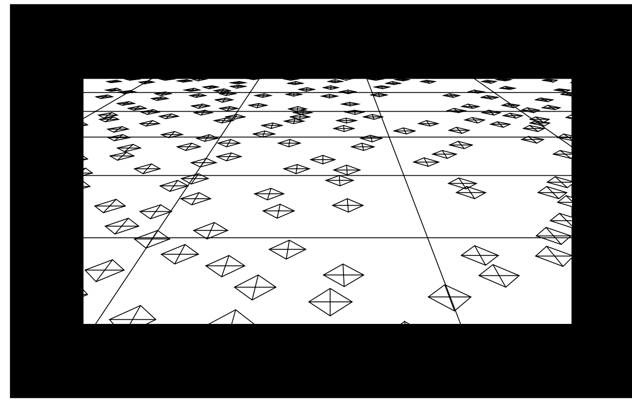

Fig. 4. *An example of a stimulus*. This is a demonstration of a stimulus for a combined-cue condition. Both of the cues (line grid and random diamond texture) are present. In the actual stimuli, the lines and texture elements were white on a gray background.

The bounding contour of a rectangular plane is potentially a perspective cue to slant. The rectangular plane was chosen to be so large that its interior filled the viewing aperture and its edges were never visible.

The choice of these two cues, texture and linear perspective, was intended to provide stimuli for which the independence assumption was likely to fail. The rectangular cells formed by the grid can be considered to be "texture elements" and the foreshortening of the diamond shaped texture elements is, in effect, a linear perspective cue. One might expect the computation of slant estimates from the grid and the texture to share common mechanisms within the visual system that would incorporate common sources of noise. This would, in turn, reveal itself as correlation between the estimates of the two cues.

4.2.3. Cue conditions

If we only consider one texture cue, one linear perspective cue and their combination, we cannot discriminate between the possibility that the cues are correlated and the possibility that observers pick weights that are sub-optimal. Accordingly, we consider stimuli formed by two density levels of each of two depth cues (Fig. 5). For the texture gradient cue, these are denoted T1 and T2, differing in density of the texture gradient of randomly placed diamonds. The average number of diamonds visible to the observer at 75° slant was 40 and 110 for the two levels of density, T1 and T2. For the linear perspective cue, the levels are denoted L1 and L2, differing in the density of lines in a linear perspective

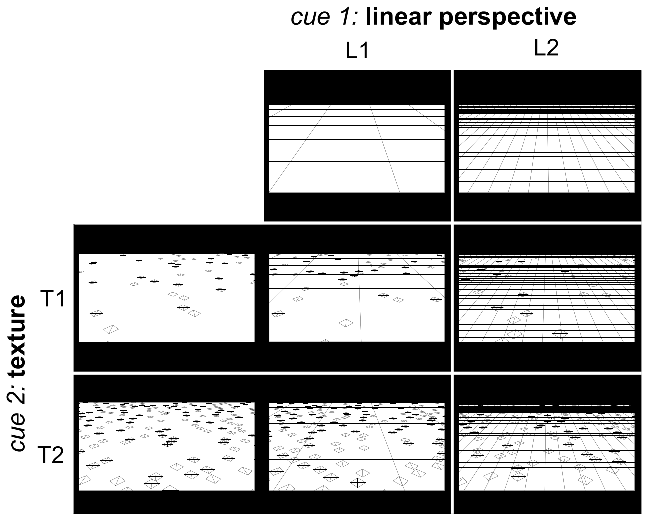

Fig. 5. *The experimental conditions*. There were four single-cue conditions consisting of two density levels of each cue type. These are shown on the left and top margins of the array of stimuli. The combined-cue conditions are also shown. Each combined-cue condition is made up of the single cues on the margin at the left end of its row and at the top of its column.

pattern. On average four horizontal and four vertical lines were visible at 75° slant for density level L1. And for L2, 30 horizontal and 28 vertical lines were visible. We will estimate the observer's reliability in each of the four possible single-cue conditions (T1, T2, L1, L2) and in each of the four possible combined-cue conditions (T1 + L1, T1 + L2, T2 + L1, T2 + L2).

In pilot testing we verified that the observers' setting reliabilities were substantially different for the different levels of each cue type. However, there was no consistency across observers as to which level of each cue (higher or lower) resulted in higher reliability. Across the range of texture and line densities used here, an ideal observer's performance improves with the additional information provided by more texture elements or lines. However, individual differences abound in depth perception studies (such as the widely ranging cue reliabilities across subjects found by Hillis et al., 2002, or the large variation in abathic distance found by Kontsevich, 1998). Thus, we never report averaged data for this sort of study, and have studied a relatively large number of subjects expecting to find a wide range of cue reliabilities.

In fitting the data, we will assume that any correlation between cues is the same in all combined-cue conditions. With this assumption and this experimental design, we will be able to estimate the correlation between cues separately from the weights chosen by observers, as described below. We can then test whether this correlation is non-zero.

4.2.4. Observers

Eight observers participated in the experiment. Four were experienced and three were inexperienced psychophysical observers all of whom were unaware of the purpose of the experiment and were paid for their participation. The last observer (IO) was one of the authors.

4.2.5. Task

The observer viewed the stimulus and adjusted its apparent slant by pressing a key. Each key press changed the slant by 0.5°. S/he adjusted the plane until its slant matched that of a remembered criterion (75°). The observer received auditory feedback on his or her setting on every trial. There were five levels of possible feedback. A correct answer (i.e. choosing the setting for 75°) was rewarded with a short melody. For settings within three steps of 75° (±1.5°), the feedback was half the melody, played in the lower octave if the setting was less then 75°, and in the higher octave if the setting was greater than 75°. If the observer's setting was further from 75°, the feedback was a pure tone, with very low pitch for lower, and with very high pitch for higher settings.

4.2.6. Procedure

Faced with the very first trial, the observer does not know what the criterion slant is. Therefore, his or her setting at the first trial is necessarily a guess. The criterion slant is learned over many trials through the feedback provided after each setting. The observers are provided with training trials to facilitate learning. A training trial is similar to an experimental trial, except that the observer gets three chances to make a setting, instead of one. In other words, after the observer makes a setting and hears the feedback, s/he can adjust their setting further, two more times, according to the feedback provided each time.

In each session, the observer completed both training and experimental blocks. Each block consisted of eight trials, one from each of the eight conditions (four single-cue, four combined-cue), in random order. The number of training blocks was relatively high in the first couple of sessions, and was decreased towards the end as the observer improved at the task. Again, the observer received feedback on all trials; on training trials, there were two additional chances to set slant correctly. A typical session consisted of 54 blocks where the observer would complete both experimental and training blocks.

The training block data were excluded from the analysis. For the experimental trials, we computed the *setting error* in each block as the squared difference between the observer's setting and 75°, summed over eight conditions in the block. We use the setting error as a measure of how well the observer performed the task, and plot it against the block number. An example of such a learning curve, for one observer, is shown in Fig. 6. We required the observer to perform the task until the setting error stabilized. Stabilization was assessed by visual inspection of the learning curves. Observers

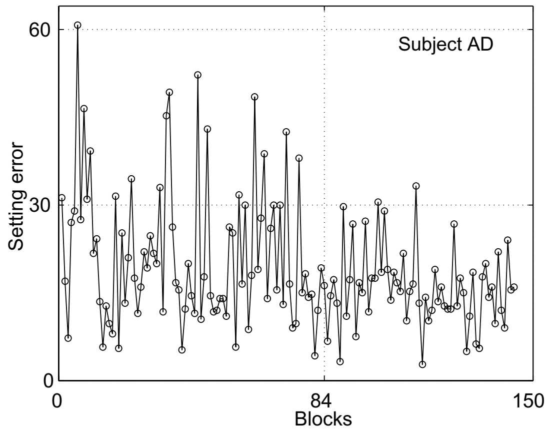

Fig. 6. *Setting error across time for one observer.* The horizontal axis is experimental block. The vertical axis is the setting error in the corresponding block. The setting error, initially high, decreases to a stable level. Observers practiced the task with feedback until their setting error stabilized (for this observer, after 84 blocks). The observer then completed 60 more experimental blocks (along with additional training blocks) and these data were used in the analysis.

typically took up to 120 blocks before setting error stabilized. After the observer was judged to have stabilized, s/he ran 60 experimental blocks, along with additional training blocks interspersed among the training blocks. For the observer in Fig. 6, the first 84 blocks were discarded.

