Statistically optimal integration of biased sensory estimates

Peter Scarfe

Department of Cognitive, Perceptual and Brain Sciences, University College London, London, UK

Paul B. Hibbard

School of Psychology, University of St. Andrews, St. Andrews, Fife, UK

Experimental investigations of cue combination typically assume that individual cues provide noisy but unbiased sensory information about world properties. However, in numerous instances, including real-world settings, observers systematically misestimate properties of the world from sensory information. Two such instances are the estimation of shape from stereo and motion cues. Bias in single-cue estimates, therefore poses a problem for cue combination if the visual system is to maintain accuracy with respect to the world, particularly because knowledge about the magnitude of bias in individual cues is typically unknown. Here, we show that observers fail to take account of the magnitude of bias in each cue during combination and instead combine cues in proportion to their reliability so as to increase the precision of the combined-cue estimate. This suggests that observers were unaware of the bias in their sensory estimates. Our analysis of cue combination shows that there is a definable range of circumstances in which combining information from biased cues, rather than vetoing one or other cue, can still be beneficial, by reducing error in the final estimate.

Keywords: 3D surface and shape perception, binocular vision, depth

Citation: Scarfe, P., & Hibbard, P. B. (2011). Statistically optimal integration of biased sensory estimates. *Journal of Vision, 11*(7):12, 1–17, http://www.journalofvision.org/content/11/7/12, doi:10.1167/11.7.12.

Introduction

Combining sensory information

Humans have access to information from multiple sensory modalities when making perceptual estimates about properties of the world. Within a single sensory modality, there are also multiple sources of information that allow us to make perceptual estimates (Hershenson, 1999). The question then arises as to how this information is integrated and combined. The visual system is not a perfect measuring device so all cues are inherently stochastic in nature. One way to combine noisy sensory estimates is described by Bayes' rule (Maloney, 2002; Mamassian, Landy, & Maloney, 2002). This prescribes a way in which the visual system can estimate the most probable state of the world given current and past sensory information (Knill & Richards, 1996; Mamassian et al., 2002).

If we consider estimating the three-dimensional (3D) shape of an object from stereo and motion cues, Bayes' equation can be written as

$$p(\hat{S}|I_S, I_M) \propto p(I_S|\hat{S})p(I_M|\hat{S})p(\hat{S}). \quad (1)$$

Here, the information provided by stereo and motion cues is represented by I_S and I_M, where the likelihood functions for stereo and motion, $p(I_S|\hat{S})$ and $p(I_M|\hat{S})$, represent the generative transfer functions producing this image data. The prior, $p(\hat{S})$, describes the probability of encountering a given shape in the world, independent of sensory data. Given this information, the most likely shape in the world, $\hat{S}$, to have produced this sensory information is given by the maximum of the posterior probability distribution, $p(\hat{S}|I_S, I_M)$.

If the cues are conditionally independent and the prior is uniform or has a much greater variance than the individual cues, the combined-cue estimate of shape, $\hat{S}_C$, can be represented by a simple weighted average of the estimates provided by the individual cues $\hat{S}_S$ and $\hat{S}_M$ (Landy, Maloney, Johnston, & Young, 1995; Oruc, Maloney, & Landy, 2003):

$$\hat{S}_C = w_S\hat{S}_S + w_M\hat{S}_M. \quad (2)$$

The weights for stereo, w_S, and motion, w_M, are determined by the relative reliabilities of the two estimators such that $w_S = \frac{r_S}{r_S+r_M}$ and $w_M = \frac{r_M}{r_S+r_M}$. The reliabilities of the estimates provided by stereo, r_S, and motion, r_M, are given by the reciprocal of their variances, $r_S = \frac{1}{v_S}$ and $r_M = \frac{1}{v_M}$. The variance of the combined-cue estimate, v_C, is given by

$$v_C = \frac{v_S v_M}{v_S + v_M}. \quad (3)$$

doi: 10.1167/11.7.12 Received December 10, 2010; published June 13, 2011 ISSN 1534-7362

This variance is the minimum possible for any linear combination of cues and can also be written as the sum of the reliabilities of the individual cues, $r_C = r_M + r_M$. A number of studies have shown that when combining sensory information, Bayes' rule, and more specifically a weighted average, provides a good account of sensory fusion both within and between modalities (Ernst & Banks, 2002; Helbig & Ernst, 2007; Hillis, Ernst, Banks, & Landy, 2002; Hillis, Watt, Landy, & Banks, 2004; Knill & Saunders, 2003; MacNeilage, Banks, Berger, & Bulthoff, 2007).

Combining and calibrating biased sensory estimators

The cue combination framework described suggests that the optimization criterion adopted by the visual system is one of reducing the variance of the combined-cue estimate. This is a desirable outcome if the cues are well calibrated and, therefore, unbiased, as the combined-cue estimator will also be unbiased. However, this is not necessarily the case if one or more of the cues are biased. We describe "bias" here as the difference between an observer's estimates of a property of the world and its actual physical value. Similarly, "accuracy" is defined as a measure of this bias. Calibration to eliminate this type of bias is what Burge, Girshick, and Banks (2010) have described as maintaining "external accuracy." They contrast this with calibration that maintains "internal consistency." This occurs when cues are calibrated such that they provide the same sensory estimate of a world property, but importantly, this estimate does not necessarily agree with the true state of the world.

In the extreme, the two types of calibration described by Burge et al. (2010) are in fact two sides of the same coin. To maintain external accuracy, the brain needs information about the physical state of the world, but the only information it has about this physical state is that provided by the senses. Because of this, there is no ground-truth estimate by which to calibrate sensory estimates (Ernst & Banks, 2002), so the visual system never has sufficient information for calibration to obtain true external accuracy. Optimizing cue combination and calibration solely in terms of variance (Burge et al., 2010; Hillis et al., 2004) could, therefore, result in perceptual bias. The prevalence of perceptual biases both with multiple-cue simulated stimuli, and in judgements about real-world objects and scenes, is indicative of the nature of this problem (Bradshaw, Parton, & Glennerster, 2000; Todd & Norman, 2003; Wagner, 1985; Watt, Akeley, Ernst, & Banks, 2005).

When measurable conflicts are large, it has been suggested that the visual system might behave in a robust manner and veto biased cues (Landy et al., 1995). This is a sensible strategy to adopt, as all cues are unlikely to be equally biased. The problem then becomes identifying which cues are biased and the relative magnitude of this bias. The cue combination framework described is unable to account for robust behavior because the visual system is modeled as blind to the absolute error of its perceptual estimates (Girshick & Banks, 2009). Despite these problems, the visual system is clearly attuned to the statistics of its environment and, under some circumstances, is seen to act to reduce the bias of its sensory estimates (Adams, Banks, & van Ee, 2001; Burge et al., 2010; Ernst, 2007).

Adams et al. (2001) demonstrated the adaptability of such sensory mappings. Their observers wore a horizontally magnifying prism in front of one eye continuously for 6 days. This systematically changed the horizontal disparity that the observers experienced during their everyday behavior. Perceived slant was tested before, during, and after imposition of the prism. Over the duration that the prism was worn, observers were shown to have remapped the relationship between retinal disparity and perceived slant. Calibration to maintain external accuracy is therefore, at least to some extent, possible. So the important questions become understanding the mechanisms the brain uses to achieve this and what those instances when it clearly fails can tell us about the process (Todd, Christensen, & Guckes, 2010).

Issues related to perceptual accuracy are clearly important for models of motor control. In many instances, it is assumed that accurate metric information is *required* for successful movement (Milner & Goodale, 1995, 2006), but rarely do such models consider how this information might be acquired. One strategy that humans seem to have adopted to control motor acts, such as prehension, is to use relative information continuously over the course of the movement (Saunders & Knill, 2003, 2004, 2005). This strategy is interesting because it actively compensates for those instances where external perceptual accuracy is not possible (Brooks, 1991a, 1991b). It is, thus, not at all obvious that veridical estimation of the metric shapes, sizes, and locations of objects is required for adaptive motor control (Brenner & Smeets, 2001; Smeets & Brenner, 2008). Furthermore, once the action is completed, endpoint errors could also be used for sensory calibration.

Variable and constant errors in cue combination

The current study addresses the combination of stereo and motion information for the estimation of 3D shape. We consider the consequences for accuracy if the visual system were to combine cues using the minimum variance strategy, when in fact one or both of the cues were biased. The first and most straightforward point to make is that, even if cues are biased, combining them in proportion to

their reliability will still result in the least variable combined-cue estimate. However, we now need to consider constant as well as variable error to assess the usefulness of this combination procedure. For cases where bias is present, a natural extension of the idea of minimizing variance is the use of mean squared error (*MSE*):

$$MSE = E[(S_C - S_T)]^2. \quad (4)$$

For our example, *MSE* is defined as the square of the expected difference between the true value of shape in the world, S_T, and the estimated value, S_C. While we do not attach any special significance to the use of the *MSE*, we have adopted it since it is a very widely used measure of the error of an estimate (Brainard & Freeman, 1997; DeGroot, 1986; Mamassian et al., 2002), closely related to variance (DeGroot, 1986). Variance is typically adopted in cue combination studies under the assumption that the estimator is unbiased. Indeed, when the estimator is unbiased, the *MSE* is equal to the variance, since the average estimated shape is equal to the true world value. In other words, combining unbiased cues to minimize variance will also minimize *MSE*. When one of more estimators is biased, this is no longer the case, and *MSE* may be expressed as the sum of terms relating to the variance and the bias of the combined-cue estimator (Berger, 1985):

$$MSE_C = v_C + b_C^2. \quad (5)$$

Here, v_C is the variance and b_C is the bias of the combined-cue estimate. In principle, it would be possible to weight cues differently so as to minimize *MSE*. However, this would require knowledge of both the variance and the bias in the relevant cues. As we have stated, we think that it is unlikely that the visual system has access to a measure of bias in each cue; we thus consider the situation in which only the variance of the cues is known. Using *MSE* in this way is a useful way of understanding the important relationship between variable and constant errors in cue combination but not a likely model of how the brain combines sensory information.