Due to the extensive training, we did not expect setting biases. Indeed, the absolute value of deviation from 75° averaged over all observers was only 0.32°. Considering that the adjustable slants in the experiment were a discrete set of angles regularly spaced at 0.5° intervals, we can safely say that the observers' settings were unbiased.

4.2.7. Results and discussion

Let $S^1, S^2, \ldots, S^{60}$ denote the observer's settings in one condition. Typically in a small percentage of trials observers are expected to make mistakes due to failures of attention and so on. Estimates of variance (and reliability) are very sensitive to such outliers. To avoid such inflation, we first converted the data for each condition into *z*-scores and removed the trials whose z-scores fell outside the interval (−2, 2). Typically, two trials were discarded from each set of 60. We computed the sample variance of the remaining ("trimmed") data.

If $S^1, S^2, \ldots, S^{60}$ are independent, identically distributed Gaussian random variables with variance σ^2, then the expected value of the trimmed sample variance is about $0.77\sigma^2$. We corrected our estimates of variance (and reliability) by dividing by 0.77. All the following analysis is based on this *corrected trimmed variance* measure, an example of a robust statistical estimator (Hampel, Ronchetti, Rousseeuw, & Stahel, 1986). We take as our estimate of reliability

cue 1: **linear perspective**

Observer IO			L1	L2
cue 2: **texture**			0.48	0.64
	T1	0.93	1.14 [1.41]	1.45 [1.57]
	T2	1.39	1.54 [1.87]	1.85 [2.03]

Fig. 7. *Results for one observer.* The observer's reliability in each of the experimental conditions is displayed in the same format as Fig. 5. In addition, the estimated maximum reliability of a weighted linear cue combination (when cues are uncorrelated) is also shown marked with a rectangle in the combined-cue cells. This maximum is simply the sum of the marginal single-cue estimates.

$$\hat{r} = \mathrm{TVar}(S^1, \ldots, S^{60})^{-1}, \tag{8}$$

where $\mathrm{TVar}(S^1, \ldots, S^{60})$ is the corrected trimmed variance measure just described. We estimated reliabilities for each of the eight conditions separately for each observer.

Fig. 7 shows the estimated reliabilities for one observer (IO) tabulated in the same format as the stimuli in Fig. 5. In the marginal cells we see the reliabilities measured for the four single-cue conditions. We know that in a combined-cue condition the optimum combination rule yields a reliability equal to the sum of the two single-cue reliabilities. These predicted optimal values are marked with a rectangle around them. The other values in these cells are the measured reliabilities for the combined-cue conditions. A quick visual comparison shows that the measured reliabilities are slightly lower than expected from an optimal combination rule, but consistently better than the reliability of the better of the two cues in isolation. We will turn to formal statistical tests in a moment but, for now, we note that IO is an observer whose performance is consistent with a weighted linear cue combination with non-negative weights and who has chosen weights that result in a higher reliability in combined-cue conditions than could be achieved with either cue alone. This observer is not simply "cue switching".

In Fig. 8, we plot the results for IO and the other seven observers in the format of Fig. 3. In discussing Fig. 3, we assumed we had perfect knowledge of single-cue reliabilities and combined-cue reliabilities and now we have only estimates of these values. Each of the estimates is shown with its 95% bootstrap confidence interval (Efron & Tibshirani, 1993). Fig. 8A shows the data for observer IO, already seen in Fig. 7, plotted in this format. There are four sub-plots, one for each of the four combined-cue conditions. In each sub-plot, filled circles indicate the reliabilities for the combined-cue

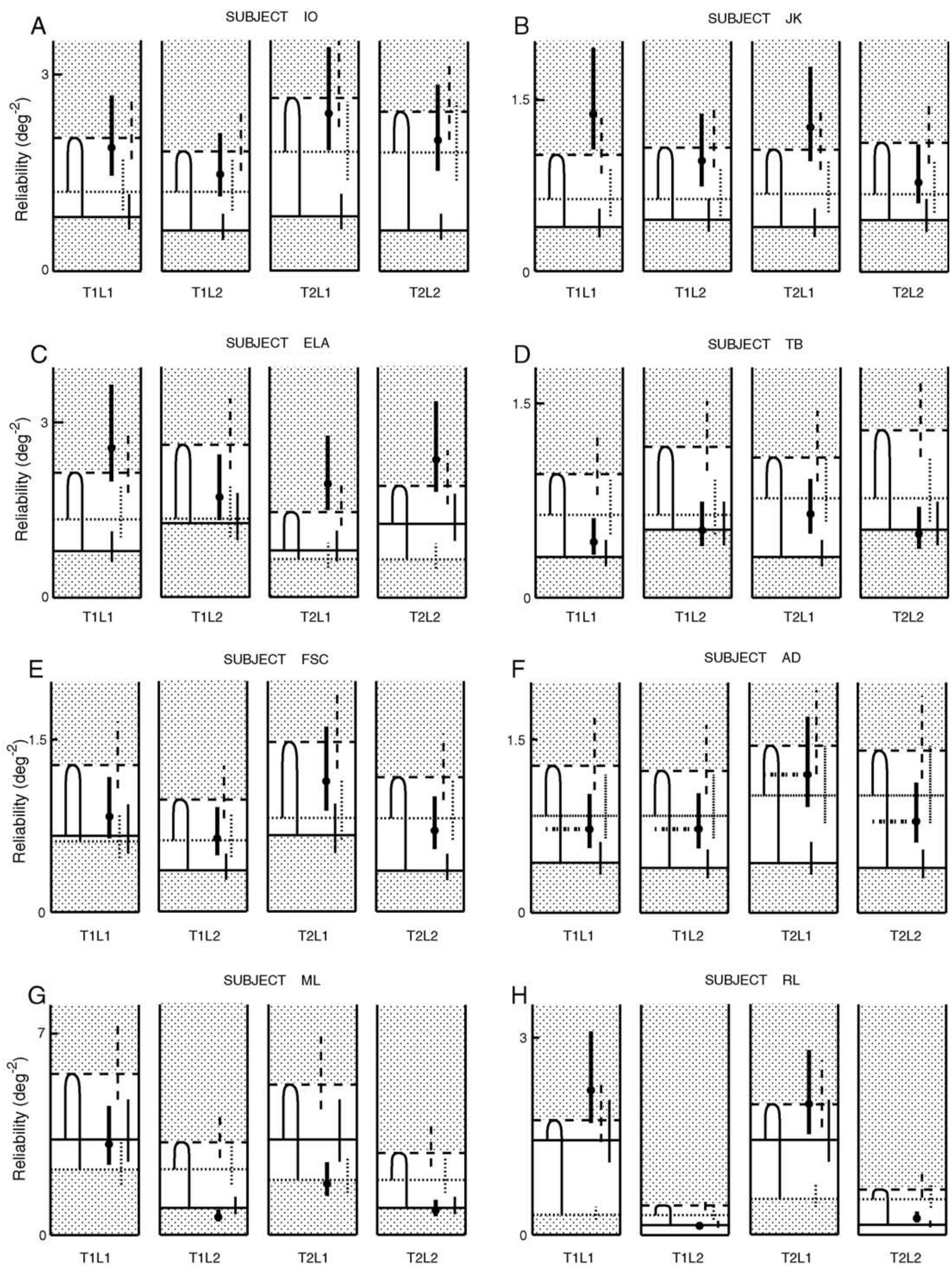

Fig. 8. *Reliability plots for all observers*. The four combined-cue conditions for each observer are plotted in the format of Fig. 3 with 95% bootstrap confidence intervals added (Efron & Tibshirani, 1993). In each sub-plot the dotted horizontal line represents the reliability level of the texture cue, the solid horizontal line marks the reliability level of the linear perspective cue, and the dashed horizontal line marks the optimal combined reliability for uncorrelated cues. The cue levels for the two cues are noted below each sub-plot. The *y*-axis indicates reliability. Each observer's data (a group of four sub-plots) are plotted using the same *y*-axis scale factor; the scale factors differ between observers.

condition, and the reliabilities of the single cues are marked with horizontal lines, dotted for the texture cue, and solid for the linear perspective cue. The sum of the single-cue reliabilities is also marked as a dashed horizontal line.

We would like to determine whether human reliability in the combined-cue case is consistent with the predictions for optimal cue combination of uncorrelated cues, that is, in each sub-plot for each observer, is the reliability in the combined-cue condition (the filled

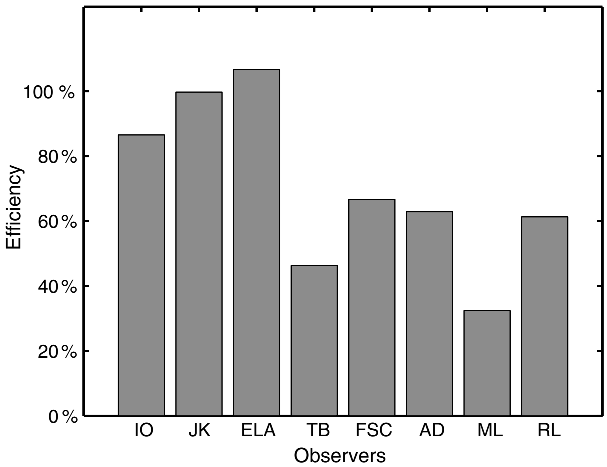

Fig. 9. *Efficiency indices: uncorrelated case*. The efficiency indices for each of the eight observers are plotted. A value of 100% indicates optimal performance: the observer is doing as well as a weighted linear combination could achieve given the observer's single-cue reliabilities. See text for details.

circle) equal to the sum of the reliabilities in the single-cue conditions (the dashed line)? A visual inspection of the plots seems to suggest that cue combination strategies of observers IO, JK and ELA are approximately optimal, whereas observers TB, FSC, AD, and ML seem to have employed less than optimal combination rules. Results for observer RL, on the other hand, are more complicated. There is a marked discrepancy between the four cases in Fig. 8H. Two of the four sub-plots suggest optimal combination, and the other two seem sub-optimal.

To quantify how sub-optimal an observer is, we define a measure of efficiency, Ω. For each observer, we have four combined-cue conditions that we will number from 1 to 4. For condition k we have the observer's measured reliability with both cues, $r_{12,k}$, and the sum of the observer's measured reliabilities for the two single-cue conditions, $r_{1,k} + r_{2,k}$. We form the *efficiency index*

$$\Omega = 100 \times \left(\prod_{k=1}^{4} \frac{r_{12,k}}{r_{1,k} + r_{2,k}} \right)^{1/4} \quad (9)$$

for each observer. [2] Fig. 9 shows the efficiency of each of the eight observers. Note that we are holding the observers to a very high criterion. If there is correlation between the cues, then even an optimal observer can appear sub-optimal by this index.

Observers IO, JK and ELA evidently have high efficiencies, above 85%, and observers TB, FSC, AD, ML, and RL all have markedly lower efficiencies, ranging between 32% and 66%. To determine whether these values are consistent with optimal behavior, we performed a bootstrap simulation. For each human observer we simulated an ideal observer with the same single-cue reliabilities as the human, but that combines cues optimally, and computed 1000 samples of Ω for this ideal observer. The rejection region [3] is the upper and lower 2.5 percentiles of the histogram formed by the 1000 samples of Ω. If the efficiency of the observer is too low compared to that of the ideal, then s/he is judged to be sub-optimal. On the other hand, if Ω is too high, then this observer is judged to be super-optimal, still inconsistent with the model. Remember that the optimal combined reliability is by definition the maximum value achievable by a linear combination rule. A higher combined reliability would suggest that the observer is not consistent with the linear model upon which we base our analysis. According to the test based on the bootstrap simulation we reject optimality for observers TB, FSC, AD, ML, and RL. For observers IO, JK, and ELA, we fail to reject optimality.

In Fig. 2, we illustrated how correlation between cues can reduce the optimal achievable reliability of an observer who combines cues by a weighted linear rule. When we find that combined-cue reliability is lower than the predicted optimal value for an observer, it might be because the observer's choice of combination rule was sub-optimal, but it could also be because the cues are correlated. We consider the possibility that some of the observers are combining cues optimally but that the cues are correlated.

When the weights applied to cues are constrained to be non-negative, the optimal reliability that can be achieved by a weighted linear combination of two correlated cues is always less than the optimal reliability that can be achieved in combining two uncorrelated cues with reliabilities identical to those of the first pair. As we show in Appendix A, this need not be the case if we allow some weights to be negative (as the weights are constrained to sum to 1, not all can be negative). In the remainder of this section, we constrain weights to be non-negative, postponing discussion of the consequences of allowing negative weights to the conclusion. To anticipate the outcome, we find no evidence either in the literature or in our experimental results that supports the claim that observers use negative weights in combining cues.

[2] The index is the geometric mean of the ratios of the reliability of the human observer in each combined-cue condition to the maximum possible reliability predicted from the corresponding single-cue conditions. If we translate reliabilities to variances, the resulting ratios of variances are the standard definition of the relative reliability of estimators employed in mathematical statistics (Freund, 1992, p. 361).

[3] For this and all subsequent tests, the null hypothesis is rejected if $p < 0.05$. The results of all the tests we performed (along with their null hypotheses, acceptance regions, test statistics and results) are given in Table 1.

Table 1
Classification of observers

Observers	Optimality test H_0: optimal uncorrelated			Correlated optimality test H_0: optimal correlated			Benefit test H_0: no benefit			Linearity test H_0: weighted linear		
	Ω	Acceptance region	Result	Ω_{corr}	Acceptance region	Result	S_1	Upper bound of the acceptance region	Result	S_2	Lower bound of the acceptance region	Result
IO	86.54	[76.04, 128.24]	Accept									
ELA	106.73	[76.93, 128.54]	Accept									
JK	99.73	[75.50, 129.97]	Accept									
FSC	66.68	[76.60, 127.61]	Reject	66.68	[57.99, 96.51]	Accept						
AD	62.89	[76.03, 128.88]	Reject	62.89	[52.24, 89.57]	Accept						
RL	61.31	[75.45, 129.37]	Reject	61.31	[56.78, 96.67]	Accept						
ML	32.40	[76.38, 130.04]	Reject	32.40	[50.34, 84.21]	Reject	49.83	126.12	Accept	93.68	79.27	Accept
TB	46.27	[76.37, 128.78]	Reject	46.27	[47.61, 81.50]	Reject	73.71	128.87	Accept	127.26	77.29	Accept

Four statistical tests are summarized. All tests are performed in the order seen in the Table from left to right. Therefore, observers who pass a given test are not tested for the remaining tests. First, optimality with uncorrelated cues is tested for all observers. The results are seen in the left-most block. Three observers pass this test. For the rest of the observers we test a slightly different hypothesis where we allow for non-zero correlation among cues. Three more observers pass this test. Thus, six observers are judged to be optimal. Passing the tests for optimality already means that these observers were consistent with the linear combination model and that they did benefit from combining cues. The only two observers who were sub-optimal, ML and TB, do not pass the test for benefiting from cue combination. The last test seen at the right-most block confirms that they are consistent with a linear combination model. All tests are performed at the 0.05 level.