In the simplest case with our stereo and motion example, one cue could be biased while the other is unbiased. If the true value for shape in the world is given by S_T, and the stereo cue (S_S) is unbiased with a variance of v_S, but the motion cue (S_M) is biased by $b_M = (S_M - S_T)$ and has a variance of v_M, the bias of the combined-cue estimator is given by

$$b_C = \frac{v_S b_M}{v_S + v_M}. \quad (6)$$

Substituting Equations 3 and 6 into Equation 5 and simplifying gives us the *MSE* of the combined-cue estimator:

$$MSE_C = \frac{v_S v_M}{v_S + v_M} + \frac{v_S^2 b_M^2}{(v_S + v_M)^2} = \frac{v_S}{(v_S + v_M)^2}\left[v_S v_M + v_M^2 + v_S b_M^2\right]. \quad (7)$$

We can then determine those conditions when it would be beneficial to combine cues, even though one is biased, compared to vetoing the biased cue, even if the bias were in fact known. This occurs if the mean squared error of the combined-cue estimate is less than the mean squared error of the stereo estimator alone, which is the case when

$$b_M^2 < v_S + v_M. \quad (8)$$

Figure 1a shows sample plots of the MSE_C for a fixed variance of 1 for the unbiased stereo cue but for varying levels of bias and variance in the motion cue. As would be expected, the lowest MSE_C is found when the motion cue is also unbiased. As the bias of the motion cue increases, so too does the MSE_C. However, this combined-cue error is less than that of the stereo cue alone for the bias levels satisfying Equation 8. These are shown by the portion of the curves that lie below the horizontal line, which shows the mean squared error of the stereo cue. This line represents the point beyond which it would be beneficial in terms of the expected error to combine cues, despite the fact that one of them is biased (see also Burge et al., 2010).

The discussion so far has focused on the case of one biased cue. We can also assess the error that will result if both cues were biased, such that $b_M = (S_M - S_T)$ and $b_S = (S_S - S_T)$. In this case, the bias in the combined-cue estimator is given by

$$b_C = \frac{v_S b_M}{v_S + v_M} + \frac{v_M b_S}{v_S + v_M}, \quad (9)$$

and the combined-cue mean squared error is given by

$$MSE_C = \frac{v_S v_M}{v_S + v_M} + b_C^2. \quad (10)$$

Figures 1b–1d follow the same format as Figure 1a and show the MSE_C for the combined-cue estimator for a range of variances and biases of the motion cue, with separate graphs for different levels of bias in the stereo cue. Here, the bias in the stereo cue is always positive, and the variance of the stereo estimator is set to 1. It can be seen that as the stereo cue becomes more biased, greater levels of bias in the motion cue can be present before the combined-cue *MSE* is greater than that of the stereo cue

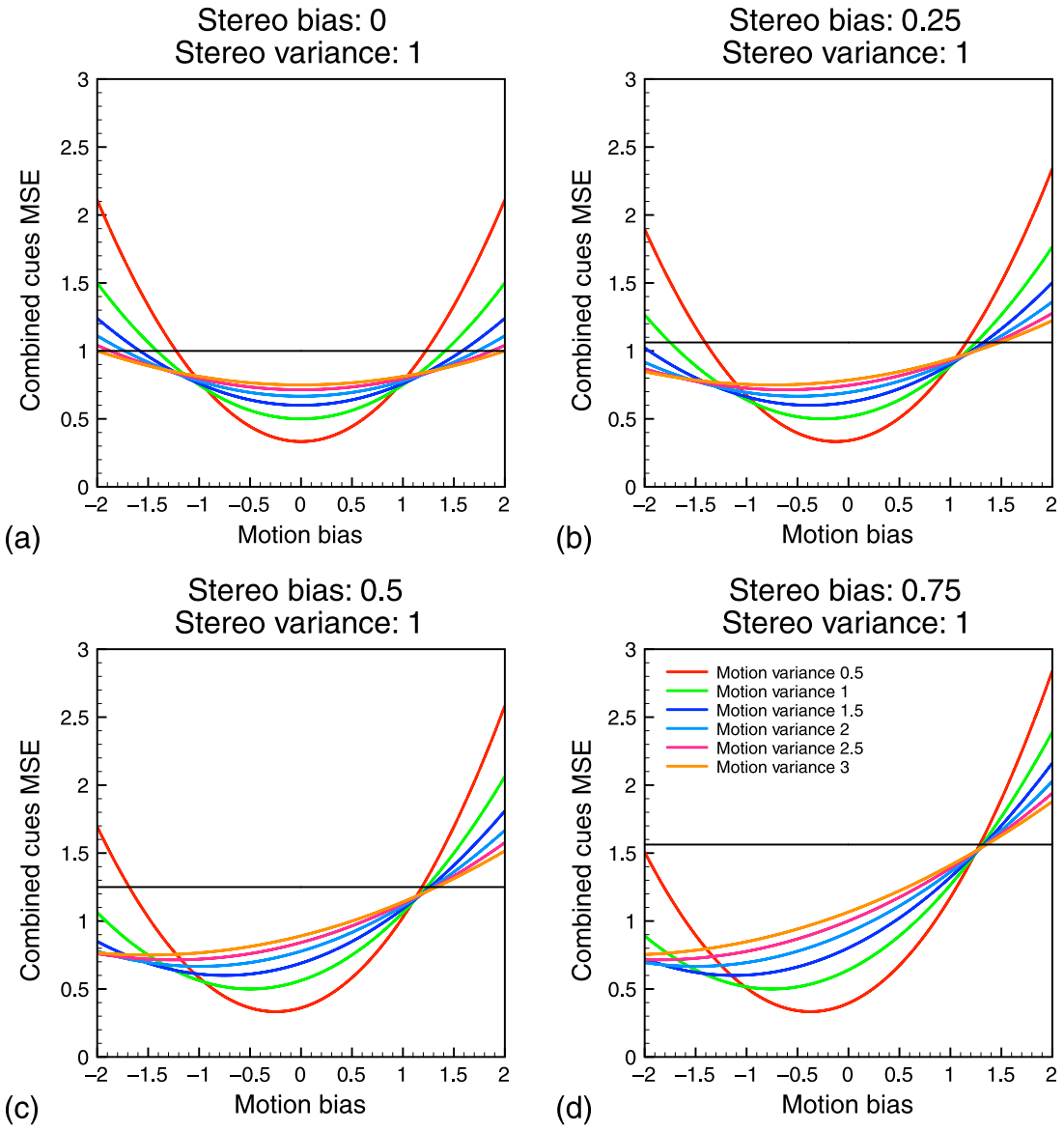

Figure 1. Plots of the mean squared error in the combined-cue estimator (ordinate) for varying levels of bias in the motion cue (abscissa). The separate plots (a) through (d) are for different levels of bias in the stereo cue, as indicated on each plot. Within each plot, the black horizontal line shows the mean squared error of the stereo estimator in isolation, and colored curves show the mean squared error of the combined-cue estimator for various levels of variance in the motion estimator. For all plots, the variance in the stereo cue was set to 1.

alone. This is especially true for negative biases in the motion cue, which serve to counteract the stereo cue's positive bias. The interplay between the variances and biases of the cues shows that there are circumstances when, even though both cues are biased, it can still pay the observer to combine cues rather than to veto one or the other.

Previous studies on stereo–motion combination

A number of previous studies have investigated how stereo and motion cues to shape might be combined. Early work tended to focus on the fact that, because shape from motion and shape from stereo scale differently with distance, it is possible for the two sources of information to conjointly specify the veridical shape of an object (Richards, 1985). Overall, the results are generally in the negative; shape is still misperceived when stereo and motion cues are available (Tittle, Todd, Perotti, & Norman, 1995; Todd, 1998; Todd, Chen, & Norman, 1998; Todd & Norman, 2003; Todd, Tittle, & Norman, 1995). One study that did find near veridical performance with stereo and motion cues is that of Johnston, Cumming, and Landy (1994). There are, however, a number of complications in interpreting their data.