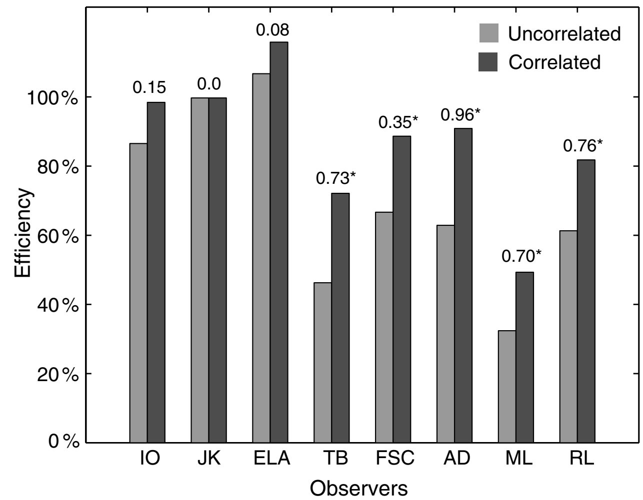

Fig. 10. *Efficiency indices: correlated and uncorrelated cases.* The efficiency indices for each of the eight observers are plotted. For each observer, the bar on the left is the efficiency index Ω with the cues assumed to be uncorrelated. The bar on the right is the revised optimality index allowing for a possible non-zero cue correlation. The estimate of cue correlation is shown and, if marked with an asterisk, it is significantly different from 0.

We fit the observers' data allowing for the possibility that there may be a non-zero correlation between cues. We assumed that this correlation was the same in all experimental conditions for a given observer, but that different observers might have different correlations. In Fig. 10, we plot each observer's measure of efficiency with correlation considered next to the corresponding measure of efficiency in the uncorrelated case (from Fig. 9). The corresponding maximum likelihood estimates of the correlation values are given above each bar.

We performed a bootstrap simulation to determine the significance of the increase in efficiency due to allowing for a non-zero correlation. For each human observer we simulated an ideal observer with zero correlation and the same single-cue reliabilities as the human, and computed 1000 samples of Ω for this ideal observer. Also, for each simulated data set, we allowed for non-zero correlation, and computed its maximum likelihood estimate. Then, we computed the efficiency again, this time taking into account our estimate of correlation. The difference between the two efficiencies is the "spurious" increase in efficiency of allowing non-zero correlation (it is a "spurious" increase because we already know that this ideal observer has zero correlation). Thus, we formed a histogram of increase in efficiency for an ideal zero correlation observer. If the human observer's increase in efficiency is above the 95th percentile of this histogram, then it is judged to be a significant increase, as marked with an asterisk in Fig. 10. We see that the optimal observers IO, JK and ELA did not show a significant increase in efficiency from allowing a non-zero correlation between the two cues.

Efficiencies for the rest of the observers were significantly increased, indicating that their data were consistent with combining correlated cues. Indeed, a test based on bootstrap as before shows that for observers FSC, AD and RL we cannot reject optimality. Still, for observers TB and ML, allowing for correlated cues does not completely account for the sub-optimality observed in their results, as the test rejects optimality for these observers.

Even if some observers are not combining cues optimally, we are still interested in whether they *benefit* from having two cues. By *benefit* we mean that they achieve a combined reliability higher than both single-cue reliabilities. This would rule out the possibility of a "cue-switching" strategy in which the observers use only one of the two cues available at each combined-cue trial, and switch from one cue to the other on different trials. We would be able to say that the observers are actually combining cues, if we can show they are benefiting from having two cues. We need to check whether the combined reliability is higher than that of the more reliable single cue or, in other words, if the combined reliability (solid circle) is in region C of Fig. 3.

To quantify how much an observer benefits from combining cues we define a statistic

$$S_1 = 100 \times \left(\prod_{k=1}^{4} \frac{r_{12,k}}{\max\{r_{1,k}, r_{2,k}\}} \right)^{1/4}. \tag{10}$$

To test whether the S_1 values for observers TB and ML are consistent with a benefit from combining cues, we performed a bootstrap simulation of a worst-case observer that cannot combine cues, but uses the more reliable cue on combined-cue trials. We computed 1000 samples of S_1 for this worst-case observer. The rejection region is the region above the 95th percentile of the histogram formed by the 1000 samples of S_1. With this test, neither observer TB nor ML is significantly more reliable than this worst-case observer. The evidence is insufficient to prove that they benefit from cue combination.

We have established that observers TB and ML do not combine cues optimally. Furthermore, they do not benefit from having two cues. Last, we would like to test, for these observers, whether the combined reliabilities are consistent with a linear combination rule. For this, we need to check whether the combined reliabilities are within the range achievable by a weighted average (regions B and C in Fig. 3). We define a statistic

$$S_2 = 100 \times \left(\prod_{k=1}^{4} \frac{r_{12,k}}{\min\{r_{1,k}, r_{2,k}\}} \right)^{1/4}. \tag{11}$$

To test whether the S_2 values for observers TB and ML are consistent with a linear combination rule, we performed a bootstrap simulation of another worst-case observer that uses the less reliable cue on combined-cue trials. Note that at this point we have already ruled out the possibility of super-optimality for these two observers. We only need to check whether the combined reliabilities are lower than that possible with a weighted average. The lowest combined reliability achievable by a weighted average is the reliability of the less reliable cue (i.e., the observer ignores the better cue). Any lower than that would suggest that adding a better cue in the scene is hurting combined reliability. Again, we computed 1000 samples of S_2 for this worst-case observer. The rejection region is the region below the 5th percentile of the histogram formed by the 1000 samples of S_2. With this test, we cannot reject the hypothesis that observers TB and ML are consistent with a linear combination rule.

5. Conclusion

The performance of all eight observers is consistent with a linear cue combination model. We remind the reader that the minimum-variance rule of combination need not be the linear rule of combination when the distributions of cues are not Gaussian (Appendix B). The observers' performance in the combined-cue conditions is neither better nor worse than predicted by a linear combination rule.

The performance of six of the eight observers could not be discriminated from optimal. For three of these optimal observers (IO, JK, ELA), we did not have to assume any degree of correlation between internal cue estimates. For the other optimal observers (FSC, AD, RL), we rejected the hypothesis that correlation was zero. Observer RL's data were peculiar, however, as the degree of optimality varied greatly between conditions. The remaining sub-optimal observers (TB, ML) passed only the test for linearity. These are substantial individual differences although, as discussed earlier, large individual differences are the rule in depth perception studies such as this.

In the analysis and discussion above, we assumed that observers would assign only non-negative weights to cues. When the cues are uncorrelated, or if correlation is negative, the optimal weights are always non-negative (Eq. (5)). For two positively correlated cues, however, the choice of weights that minimize variance is (Appendix A), $w_i \propto r_i - \rho_{12}\sqrt{r_1 r_2}$. The optimal choice of weight w_1 will be negative whenever $\rho_{12} > \sqrt{r_1/r_2}$. The optimal choice of weight w_2 will be negative whenever, $\rho_{12} > \sqrt{r_2/r_1}$. At most one of the weights can be negative (they must sum to 1) and the negative weight must correspond to the cue with the lower reliability. Intuitively, if the cues are highly correlated, the cue with lower reliability can be used to cancel a portion of the noise in the more reliable cue (by giving the less reliable cue a negative weight). We have established the following proposition: The optimal weights will both be non-negative if and only if

$$\rho_{12} \leqslant \min\{\sqrt{r_2/r_1},\ \sqrt{r_1/r_2}\}. \tag{12}$$

Even when there is a positive correlation between the cues, the optimal choice of weights may be all non-negative (see, e.g., Fig. 2). When $r_1 = r_2$, for example, the above condition is always satisfied (correlation cannot exceed 1), and the weights will be non-negative.