First, their stimuli also contained a relatively strong texture cue, which was always consistent with the motion cue. Second, focus cues were in conflict with the geometric cues used to render the stimuli, as the screen was positioned a fixed distance from the observer. This makes interpretation of the results in terms of weighted averaging of stereo and motion difficult. Finally, Todd and Norman (2003) identified a heuristic strategy that observers could have used to estimate shape in this study without the need to use a weighted averaging scheme. An overall assessment of the literature, therefore, suggests that stereo and motion cues are subject to systematic bias in most circumstances. This bias has caused some investigators to look beyond weighted averaging to model stereo–motion combination (Domini, Caudek, & Tassinari, 2006).

The intrinsic constraint (IC) model (Tassinari & Domini, 2008) proposes that stereo and motion information are not processed in isolation as in weighted averaging, but instead both cues are used to determine shape up to an affine level, and then the most likely Euclidean interpretation consistent with this affine structure is estimated. There is active debate as to whether the IC model provides a good description of cue combination (Domini & Caudek, 2009; MacKenzie, Murray, & Wilcox, 2008), and there are a number of differences between it and the Bayesian weighted averaging scheme (Domini & Caudek, 2009). As will become clear, once it is acknowledged that cues may provide biased sensory estimates, the weighted averaging framework provides a good account of stereo–motion cue combination. We discuss the IC model further in the Discussion section.

Summary of the current study

We investigated the way in which human observers combine information from motion and stereo for the estimation of three-dimensional shape. Rather than introduce small conflicts between the cues, while assuming that each provides veridical information (e.g., Hillis et al., 2004), we exploit the fact that observers show biases in estimating shape from motion and stereo, and that these biases typically result in different estimates of the same three-dimensional shape from each cue at a given viewing distance. This provides an ideal way to see whether the visual system combines discrepant sensory estimates in order to minimize variance and to examine the consequences of these discrepancies for the accuracy of perceived 3D shape.

Methods

Participants

Five observers took part in the experiment. These were one of the authors (PS) and four people naive to the purpose of the experiment. All had normal or corrected-to-normal vision and good stereopsis.

Apparatus

The stimuli were viewed in a Wheatstone stereoscope, such that the observer's eyes viewed two identical monitors through two front-surface mirrors orientated at 45 deg relative to a line of sight defined by zero vergence, i.e., eyes looking at optical infinity. The center-to-center distance of the mirrors was matched to the interocular distance of the observer. Observers were positioned in chin- and headrests to minimize head movement. Eye height was adjusted to match that of the vertical center of the monitor screens. The two monitors comprising the stereoscope were spatially calibrated and gamma corrected. The screen resolution of each monitor was 1152 by 864, running with a refresh rate of 85 Hz. Each monitor was attached to a rail-mounted, custom-built, metal enclosure. This allowed us to set the path length between the eye and each monitor to match the vergence specified distance of the rendered stimuli, while at the same time maintaining highly accurate monitor orientation.

Stimuli

The stimuli were perspective projections of horizontally orientated elliptical hemi-cylinders, 10 cm in length and 6 cm in height. The radius of a physically circular cylinder would, therefore, be 3 cm, i.e., the cylinder's nearest point would appear 3 cm in front of the monitor's surface. The projection took account of each observer's interocular distance. The surface of each cylinder was defined by anti-aliased red dots that were positioned with subpixel accuracy. The diameter of the dots was 4 pixels (approximately 1.3 mm). The dot density of the cylinder's surface was 6 dots/cm^2, and cylinders were positioned centrally on the screen. Figure 2 shows a diagrammatic side view of the viewing arrangement, and Figure 3 shows a stereogram of the stimulus.

There were three stimulus conditions: (1) stereo only, (2) motion only, and (3) stereo and motion. In the motion-only condition, the scene was viewed with the right eye alone. Motion was produced by sinusoidally oscillating the cylinder centrally around its major axis (Figure 2). The amplitude of the oscillation was ±35 deg and the cylinder oscillated with a frequency of 1 Hz. The stereo and motion stimuli always contained physically consistent stereo and motion information. The initial direction of the cylinder's movement was determined randomly on each trial. During piloting, we found that, despite there being sufficient geometric information available in the motion-only condition, some observers occasionally perceived a depth reversal of the cylinder. This gave them an ambiguous and somewhat nonrigid percept. In order to

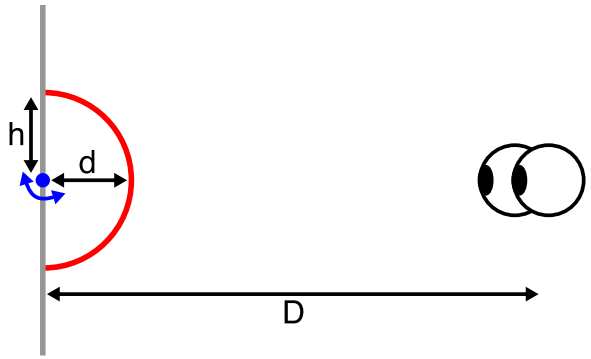

Figure 2. Diagrammatic side view of the experimental stimulus (not to scale). Observers viewed a horizontally orientated disparity-defined hemi-cylinder positioned at eye height, at a distance *D*. Their task was to judge whether the cylinder was circular, i.e., whether $h = d$. Throughout the experiment, a vertically orientated gray bar with a central fixation point was positioned at eye height, at the viewing distance *D*. In the motion and stereo–motion conditions, the cylinder sinusoidally rotated back and forth around its major axis, as indicated by the blue dot and curved blue arrow in the diagram. The magnitude of oscillation was ±35 degrees. With no rotation applied, the upper and lower edges of the cylinder abutted a vertical gray bar (as in the diagram). When the cylinder rotated, its upper and lower edges could pass "through" the gray bar, providing observers with an occlusion cue that successfully disambiguated depth polarity (see main text for more details). See also Figure 3, for a stereogram of the experimental stimulus.

eliminate this possibility during the experiment, a gray bar 10 cm in height and 3 cm in width was rendered centrally at the screen distance to disambiguate the motion (Figures 2 and 3).

When the cylinder oscillated, parts of its upper and lower edges could pass "through" the bar (Figures 2 and 3). This provided an occlusion cue that successfully disambiguated the cylinder's depth polarity for all observers. The bar had a white fixation point approximately 3.3 mm in diameter positioned centrally. The cylinder's surface was otherwise transparent; this meant that the observers could at all times see the bar and fixation point, which were present throughout the experiment. The background of the screen during the experiment was black. Stimuli were presented at 40, 60, 80, or 100 cm (which included the path between the eye and mirrors). Given the fixed dimensions of the cylinder, this meant that the visual angle subtended by the stimuli varied naturally with distance. All stimuli were rendered online in OpenGL using Matlab and the Psychophysics toolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007). The room in which the experiment was conducted was otherwise completely dark.

Procedure

Observers completed an apparently circular cylinder task (Johnston, 1991), in which they were asked to judge whether the cylinder they were presented with was squashed or stretched in depth extent relative to a physically circular cylinder (Figure 2). Trials were blocked by viewing distance (40, 60, 80, and 100 cm) and the cue(s) defining the cylinder (motion, stereo, or stereo and motion) and were completed in a randomized order. Within a block of trials, the depth of the cylinder was varied using the method of constant stimuli. There were 9 depths and each was presented 30 times, in a randomized order. The exact depths depended on the

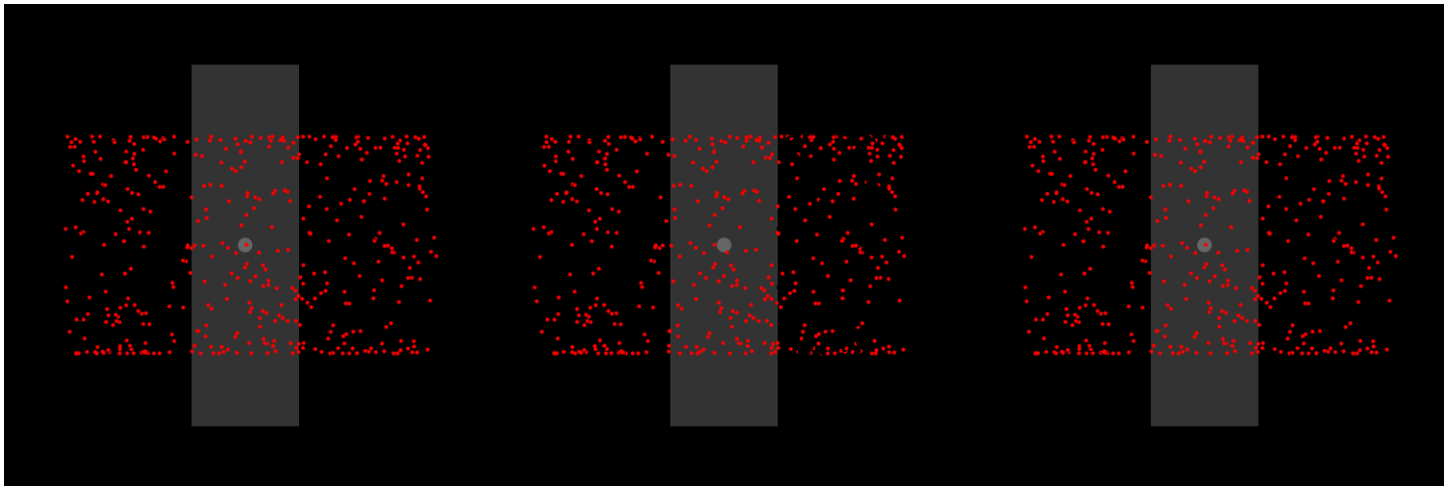

Figure 3. Stereogram of the experimental stimuli. Left and middle images are for divergent fusion, and the middle and right images are for cross-fusion. Note that the sizes of dots and luminance of aspects of the stimuli have been changed to make the images easier to free fuse.

person, the viewing distance, and the available cues and were determined on the basis of pilot experiments. Prior to the start of the experiment, observers were given a chance to familiarize themselves with the stimuli and task.

On each trial, the cylinder was presented for 2 s, after which it was extinguished, leaving only the gray bar and fixation point. There was no time limit on observers making a response; however, they typically responded in around 1 s. The next trial in the block was presented automatically after a 1 s interstimulus interval, which started upon the observer making their response with a keyboard button press. New dot coordinates for the cylinder's surface were generated on each trial.

Results

Cumulative Gaussian functions were fit to observers' data using a bootstrapping technique (Wichmann & Hill, 2001a, 2001b); from this fitted function, we were able to determine the point of subjective equality (PSE) and the just noticeable difference (JND). The PSE is defined as the 50% point of the psychometric function and provides a measure of the cylinder that would appear circular to the observer, in units of depth-to-half-height ratio. A PSE of one represents a circular cylinder, whereas a PSE greater than one indicates that, to be perceived as circular, cylinders needed to be *stretched* in depth extent. Conversely, a PSE less than one indicates that, to be perceived as circular, cylinders needed to be *squashed* in depth extent. The JND was defined as the standard deviation of the cumulative Gaussian fitted to the observer's data. This is equivalent to the difference between the cylinder depth-to-half-height ratios corresponding to the 50% and 84% of the psychometric function. Figure 4 shows the PSEs and Figure 5 shows the JNDs for each observer across our range of viewing distances.

The PSE data show that for our observers a perceptually circular cylinder, for both single-cue conditions, was generally one that was squashed in depth extent. This indicates that observers were overestimating the depth in our displays. However, as can be seen, the PSEs and JNDs for the single-cue data depended on both the viewing distance and whether the cylinder was defined by motion or stereo information. This variation in accuracy and precision allowed us to test the weighted averaging model. From the single-cue JNDs and PSEs, we predicted those for the combined-cue condition using Equations 2 and 3. Figure 6 shows the stereo–motion PSEs and those predicted by the model. Similarly, Figure 7 shows the stereo–motion JNDs and those predicted by the model. In both plots, we show 95% confidence intervals around the PSE and JND predictions. These were determined using the bootstrapped 95% confidence limits around the PSEs and JNDs of the single-cue conditions. For example, to determine the confidence intervals around predicted combined-cue PSEs, we used the weights determined by the stereo and motion JNDs, as normal, but used the upper or lower bounds of the bootstrapped 95% confidence intervals around the stereo and motion PSEs, instead of the PSEs themselves. This gave an upper or lower bound on the predicted combined-cue PSE. Confidence intervals around the predicted combined-cue JNDs were calculated in a similar manner.

As can be seen, the weighted averaging model provides a good fit to the combined-cue data for both the PSEs (Figure 6) and JNDs (Figure 7). We fit a least squares linear model to the predicted and observed PSEs, and JNDs, for each observer to get an overall idea as to the fit of the data to predictions of weighted averaging. The mean R^2 values for linear fits were 0.63 for the PSEs and 0.62 for the JNDs. Surprisingly, few cue combination studies provide explicit statistically assessments of the fit of the model, leaving the reader to judge this visually. One such study that has is that of Burge et al. (2010). For the combination of vision and haptic cues to slant, they found an overall R^2 value of 0.60 for both observed and predicted JNDs and observed and predicted PSEs.

A number of assumptions are implicit in the analysis of data from tasks where observers judge stimuli relative to an internal standard, such as a circular cylinder (Johnston, 1991) or a 90-degree dihedral angle (Watt, Akeley, Ernst et al., 2005). We assume that observers used the same internal standard to judge object properties across all conditions, such that over conditions observers did not change their mind as to what constituted a circular cylinder. We also assume that observers scaled the retinal size of the object in the same way for both cues at a given distance. If they did not do this, cylinders defined by motion or stereo, presented at the same distance, would look to be of different sizes. This might cause the single-cue data to poorly predict the combined-cue data. Finally, we assume that the functions relating perceived to physical shape are linear across the range of ratios covering the differences in perceived shape from the individual cues.

Would it have been better to veto a cue?

The analysis of mean squared error presented in the Introduction section allows us to gain an understanding of whether combining stereo and motion cues using weighted averaging resulted in more or less error than would have occurred with vetoing one or other cue (Landy et al., 1995). This is because mean squared error takes account of both the constant and variable errors in the combined-cue estimate. Figure 8 plots the mean squared error for the single-cue and combined-cue conditions and that predicted for the combined-cue condition given the PSE's and JNDs of the single-cue data. Although *MSE* for the single-cue and combined-cue conditions varies across observers, in many

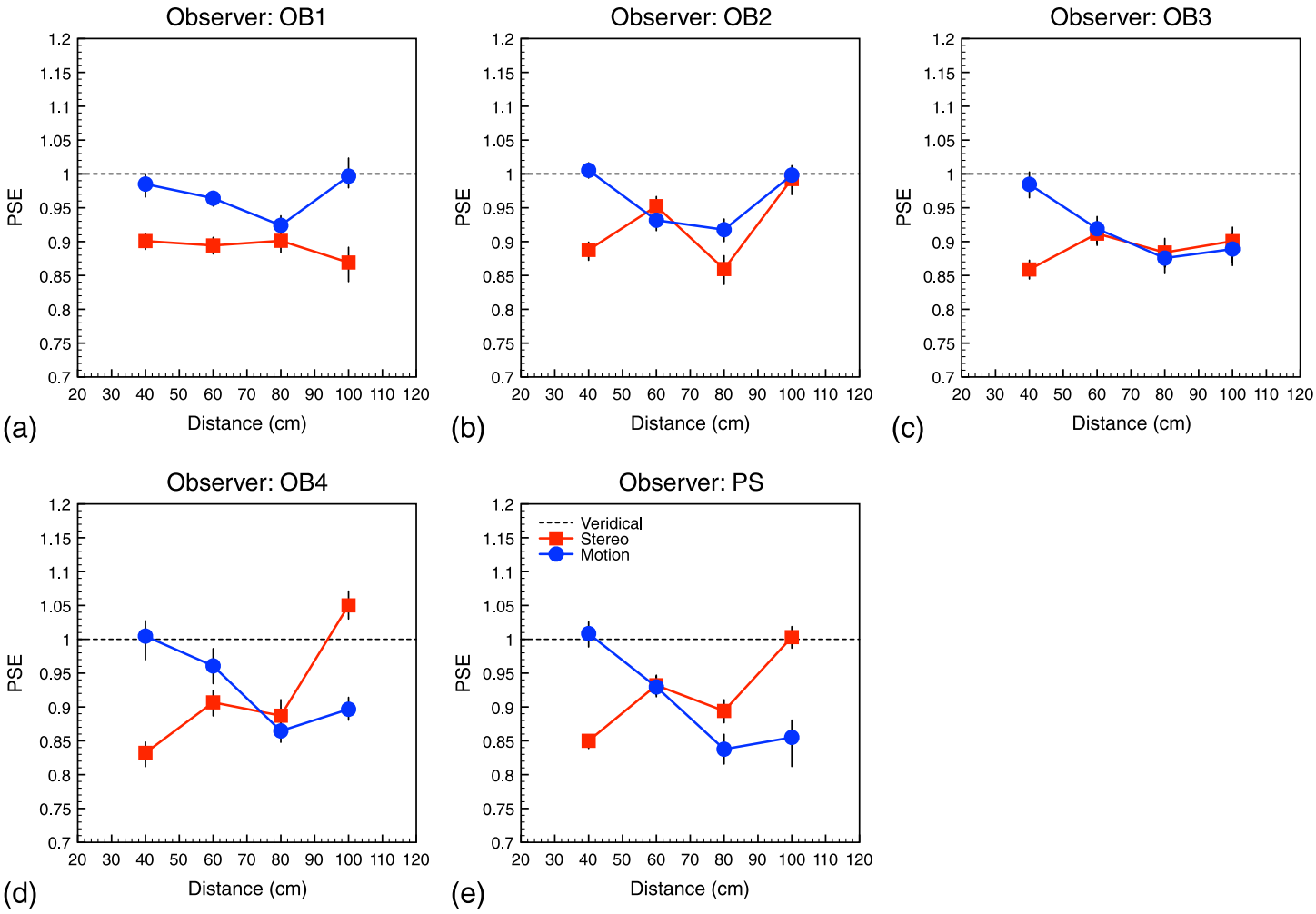

Figure 4. Plots (a) through (e) show each observer's stereo and motion PSEs across distance. Error bars show 95% confidence intervals derived from the psychometric function fitting procedure. PSEs below one mean that to be perceived as circular, cylinders needed to be *squashed* in depth relative to their height. PSEs greater than one mean that cylinders needed to be *stretched* in depth relative to their height to be perceived as circular. These are consistent with an over- and underestimation of depth, respectively.

instances, there is a clear advantage, in terms of reduced *MSE*, of combining cues rather than vetoing one or other cue. The mean R^2 value of a linear fit to the predicted and observed combined-cue *MSE*s for each observer was 0.78, which represents a good correspondence to that predicted.