Moreover, the reliability of the optimal weighted linear combination (from Eq. (7)) with two correlated cues can exceed the reliability of the optimal combination of two uncorrelated cues with the same reliabilities. It is easy to show that when the correlation ρ is positive,

$$r_1 + r_2 < \frac{r_1 + r_2 - 2\rho\sqrt{r_1 r_2}}{1 - \rho^2} \quad \text{if and only if} \quad \rho > 2\sqrt{r_1 r_2}/(r_1 + r_2). \tag{13}$$

The left-hand side of the first inequality is the maximum possible reliability of a weighted linear combination of two uncorrelated cues with reliabilities r_1 and r_2, while the right-hand side is the maximum possible reliability if the same cues are correlated with correlation ρ. To illustrate, we have plotted the reliability (Fig. 11) of a weighted linear combination of two cues versus the weight of the less reliable cue with $r_1 = 0.5$ and $r_1 = 2$ for the case where $\rho = 0$ and the case where $\rho = 0.9$. Note that $\rho > \sqrt{r_1/r_2}$, so we expect that the optimal weight for the less reliable cue to be negative. Further, $\rho > 2\sqrt{r_1 r_2}/(r_1 + r_2)$, so we expected that the maximum reliability with the correlated cues will exceed the reliability with the uncorrelated cues. The plots in Fig. 11 bear out both predictions.

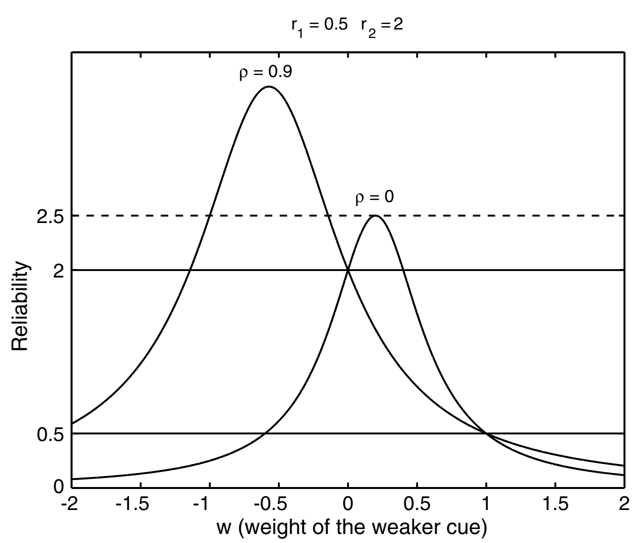

Fig. 11. *High correlations and negative weights*. For large positive correlations, the weights that maximize the reliability of the weighted linear may not fall between 0 and 1. The weights still sum to 1 but one is negative, the other greater than 1. This figure is a plot of reliability versus the weight of the weaker cue when the cue reliabilities are $r_1 = 0.5$ and $r_2 = 2$. Two curves are plotted, one for the uncorrelated case ($\rho = 0$) and one for the case where $\rho = 0.9$. In the uncorrelated case, maximum reliability occurs when $w = 0.2$ and the maximum achieved is the sum of the reliabilities of the two individual cues (i.e., 2.5). In the correlated case, the maximum occurs at $w_1 = -0.57$. Note that the maximum achieved $r_c = 3.68$ is greater than the sum of the reliabilities of the two cues (2.5).

The implication is that we may encounter negative weights in cue combination studies when correlated cues are combined optimally, at least when correlations are large and the single-cue reliabilities differ considerably. In effect, the observer uses a less reliable cue to cancel "noise" from a more reliable cue when the cues are correlated. We know of no evidence for negative weights in studies that explicitly measure the weights assigned to individual cues. We have based our analysis on the assumption that observers are not able to use negative weights and it is of interest to inquire whether allowing for the possibility of negative weights would alter our conclusions in any respect.

For the data collected in this experiment, relaxing the constraint that weights be non-negative has very little effect on results of our analysis. Here is why: the optimal weights for combining uncorrelated cues are always non-negative (Eq. (2)). Therefore, our optimality tests for uncorrelated cues are not affected. An optimal uncorrelated observer has to use non-negative weights.

Only three observers failed the condition in Eq. (12) for at least one of the combined-cue conditions: RL, AD, and ML. For the remainder, the optimal choices of weights are still non-negative even when the possibility of correlation is taken into account. Of these three observers, observer RL's data already exhibited internal inconsistency. Observer ML did not pass the test for correlated optimality: his reliability in the combined-cue conditions is simply too low to count as optimal. Allowing for negative weights can only increase the criterion value (optimal reliability), making the test harder to pass, and therefore the outcome would be the same. Similarly, for observer AD, allowing for negative weights will make the test harder to pass; but since this observer has passed this test we performed the bootstrap simulation again, to check. When we allow for negative weights, the acceptance region for Ω for correlated optimal combination is [112.30, 194.53]. Remember that Ω for AD is 62.89. Thus if we were to assume that observers are able to use negative weights we reject optimality for observer AD. This is the only change that would be caused by allowing for negative weights: AD would no longer be judged to be optimal.

The benefits test simply checks whether the reliabilities in the combined-cue conditions are larger than that of the more reliable single-cue conditions. Thus, this test's outcome has no relation to the particular weights that were used.

The linearity test was performed for two observers, ML and TB, and only for the lower bound. Allowing for negative weights will only relax the lower bound for this test. Since both observers passed this test, the results would be unchanged.

In summary, we raise the intriguing possibility that human observers may employ negative weights in some cue combination tasks, using a less reliable cue to cancel "noise" from a more reliable cue when the cues are highly correlated. However, in our particular case, allowing for this possibility has very little effect on our results and conclusions and we find no evidence that weights are ever negative.

In this study we have found evidence confirming that depth cue combination is linear. We have introduced the concept of correlated cues. We established that under certain conditions, apparently sub-optimal cue combination performance might, in fact, be explained by the cues being non-independent. We found that the degree of cue correlation as well as the optimality of cue combination varied from subject to subject. It remains to be seen whether such cue correlations or sub-optimal behavior might be eliminated with practice.

Acknowledgements

The research reported here was supported by Grant EY08266 from the National Institute of Health, National Eye Institute, and Grant RG0109/1999-B from the Human Frontiers Science Program.

Appendix A. Minimum variance unbiased linear estimation

In this and the following appendices, we have gathered results from the mathematical statistics literature relevant to cue combination that are usually stated without proof, the proof being left to the reader. The results concerning correlated cues in this appendix are based on Keller and Olkin (2002, p. 5). We make no assumptions [4] about the distributions of the random variables that represent estimates from different cues. In particular, we do not assume that the variables are Gaussian.

Notation. The identity matrix will be denoted I and the transpose of any vector or matrix is V is denoted V'. Let $S = [S_1, \dots, S_n]'$ be a column vector of random variables, $S_1, \dots, S_n$. These random variables are assumed to share a common *expected value*, $E(S_i) = s$, but can have distinct *variances*, $\mathrm{Var}(S_i) = \sigma_i^2 > 0$. The correlation between S_i and S_j is denoted ρ_{ij} and the correlation matrix is the matrix R whose ijth entry is ρ_{ij}. The *covariance* of S_i and S_j is $\rho_{ij}\sigma_i\sigma_j$ and the covariance matrix of $S_1, \dots, S_n$ is

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & & \rho_{2n}\sigma_2\sigma_n \\ \vdots & & \ddots & \vdots \\ \rho_{1n}\sigma_1\sigma_n & \cdots & & \sigma_n^2 \end{bmatrix}. \tag{A.1}$$

Both the correlation matrix R and the covariance matrix Σ are symmetric (as $\rho_{ij} = \rho_{ji}$). We assume throughout that Σ is invertible, i.e. that there really are n uncorrelated cues.