Overall, the *MSE* data support the analysis presented in the Introduction section. If the visual system were able to calibrate cues so as to eliminate bias and maintain external accuracy, combining cues so as to minimize variance using weighted averaging would also minimize *MSE*. However, when bias is present, this is not necessarily the case (Figure 1). Observers exhibited clear perceptual bias but combined cues so as to minimize the variance of the combined-cue estimate. Despite this fact, there are clear instances where the increase in bias observers accrued from combining biased cues was more than compensated for by a reduction in variance, leading to a lower overall mean squared error. This suggests that weighted averaging can be a robust strategy to adopt in the face of unknown perceptual bias.

Deviations from weighted averaging

While the data are well fit by the weighted averaging model, some deviations from the predictions are evident, especially for observer OB3. This observer was able to near veridically estimate shape when provided with stereo and motion information, which deviates from the model's predictions. In contrast, this observer's JNDs were well fit by the model. There are a number of reasons why deviations from weighted averaging might occur. The first is in terms of unmodeled cues or the use of perceptual priors. For reasons detailed in the Discussion section, we feel that these cues are unlikely to have significantly affected performance in our task. The second is the

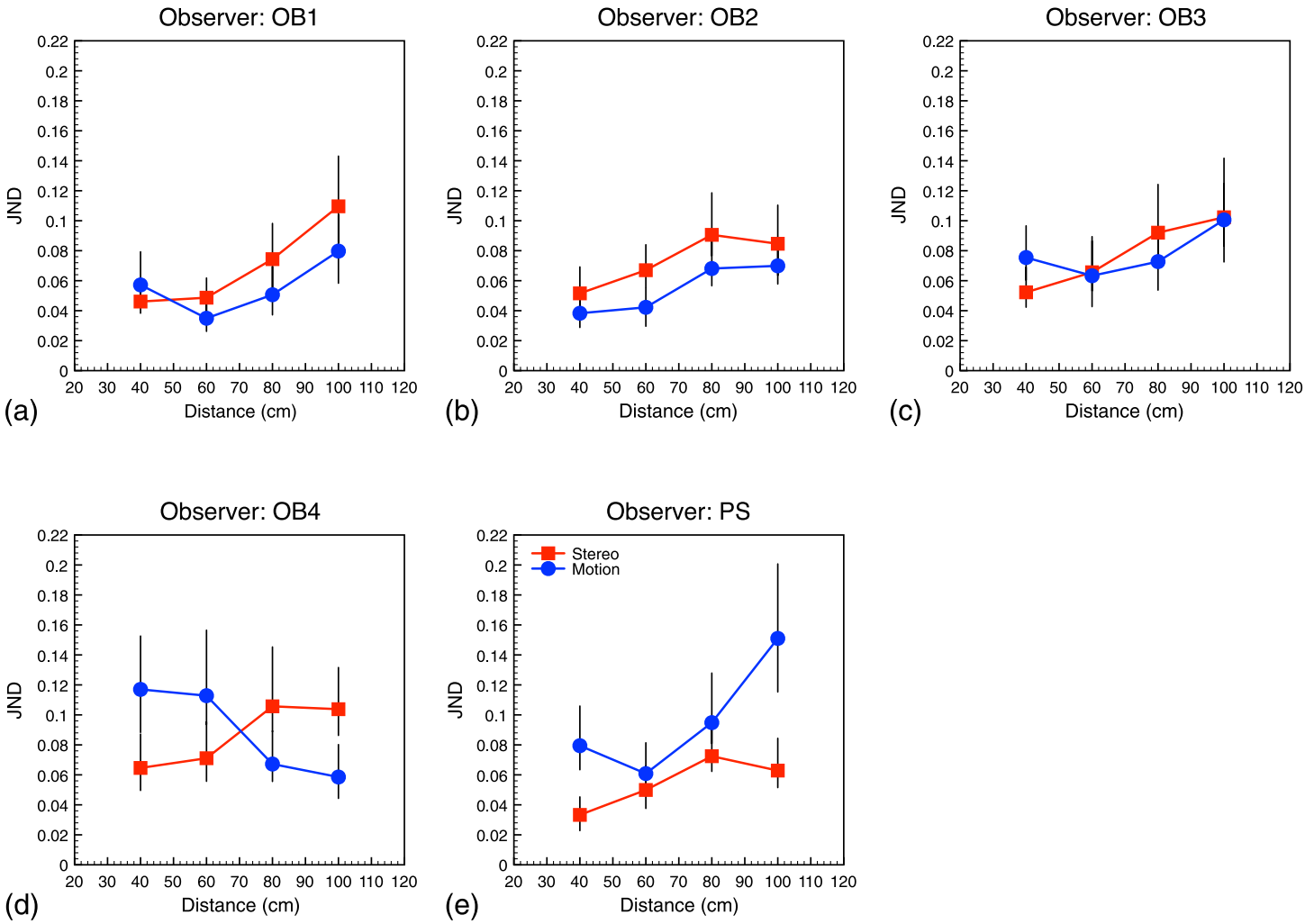

Figure 5. Plots (a) through (e) show each observer's stereo and motion JNDs across distance. Error bars show 95% confidence intervals derived from the psychometric function fitting procedure.

possibility that, at least for some observers, promotion of stereo and motion information might be possible (Richards, 1985). If this is the case, these observers are clearly in the minority (Tittle et al., 1995; Todd, 1998; Todd & Norman, 2003; Todd et al., 1995). A more likely explanation is simplifications and assumptions underlying the weighted averaging model. We discuss these at greater length below.

Discussion

Integrating biased sensory estimates

In the current paper, we provide evidence that, in estimating three-dimensional shape, human observers combine stereo and motion so as to minimize the variance of the final combined-cue estimate (Ernst, 2006; Ernst & Banks, 2002; Ernst & Bülthoff, 2004). In isolation, both stereo and motion information typically result in biased estimates of shape that depend on the distance at which the object is viewed (Tittle et al., 1995; Todd et al., 1995). This means that combining stereo and motion cues in proportion to their reliability does not necessarily result in a more accurate percept. The transfer functions that relate sensory cues to properties of the world are likely to be highly nonlinear (Hogervorst & Eagle, 1998; Scarfe & Hibbard, 2004), so bias could be introduced into perceptual cues as a natural consequence of the way they are sensed. This means that it is a nontrivial problem for the visual system to know when a cue is biased.

Bias in perceptual estimates is not inevitable if observers are able to calibrate their sensory data. However, the prevalence of perceptual bias in the estimation of metric object properties, even with real-world stimuli (Bradshaw et al., 2000; Cuijpers, Kappers, & Koenderink, 2000; Koenderink, van Doorn, Kappers, & Todd, 2002; Koenderink, van Doorn, & Lappin, 2000; Wagner, 1985; Watt, Akeley, Ernst et al., 2005), suggests that in many circumstances calibration to maintain external accuracy