When the random variables $S = [S_1, \dots, S_n]'$ are uncorrelated ($\rho_{ij} = 0$ whenever $i \neq j$), the correlation matrix is the identity matrix, and the covariance matrix is the diagonal matrix $\mathrm{Diag}(\sigma_1^2, \dots, \sigma_n^2)$.

We wish to estimate s given $S_1, \dots, S_n$ using a weighted linear combination rule. Given a vector of non-negative weights $w = [w_1, \dots, w_n]'$ with $\sum_{i=1}^n w_i = 1$, the weighted linear combination,

$$F_w(S) = w'S = \sum_{i=1}^n w_i S_i \tag{A.2}$$

is an unbiased estimate of s:

$$E(F_w(S)) = \sum_{i=1}^n w_i E(S_i) = \sum_{i=1}^n w_i s = s. \tag{A.3}$$

The variance of the estimate is

$$\mathrm{Var}(F_w(S)) = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \rho_{ij}\sigma_i\sigma_j, \tag{A.4}$$

or, in matrix notation,

$$\mathrm{Var}(F_w(S)) = w'\Sigma w. \tag{A.5}$$

We wish to determine the choice of weights $w = [w_1, \dots, w_n]'$ that minimizes the variance above. We will first consider the uncorrelated case.

Proposition A.1. *When the random variables are uncorrelated, then the variance of the weighted linear combination is*

$$\mathrm{Var}(F_w(S)) = w'\,\mathrm{Diag}(\sigma_1^2, \dots, \sigma_n^2)w = \sum_{i=1}^n w_i^2\sigma_i^2. \tag{A.6}$$

The variance is minimized when the weights are $w_i = \sigma_i^{-2} / \sum_{j=1}^n \sigma_j^{-2} = r_i / \sum_{j=1}^n r_j$ *(using the alternative parameterization,* $r_i = \sigma_i^{-2}$ *described in the main text).*

Proof. The earliest reference for this result that we have found is Cochran (1937) who proved it for the case where the random variables $S = [S_1, \dots, S_n]'$ are Gaussian. The Cauchy–Schwarz inequality (Hardy, Littlewood, & Pólya, 1952 , p. 16) states that, for any real numbers $a_1, \dots, a_n$ and $b_1, \dots, b_n$,

[4] Other than that they have finite means and variances.

$$\sum_{i=1}^{n} a_i^2 \sum_{i=1}^{n} b_i^2 \geqslant \left(\sum_{i=1}^{n} a_i b_i \right)^2. \tag{A.7}$$

Further, the inequality is an equality if and only if the $a_1, \ldots, a_n$ and the $b_1, \ldots, b_n$ are proportional. More precisely, equality holds only if the $b_1, \ldots, b_n$ are all zero or there is a constant c with $a_i = cb_i$ for $i = 1, 2, \ldots, n$.

Letting $a_i = w_i \sigma_i$ and $b_i = \sigma_i^{-1}$ in the Cauchy–Schwarz inequality, we have

$$\sum_{i=1}^{n} w_i^2 \sigma_i^2 \sum_{i=1}^{n} \sigma_i^{-2} \geqslant \left(\sum_{i=1}^{n} w_i \right)^2 = 1. \tag{A.8}$$

The first term on the left-hand side is the variance we wish to minimize and the other terms do not depend on the choice of weights. The variance is evidently minimized when the inequality is an equality, which (as the $b_i = \sigma_i^{-1}$ cannot be 0) occurs precisely when the $a_i = w_i \sigma_i$ are equal to a constant times the $b_i = \sigma_i^{-1}$. This occurs when $w_i \propto \sigma_i^{-2}$. The constant of proportionality is determined by the constraint $\sum_{i=1}^{n} w_i = 1$, giving

$$w_i = \sigma_i^{-2} \bigg/ \sum_{j=1}^{n} \sigma_j^{-2} = r_i \bigg/ \sum_{j=1}^{n} r_j, \tag{A.9}$$

which is what we set out to prove. □

We next wish to determine the variance-minimizing weights when the random variables $S = [S_1, \ldots, S_n]'$ may be correlated. The variance we wish to minimize is still $w'\Sigma w$. We will first determine the solution in the following proposition and then show how it can be simplified.

Proposition A.2. *Consider the case where the correlations between the random variables $S = [S_1, \ldots, S_n]'$ may be non-zero. Then the variance-minimizing weights are proportional to the sums of the elements in the corresponding rows of the inverse of the covariance matrix Σ^{-1}:*

$$w \propto \Sigma^{-1} e, \tag{A.10}$$

where $e = [1, 1, \ldots, 1]'$.

Example. Before proving the proposition, we illustrate it for the case $n = 2$, the case considered in the main text. We can readily calculate the inverse of the covariance matrix by cofactors (described below) or Cramer's rule (Strang, 1988, p. 231ff); it is

$$\Sigma^{-1} = \frac{1}{1 - \rho_{12}^2} \begin{bmatrix} \sigma_1^{-2} & -\rho_{12}\sigma_1^{-1}\sigma_2^{-1} \\ -\rho_{12}\sigma_1^{-1}\sigma_2^{-1} & \sigma_2^{-2} \end{bmatrix}. \tag{A.11}$$

If we change parameterization to reliabilities and sum the rows (ignoring the constant in front of the matrix), we have

$$w_i \propto r_i - \rho_{12}\sqrt{r_1 r_2} \tag{A.12}$$

as in the main text (where ρ_{12} is denoted ρ). We now prove this result for all n. As discussed in the main text, we can think of the weights in the correlated case as proportional to corrected reliabilities, $r_i' = r_i - \rho_{12}\sqrt{r_1 r_2}$, corrected for the correlational structure in the random variables $S = [S_1, \ldots, S_n]'$. In Proposition A.3 we will show that we can extend this way of thinking to the general case. But first, the proof of Proposition A.2.

Proof. The covariance matrix Σ is symmetric and positive definite (Strang, 1988, Chap. 6). By the spectral decomposition theorem, it can be factored into the form $V'DV$ where $D = \mathrm{Diag}(\tau_1^2, \ldots, \tau_n^2)$ and V is an orthogonal matrix satisfying $V'V = I$ (Mardia, Kent, & Bibby, 1979, p. 469).