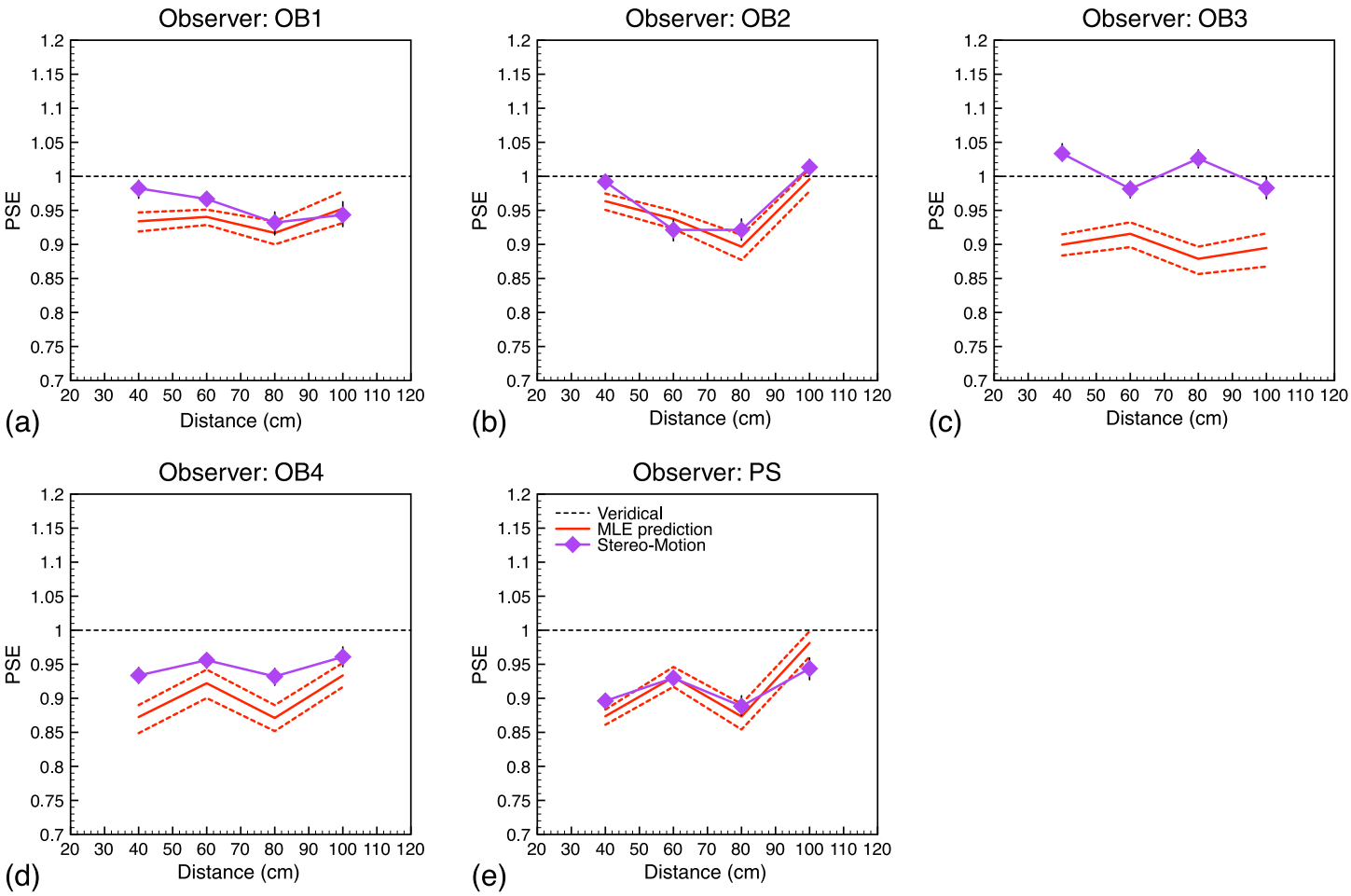

Figure 6. Plots (a) through (e) show observers stereo–motion PSEs with 95% confidence intervals derived from the psychometric function fitting procedure. The predictions from the MLE model are shown as the solid red line and 95% confidence intervals around these predictions are shown as the red dashed lines.

has not been possible. To maintain accuracy with respect to the world, the visual system needs to have information regarding the accuracy of its cues. This information may be unobtainable because the only way to judge accuracy is by using the very cues that one might need to calibrate (Ernst & Banks, 2002).

We derived equations for the level of mean squared error in the combined-cue estimate that would result from combining cues using weighted averaging, when in fact one or more cues were biased. While minimizing *MSE* is unlikely to be a viable strategy for the visual system, given that the bias in individual cues is unknown, *MSE* allows us to gain some understanding of when it would be beneficial to combine biased cues rather than veto one or the other (Landy et al., 1995). This is because it incorporates the constant error as well as variable error in an observer's estimates (Berger, 1985). Across the range of biases found in the present study, there were clear instances where the mean squared error in the combined-cue estimate was less than that of the individual cues. This suggests that optimizing cue combination for variance might be a reasonably robust strategy for the visual system to adopt.

This is not to say recalibration of perceptual attributes in response to our actions in the world is not possible or does not occur. The brain is clearly highly attuned to the statistical structure of the environment. Evidence for this comes from its ability to adaptively remap the relationship between sensory information and properties of the world (Adams et al., 2001) and to learn completely new sensory mappings between arbitrary sensory inputs (Ernst, 2007). However, cue calibration clearly fails to eliminate perceptual bias under many circumstances (Todd & Norman, 2003). Interestingly, evidence has shown that when we make movements, the brain tends to adopt control strategies that continuously sample relative information over the course of a movement (Saunders & Knill, 2003, 2005). This removes the need to veridically estimate metric properties of the world for adaptive and skillful behavior (Smeets & Brenner, 2008). This is a stark contrast to the assumption that, *because* behaviors are

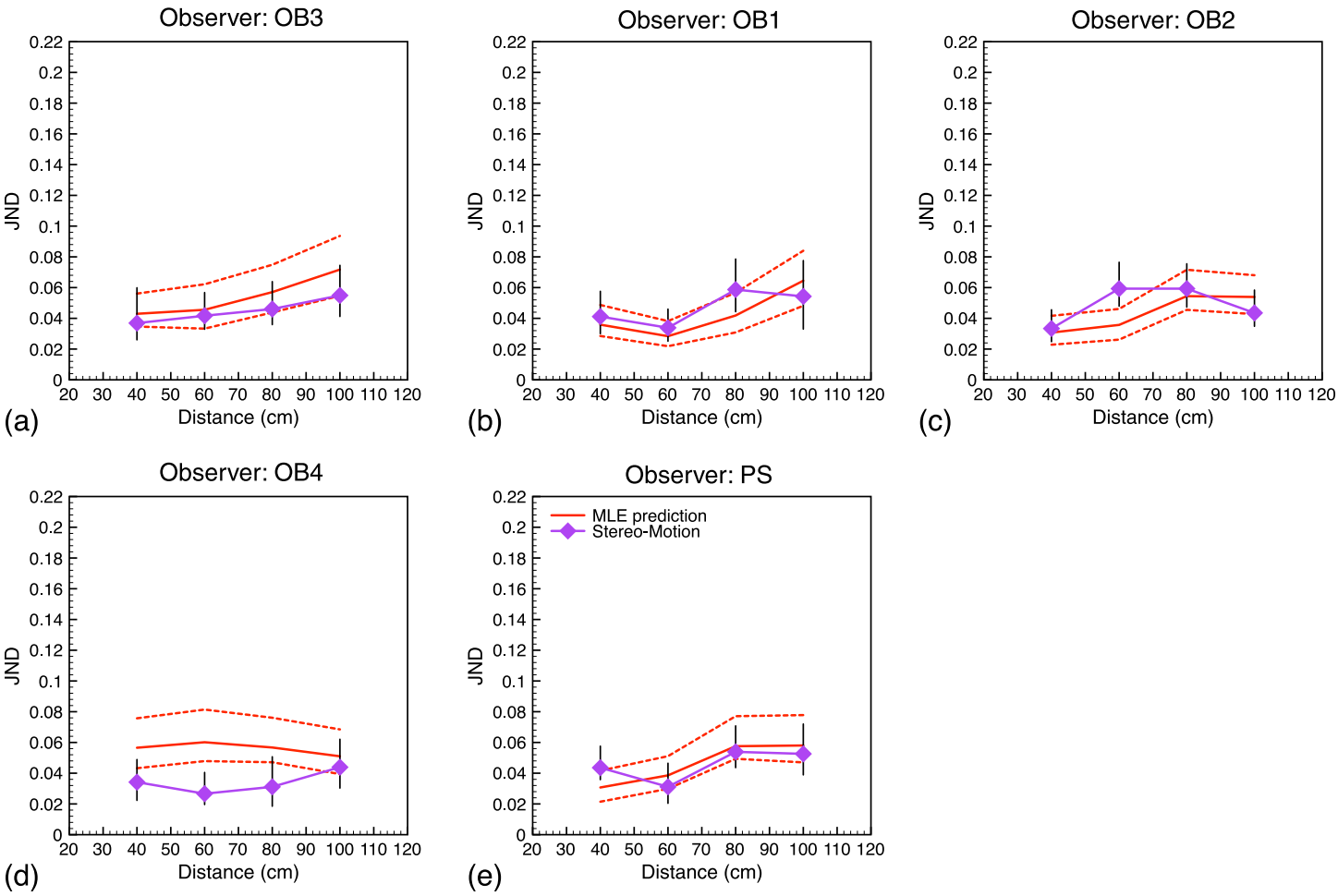

Figure 7. Plots (a) through (e) show observers' stereo–motion JNDs with 95% confidence intervals derived from the psychometric function fitting procedure. The predictions from the MLE model are shown as the solid red line and 95% confidence intervals around these predictions are shown as the red dashed lines.

skilled and adept, they must be controlled by accurate metric representations (Milner & Goodale, 1995, 2006).

The role of unmodeled cues and perceptual priors

Computer-generated 3D stimuli typically contain uncontrolled cues that conflict with the cues being manipulated to render the stimuli (Akeley, Watt, Girshick, & Banks, 2004; Hoffman, Girshick, Akeley, & Banks, 2008; Watt, Akeley, Ernst et al., 2005; Watt, Akeley, Girshick, & Banks, 2005). One of the main reasons for this is that the light rendering the scene emanates from a single display surface (Watt, Akeley, Ernst et al., 2005). This means that focus cues, such as accommodation and blur, signal flatness rather than the intended 3D properties of the scene. In a slant estimation task, Watt, Akeley, Ernst et al. (2005) demonstrated the importance of such cues by rotating their display surface so that focus cues were either consistent or inconsistent with the amount of slant specified by binocular and texture information. While conflicting cues had no measurable effect on the perceived slant of disparity-defined surfaces, the perceived slant of texture-defined surfaces was significantly reduced, consistent with cues to flatness.