The inverse of the covariance matrix is $V'D^{-1}V$ where $D^{-1} = \mathrm{Diag}(\tau_1^{-2}, \ldots, \tau_n^{-2})$ is the inverse of the diagonal matrix D. If we let $v = Vw$, then we can write the variance of the weighted linear combination in terms of v,

$$w'\Sigma w = w'V'DVw = v'Dv \tag{A.13}$$

and we seek to minimize $v'Dv = \sum_{i=1}^{n} v_i^2 \tau_i^2$ subject to the constraint $\sum_{i=1}^{n} w_i = 1$. Define $z = Ve$. As in the uncorrelated case, we let $a_i = \tau_i v_i$ and $b_i = \tau_i^{-1} z_i$ and write out the Cauchy–Schwarz inequality,

$$\sum_{i=1}^{n} v_i^2 \tau_i^2 \sum_{i=1}^{n} z_i^2 \tau_i^{-2} \geqslant \left(\sum_{i=1}^{n} v_i \tau_i z_i \tau_i^{-1} \right)^2 = (v'z)^2$$
$$= (w'V'Ve)^2 = (w'e)^2 = 1, \tag{A.14}$$

where we use the fact that $V'V = I$. As in the uncorrelated case, the variance is minimized when $v_i \tau_i \propto z_i \tau_i^{-1}$ or equivalently, $v_i \propto z_i \tau_i^{-2}$. In vector notation, the condition for minimum variance is that $v = cD^{-1}z$ for some constant c. Substituting for v and z,

$$Vw = cD^{-1}Ve \tag{A.15}$$

and, multiplying both sides of the above by the non-singular matrix V',

$$w = cV'D^{-1}Ve = c\Sigma^{-1}e, \tag{A.16}$$

the result we wished to prove. The variance-minimizing weights are proportional to the sums of the rows of the inverse of the covariance matrix, Σ^{-1}. □

We now wish to characterize the weights $c\Sigma^{-1}e$ and, to do so, we factor the covariance matrix as

$$\Sigma = \mathrm{Diag}(\sigma_1, \ldots, \sigma_n) R \, \mathrm{Diag}(\sigma_1, \ldots, \sigma_n), \tag{A.17}$$

where $R = (\rho_{ij})$ is the correlation matrix. The inverse covariance matrix Σ^{-1} can be expressed as

$$\Sigma^{-1} = \mathrm{Diag}(\sigma_1^{-1}, \ldots, \sigma_n^{-1}) R^{-1} \, \mathrm{Diag}(\sigma_1^{-1}, \ldots, \sigma_n^{-1}). \tag{A.18}$$

Let the elements of R^{-1} be denoted δ_{ij}. The first row of Σ^{-1} can then be written as (switching to reliabilities r_i)

$$\begin{bmatrix} \delta_{11}r_1 & \delta_{12}\sqrt{r_1r_2} & \cdots & \delta_{1n}\sqrt{r_1r_n} \end{bmatrix} \tag{A.19}$$

and

$$w_1 \propto \delta_{11}r_1 + \sum_{i=2}^{n} \delta_{1i}\sqrt{r_1r_i}. \tag{A.20}$$

The jth weight is

$$w_j \propto \delta_{jj}r_j + \sum_{i\neq j} \delta_{ij}\sqrt{r_ir_j} \tag{A.21}$$

and if we define the right-hand side of the above as the reliability corrected for correlation, then we find that the weights are proportional to the corrected reliability.

The δ_{ij} are readily computed by cofactors (Strang, 1988, pp. 231ff) as

$$\delta_{ij} = (-1)^{i+j}|R(i,j)|/|R|, \tag{A.22}$$

where $R(i,j)$ is the correlation matrix R with the ith row and jth column removed and || denotes the determinant. For example, when

$$R = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}, \tag{A.23}$$

the inverse of the correlation matrix is

$$(\delta_{ij}) = \begin{bmatrix} 1-\rho_{23}^2 & \rho_{13}\rho_{23}-\rho_{12} & \rho_{12}\rho_{23}-\rho_{13} \\ \rho_{13}\rho_{23}-\rho_{12} & 1-\rho_{13}^2 & \rho_{12}\rho_{13}-\rho_{23} \\ \rho_{12}\rho_{23}-\rho_{13} & \rho_{12}\rho_{13}-\rho_{23} & 1-\rho_{12}^2 \end{bmatrix} \times \frac{1}{|R|} \tag{A.24}$$

and

$$w_1 \propto (1-\rho_{23}^2)r_1 + (\rho_{13}\rho_{23}-\rho_{12})\sqrt{r_1r_2} + (\rho_{12}\rho_{23}-\rho_{13})\sqrt{r_1r_3}, \tag{A.25}$$

where the reciprocal of the determinant $|R|$ has been absorbed into the constant of proportionality.

Substituting Eq. (A.22) into Eq. (A.21), we have

Proposition A.3. *The variance-minimizing weights are, with the notation above, defined by*

$$w_j \propto |R(j,j)|r_j + \sum_{i\neq j}(-1)^{i+j}|R(i,j)|\sqrt{r_ir_j}. \tag{A.26}$$

If we define the corrected reliability of the ith cue to be the expression on the right-hand side, then the optimal weights are proportional to reliability corrected for correlational structure.

The minimum variance achieved by the optimal rule can be computed by substituting the expression for the weights, $w \propto \Sigma^{-1}e$ into the expression for the variance, $w'\Sigma w$. First, we need to determine the constant of proportionality in the expression, $w \propto \Sigma^{-1}e$. But this is just the inverse of the sum of the sum of the rows: $(e'\Sigma^{-1}e)^{-1}$. Thus,

$$\begin{aligned}\mathrm{Var}(F_w(S)) &= e'\Sigma^{-1}\Sigma\Sigma^{-1}e(e'\Sigma^{-1}e)^{-2} \\ &= (e'\Sigma^{-1}e)^{-1}\end{aligned} \tag{A.27}$$

or, in terms of reliability,

$$r = e'\Sigma^{-1}e. \tag{A.28}$$

In the uncorrelated case, $\Sigma^{-1} = \mathrm{Diag}(r_1,\ldots,r_n)$, and we can verify once again that $r = r_1 + r_2 + \cdots + r_n$.

Appendix B. Other approaches

B.1. Non-linear rules

As in the previous appendix, let $S = [S_1,\ldots,S_n]'$ be a column vector of random variables, $S_1,\ldots,S_n$, with a common *expected value*, $E(S_i) = s$, and *variances*, $\mathrm{Var}(S_i) = \sigma_i^2 > 0$. The correlation between S_i and S_j is denoted ρ_{ij} and the correlation matrix is the matrix R whose ijth entry is ρ_{ij}. The *covariance matrix* Σ is the matrix whose ijth entry is $\rho_{ij}\sigma_i\sigma_j$. In the previous appendix, we determined the minimum-variance weighted linear combination rule of $S = [S_1,\ldots,S_n]'$ subject to the constraint that the weights be positive and sum to 1. The constraint on the weights guarantees the expected value of the weighted linear cue combination to be s, i.e. that it is unbiased.

We might wonder, is there some non-linear combination rule that is unbiased and that has even lower variance than the optimal linear rule? The answer, in general, is that there can be. An example of a non-linear cue combination rule that is unbiased and that has lower variance than the optimal linear rule is given in Freund (1992, p. 368, Problem 10.30). However, when the $S = [S_1,\ldots,S_n]'$ are Gaussian, the linear combination rule derived in the previous appendix is the minimum variance unbiased estimator of the common mean s (Freund, 1992, pp. 360–361). There are other choices of distributions for which a weighted linear rule has minimum variance as well (Mood, Graybill, & Boes, 1974, pp. 318–319).

B.2. Bayesian approaches

There has been a great deal of work recently applying Bayesian decision-making models to perceptual behavior (Mamassian & Landy, 1998, 2001; Mamassian, Landy, & Maloney, 2002; Trommershäuser, Maloney, & Landy, 2003a, 2003b; see Knill & Richards, 1996; Maloney, 2002; Rao, Olshausen, & Lewicki, 2002). The Bayesian approach begins with the assumption that the true value of slant s is the realization of a random variable that has a known prior distribution, $P(s)$. The Bayesian observer uses knowledge of that distribution to improve the estimate of slant based on sensory data alone. Given two cues, S_1 and S_2, the Bayesian observer chooses as the slant estimate S that value of slant that maximizes the a posteriori (MAP) probability,

$$P(s|S_1, S_2) \propto P(S_1, S_2|s)P(s). \tag{B.1}$$

If the cues are conditionally independent, [5] this simplifies to

$$P(s|S_1, S_2) \propto P(S_1|s)P(S_2|s)P(s). \tag{B.2}$$

Eq. (B.2) is the Bayesian analogue of the uncorrelated case treated in the previous appendix, with the difference that we cannot reduce Eq. (B.2) to an estimate without explicit knowledge of the distributional forms of the *likelihood functions* $P(S_i|s)$ and the *prior distribution* $P(s)$. If all are Gaussian in form, the MAP estimate is identical to the minimum variance estimate of Eqs. (1) and (2) with the mean of the prior effectively treated as an additional non-sensory cue weighted in proportion to its own reliability. To prove this claim, we need the following result:

Proposition B.1. *The product of the n probability density functions (pdf) of n Gaussian variables with respective means* $\mu_1, \ldots, \mu_n$ *and reliabilities* $r_1, \ldots, r_n$ *is proportional to a Gaussian probability density function with mean* $\mu = \sum_{i=1}^{n} w_i\mu_i$ *with* $w_i = r_i / \sum_{j=1}^{n} r_j$, *and reliability* $r = \sum_{i=1}^{n} r_i$.