They also measured the effect of conflicting cues on disparity scaling by inducing greater cue conflict by positioning the front-parallel screen at a distance different from that used to render the stimuli. Under these conditions, they found that conflicting cues reduced depth constancy in stereo-defined objects, suggesting that focus cues can affect disparity scaling by influencing the distance used to scale image properties (Brenner & Landy, 1999; Brenner & van Damme, 1999). It is, therefore, important to consider unmodeled cues such as these in the current experiment, in particular whether they can account for the pattern of biases that we observed.

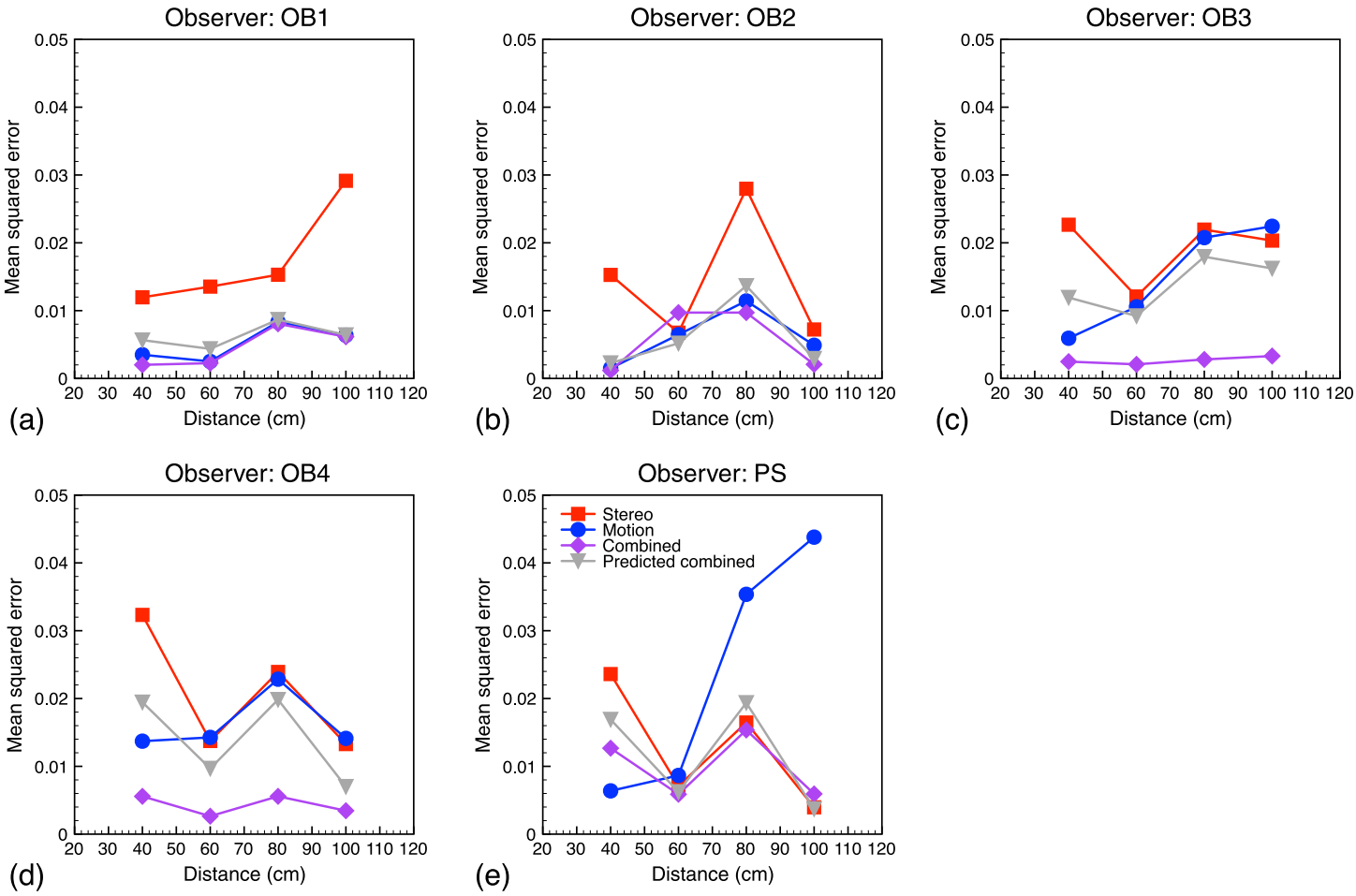

Figure 8. Plots (a) to (e) show the mean squared error (*MSE*) of the stereo, motion, and combined-cue conditions for each observer. In addition, we show the predicted *MSE* for the combined-cue condition. For more details, see the accompanying text.

We matched the distance to our monitors to the vergence specified distance of the stimuli so as to minimize the effects of conflicting focus cues. However, it remains possible that the absence of a gradient of accommodative blur over the surface of our cylinders may have been detectable. In addition, other cues could have signaled stimulus flatness, specifically: (1) motion parallax from residual head movements, (2) texture cues from the pixel grid of the screen, and (3) the uniformly circular dot size of the points defining the cylinder. We can, therefore, make two predictions regarding these cues. First, observers should underestimate depth in the 3D scene because all uncontrolled cues signal stimulus flatness (Watt, Akeley, Ernst et al., 2005). Second, this effect should be most prominent in the single-cue conditions and least prominent in the combined-cue condition. This is because in the combined-cue condition observers have two cues signaling the intended depth percept rather than one.

As regards the first prediction, that depth should be underestimated in our stimuli, Figure 4 shows that this was clearly not the case. For both the stereo and motion single-cue conditions, depth was near universally overestimated. This means that contrary to the predictions of conflicting cues to flatness (Watt, Akeley, Girshick et al., 2005), a perceptually circular cylinder for our observers was one that was squashed in depth extent. Out of 40 data points, the single point for which this does not hold is for OB4 at the 100-cm viewing distance, with the stereo cue. These results are consistent with previous studies that have shown the depth of stereo-defined objects placed below 80–100 cm to be overestimated with both simulated and real-world objects (e.g., Johnston, 1991). We now consider the second prediction, that the underestimation of depth should be largest in our single-cue conditions.

Figures 4 and 5 show that this was also not the case. Because the combined-cue PSEs were well fit by the weighted averaging model, they typically fell between the PSEs of the single-cue conditions. This means that with both cues, the depth perceived was typically greater than that in one of the single-cue conditions and less than that in the other. The single observer who clearly deviated

from the predictions of weighted averaging was OBS3. However, this observer's data are also inconsistent with the predictions of cues to flatness, because with both cues this observer, although perceiving shape near veridically, perceived less depth than with each cue in isolation. We can, therefore, be confident that cues to flatness (e.g., Watt, Akeley, Ernst et al., 2005) cannot explain the pattern of results in our data. In fact, they predict the opposite pattern of results to that which we find in nearly all instances.

Another potential source of information that should be considered is prior knowledge of the probable structure of the environment. Within the Bayesian framework, this knowledge is instantiated in the form of prior probability distributions. Priors can have a significant effect on perception. Specifically, as the variance of the sensory data increases, priors are predicted to have a more pronounced effect. This is a sensible strategy, as when faced with poor information the system places more weight on its past experience of the structure of the environment. As with uncontrolled cues, many studies have left priors unmodeled (Ernst & Banks, 2002; Johnston et al., 1994) or assumed that they will have little influence on the observed results, since the variance of the prior might be expected to be large in comparison with that of sensory cues (Hillis et al., 2004). These approaches are clearly simplifications of a more complicated picture.

There are currently no direct measurements of the statistical likelihood of different shapes in the environment, but we can make some inferences on the basis of psychophysical studies. When interpreting an elliptical projection at the retina, observers generally assume that the object underlying the projection is circular. This allows the observer to use the aspect ratio of the projection to make an estimate of the 3D orientation of the object (Knill, 2007; Muller, Brenner, & Smeets, 2009; Seydell, Knill, & Trommershauser, 2010). We might infer, therefore, that a prior for 3D shape in our cylinders may bias observers to see our stimuli as circular. Like focus cues, this prior should be most noticeable in the single-cue conditions, as with both stereo and motion cues the prior should receive less weight. It becomes immediately apparent, however, that a circularity prior cannot provide an alternative account for our data. If we consider the case where a cue provides a biased estimate of cylinder shape, any action of a circularity prior can only act to decrease, but not eliminate, the magnitude of this bias. As such, a circularity prior fails to provide a valid account of the bias that we observe (Figures 4 and 6).

Cue conflicts and perceptual unity

An important consideration for the visual system is when it should combine sensory information provided by different cues (Shams & Beierholm, 2010). In a cue combination study, Gepshtein, Burge, Ernst, and Banks (2005) varied the spatial proximity of visual and haptic cues. Discrimination performance was consistent with statistical optimality when the cues were spatially coincident, but as the spatial conflict increased, precision decreased, such that at the largest conflict it was consistent with that of one cue alone. This pattern of results is sensible as cues are more likely to arise from different objects as the spatial separation between them increases (Ernst, 2006). Similarly, instances of sensory bias and cue conflict are interesting because they probe the circumstances under which the visual system combines discrepant information, presumably because it believes that even though the information is inconsistent, it in fact arises from a single object.