Proof. Written in terms of reliability, a Gaussian pdf with mean μ_i and reliability r_i is of the form $\phi(s; \mu_i, r_i) = \sqrt{r_i/2\pi}\exp(-\frac{1}{2}r_i(s-\mu_i)^2)$. The product of n of them is

$$\prod_{i=1}^{n} p(s; \mu_i, r_i) = C\exp\left[-\frac{1}{2}r(s^2 - 2\mu s)\right], \tag{B.3}$$

where C does not depend on s. Completing the square in the exponent,

$$\begin{aligned}\prod_{i=1}^{n} p(s; \mu_i, r_i) &= C\exp\left[-\frac{1}{2}r(s^2 - 2\mu s + \mu^2)\right]\exp\left[\frac{1}{2}r\mu^2\right] \\ &= C'\exp\left[-\frac{1}{2}r(s-\mu)^2\right]\end{aligned} \tag{B.4}$$

and, as the last term is proportional to a Gaussian pdf with mean μ and reliability r, we have proven the proposition. □

When the likelihood functions in Eq. (B.2) are Gaussian, we can write them in the form $P(S_i|s) = \phi(s; S_i, r_i)$ for $i = 1, 2$. When the prior is also Gaussian, we can write it in the form $P(s) = \phi(s; S_0, r_0)$. Then, the posterior distribution in Eq. (B.2) is the product of three Gaussian pdfs and, applying the proposition above, the MAP is,

$$S = w_0S_0 + w_1S_1 + w_2S_2, \tag{B.5}$$

where the weights sum to 1 and $w_i \propto r_i$. The prior is, in effect, a cue with a fixed value S_0 and reliability r_0. While S_1 and S_2 are random variables that, by assumption, have expected values s (the true slant), S_0 is not a random variable and its expected value is not, in general, s. The expected value of the MAP estimator in Eq. (B.5) is

$$E\left[\sum_{i=0}^{2} w_iS_i\right] = s + w_0(S_0 - s), \tag{B.6}$$

and the estimator is biased toward S_0 in any particular scene unless the prior mean is exactly the true value s in that scene. This bias represents an intentional trade-off between the non-sensory information in the prior and the unbiased sensory information available from the cues (Maloney, 2002) and the magnitude of the bias is proportional to $w_0 = r_0/(r_0 + r_1 + r_2)$, the relative magnitude of the prior reliability r_0 and the reliability $r_1 + r_2$ of the sensory cues.

Finally, if the individual cues have Gaussian distributions but are not conditionally independent, there is a Bayesian analogue to the treatment of the correlated case in the previous appendix, treating S_0 as an equivalent, independent cue. The covariance matrix, Σ, is augmented by one column and one row corresponding to the prior. The derivation of the MAP estimator for this case is then identical to the derivation in the previous appendix.

[5] Yuille and Bülthoff (1996) consider the case where the posterior distribution for two cues can be expressed as the product of the posterior distributions for the individual cues: $P(s|S_1, S_2) \propto P(s|S_1)P(s|S_2)$ and state that, in the Gaussian case, the MAP is a weighted linear combination.

References

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation, 1*, 295–311.

Barlow, H. B., & Földiák, P. (1989). Adaptation and decorrelation in the cortex. In C. Miall, R. M. Durbin, & G. J. Mitchison (Eds.), *The computing neuron* (pp. 54–72). Wokingham, England: Addison-Wesley.

Clark, J. J., & Yuille, A. L. (1990). *Data fusion for sensory information processing systems*. Boston, MA: Kluwer.

Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society, 4*(Suppl.), 102–118.

Cohen, A., & Sackrowitz, H. B. (1974). On estimating the common mean of two normal distributions. *Annals of Statistics, 2*, 1274–1282.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman Hall.

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

Freund, J. E. (1992). *Mathematical statistics* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology, 13*, 483–488.

Ghahramani, Z., Wolpert, D. M., & Jordan, M. I. (1997). Computational models of sensorimotor integration. In P. G. Morasso, & V. Sanguineti (Eds.), *Self-organization, computational maps and motor control* (pp. 117–147). Amsterdam: North-Holland.

Graybill, F. A., & Deal, R. B. (1959). Combining unbiased estimators. *Biometrics, 15*, 543–550.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. New York: Wiley.

Hardy, G. H., Littlewood, J. E., & Pólya, G. (1952). *Inequalities*. Cambridge, UK: Cambridge University Press.

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.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research, 39*, 3621–3629.

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

Kaufman, L. (1974). *Sight and mind*. London: Oxford University Press.

Keller, T., & Olkin, I. (2002). Combining correlated unbiased estimators of the mean of a normal distribution. *Technical Report No. 2002-5*. Department of Statistics, Stanford University.

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

Knill, D. C., & Saunders, J. (2002). Humans optimally weight stereo and texture cues to estimate surface slant [abstract]. *Journal of Vision, 2*(7), 400.

Kontsevich, L. L. (1998). Defaults in stereoscopic and kinetic depth perception. *Proceedings of the Royal Society of London B, 265*, 1615–1621.

Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of the Optical Society of America A, 18*, 2307–2320.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. J. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research, 35*, 389–412.

Maloney, L. T. (2002). Statistical decision theory and biological vision. In D. Heyer, & R. Mausfeld (Eds.), *Perception and the physical world* (pp. 145–189). New York: Wiley.

Mamassian, P., & Landy, M. S. (1998). Observer biases in the 3D interpretation of line drawings. *Vision Research, 38*, 2817–2832.

Mamassian, P., & Landy, M. S. (2001). Interaction of visual prior constraints. *Vision Research, 41*, 2653–2688.

Mamassian, P., Landy, M. S., & Maloney, L. T. (2002). Bayesian modeling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of perception and brain function* (pp. 13–36). Cambridge, MA: MIT Press.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. New York: Academic Press.

Mood, A., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.

Rao, R. P. M., Olshausen, B. A., & Lewicki, M. S. (Eds.). (2002). *Probabilistic models of perception and brain function*. Cambridge, MA: MIT Press.

Strang, G. (1988). *Linear algebra and its applications* (3rd ed.). San Diego: Harcourt Brace Jovanovich.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003a). Statistical decision theory and trade-offs in the control of motor response. *Spatial Vision, 16*, 255–275.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003b). Statistical decision theory and the selection of rapid, goal-directed movements. *Journal of the Optical Society of America A, 20*, 1419–1433.

Van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1998). The precision of proprioceptive position sense. *Experimental Brain Research, 122*, 367–377.

Van Beers, R. J., 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.

Watt, R. J., Morgan, M. J., & Ward, R. M. (1983). Stimulus features that determine the visual location of a bright bar. *Investigative Ophthalmology & Vision Science, 24*, 66–71.

Young, M. J., 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.

Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill, & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123–161). Cambridge, England: Cambridge University Press.

Zacks, S. (1966). Unbiased estimation of the common mean of two normal distributions based on small samples of equal size. *Journal of the American Statistical Association, 61*, 467–476.