The most extensively investigated consequence of large cue conflicts is that of bistability (van Ee, van Dam, & Erkelens, 2002). In this situation, perception can alternate between that defined by each cue. During debriefing, none of our observers ever reported bistability. To some extent, this is to be expected, as in the combined-cue condition the points viewed in stereo were *carrying* the motion signal. A more likely consequence of cue conflict in our study would have been for the combined-cue stimuli to look nonrigid. This was also not reported by any of our observers. If anything, the observers commented that the combined-cue stimuli looked the most "real." Our observers, therefore, seem to have treated the cues as belonging to the same object. As such, these results are consistent with those of Girshick and Banks (2009), who also observed perceptual unity with large cue-conflict stimuli. Finally, it is interesting to note that under many situations the visual system also seems quite unperturbed by highly discrepant sensory inputs arising from the same object (Smeets & Brenner, 2008).

Modifying models of cue combination

Minimizing variance is just one of a set of possible strategies that the visual system might adopt when combining sensory information (Clark & Yuille, 1990). While the weighted averaging model predicted our data well, the fit to this model was not perfect. This is the case for the literature at large. The fit is generally good, but not perfect. A number of other studies have shown that while their data might show sensitivity to the variance of cues, weighted averaging does not fit their data (Butler, Smith, Campos, & Bulthoff, 2010; Rosas, Wagemans, Ernst, & Wichmann, 2005). Deviations from predictions are clearly important because they allow us to identify simplifications and flaws in the models. One assumption addressed in the current study is that individual cues provide unbiased estimates of world properties. Other common assumptions that we have also adopted are that the information provided by each cue is well modeled by a Gaussian distribution (Hillis et al., 2002, 2004) and that information from different cues is conditionally independent (Oruc et al., 2003).

The intrinsic constraint (IC) model of cue combination was in part proposed to account for the large biases of perceived shape exhibited by observers (Domini et al., 2006; Tassinari & Domini, 2008). This is because when individual cues are assumed to provide unbiased estimates, the weighted averaging framework has problems accounting for biases in the combined-cue percept without evoking the role of conflicting information such cues to flatness (Todd et al., 2010; Watt, Akeley, Ernst et al., 2005). We have shown that under conditions where cues to flatness predict the opposite pattern of bias to that observed, the weighted averaging framework can account for performance as long as one accepts that individual cues can provide biased estimates of world properties, such as three-dimensional shape. We used the weighted averaging framework rather than the IC model for a number of reasons.

The primary reason is that, within the distance range that we used (distances up to 1 m), observers are readily able to use vergence information to estimate distance and scale retinal disparity (Brenner & Smeets, 2000; Brenner & van Damme, 1998). With around 90% of the vergence range being used up for distances below 1 m, it is in this near distance range where stereo information should be of maximum utility (Howard, 2002). The IC model currently has no way to model the role of extraretinal cues such as vergence or retinal cues such as vertical disparity, so in its current form it cannot model performance where these cues have a clear and demonstrable effect (Domini et al., 2006). Second, it is not clear whether the IC model can be readily generalized to model the full gamut of multimodal cues available to the observer, which the Bayesian weighted averaging framework has had considerable success in doing (Ernst & Bülthoff, 2004). Interesting, our data show that the weighted averaging framework can easily model the effects of perceptual bias.

Some of the additional assumptions used in weighted averaging are also starting to be tackled. Girshick and Banks (2009) have modeled the combination of texture and disparity cues to slant with "heavy-tailed" Gaussians and proposed that combination with these distributions could account for robust vetoing when cue conflicts are large. Others have taken a more direct approach and modeled the transfer function between world and cue, in order to directly assess the shape of the likelihood distribution and the bias that this might introduce into the estimation process (Hogervorst & Eagle, 1998; Scarfe & Hibbard, 2004). A further, but important, point to consider is that estimation strategies may be highly cue specific or specific to certain environmental circumstances (Glennerster, Rogers, & Bradshaw, 1996; Scarfe & Hibbard, 2006; Todd, 2004; Todd et al., 2010; Todd & Norman, 2003). While this makes it difficult to derive single unified rules for cue combination, and sensory processing in general, it is exactly what might be expected for an evolved system attuned to those aspects of its environment that allow for adaptive behavioral control.

Acknowledgments

The authors would like to thank the anonymous referees and the editor for their constructive comments during the review process.

Commercial relationships: none.
Corresponding author: Peter Scarfe.
Email: p.scarfe@ucl.ac.uk.
Address: Department of Cognitive, Perceptual and Brain Sciences, University College London, 26 Bedford Way, London, WC1H 0AP, UK.

References

Adams, W. J., Banks, M. S., & van Ee, R. (2001). Adaptation to three-dimensional distortions in human vision. *Nature Neuroscience, 4,* 1063–1064.

Akeley, K., Watt, S. J., Girshick, A. R., & Banks, M. S. (2004). A stereo display prototype with multiple focal distances. *ACM Transactions on Graphics, 23,* 804–813.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. New York: Springer-Verlag.

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

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436.

Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. [Research Support, U.S. Gov't, P.H.S. Review]. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 14,* 1393–1411.

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

Brenner, E., & Smeets, J. B. J. (2000). Comparing extra-retinal information about distance and direction. *Vision Research, 40,* 1649–1651.

Brenner, E., & Smeets, J. B. J. (2001). We are better off without perfect perception. *Behavioral and Brain Sciences, 24,* 215–216.

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

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

Brooks, R. A. (1991a). Intelligence without representation. *Artificial Intelligence, 47,* 139–159.

Brooks, R. A. (1991b). New approaches to robotics. *Science, 253,* 1227–1232.

Burge, J., Girshick, A. R., & Banks, M. S. (2010). Visual-haptic adaptation is determined by relative reliability. *Journal of Neuroscience, 30,* 7714–7721.

Butler, J. S., Smith, S. T., Campos, J. L., & Bulthoff, H. H. (2010). Bayesian integration of visual and vestibular signals for heading. *Journal of Vision, 10*(11):23, 1–13, http://www.journalofvision.org/content/10/11/23, doi:10.1167/10.11.23. [PubMed] [Article]

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

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

DeGroot, M. H. (1986). *Probability and statistics* (2nd ed.). Reading, MA: Addison-Wesley.

Domini, F., & Caudek, C. (2009). The intrinsic constraint model and Fechnerian sensory scaling. [Comparative Study Research Support, U.S. Gov't, Non-P.H.S.]. *Journal of Vision, 9*(2):25, 1–15, http://www.journalofvision.org/content/9/2/25, doi:10.1167/9.2.25. [PubMed] [Article]

Domini, F., Caudek, C., & Tassinari, H. (2006). Stereo and motion information are not independently processed by the visual system. *Vision Research, 46,* 1707–1723.

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. In G. Knoblich, I. M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Human body perception from the inside out* (pp. 105–131). New York: Oxford University Press.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision, 7*(5):7, 1–14, http://www.journalofvision.org/content/7/5/7, doi:10.1167/7.5.7. [PubMed] [Article]

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

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8,* 162–169.

Gepshtein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision, 5*(11):7, 1013–1023, http://www.journalofvision.org/content/5/11/7, doi:10.1167/5.11.7. [PubMed] [Article]

Girshick, A. R., & Banks, M. S. (2009). Probabilistic combination of slant information: Weighted averaging and robustness as optimal percepts. *Journal of Vision, 9*(9):8, 1–20, http://www.journalofvision.org/content/9/9/8, doi:10.1167/9.9.8. [PubMed] [Article]

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

Helbig, H. B., & Ernst, M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research, 179,* 595–606.

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

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision, 4*(12):1, 967–992, http://www.journalofvision.org/content/4/12/1, doi:10.1167/4.12.1. [PubMed] [Article]

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision, 8*(3):33, 1–30, http://www.journalofvision.org/content/8/3/33, doi:10.1167/8.3.33. [PubMed] [Article]

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 B: Biological Sciences, 265,* 1587–1593.

Howard, I. P. (2002). *Seeing in depth: Basic mechanisms* (vol. 1). Toronto, ON, Canada: I Porteous.

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

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

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? [Meeting Abstract]. *Perception, 36,* 14.

Knill, D. C. (2007). Learning Bayesian priors for depth perception. *Journal of Vision, 7*(8):13, 1–20, http://www.journalofvision.org/content/7/8/13, doi:10.1167/7.8.13. [PubMed] [Article]

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

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.

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

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

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.

MacKenzie, K. J., Murray, R. F., & Wilcox, L. M. (2008). The intrinsic constraint approach to cue combination: An empirical and theoretical evaluation. *Journal of Vision, 8*(8):5, 1–10, http://www.journalofvision.org/content/8/8/5, doi:10.1167/8.8.5. [PubMed] [Article]

MacNeilage, P. R., Banks, M. S., Berger, D. R., & Bulthoff, H. H. (2007). A Bayesian model of the disambiguation of gravitoinertial force by visual cues. *Experimental Brain Research, 179,* 263–290.

Maloney, L. T. (2002). *Statistical decision theory and biological vision*. New York: Wiley.

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

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

Milner, A. D., & Goodale, M. A. (2006). *The visual brain in action* (2nd ed.). Oxford, UK: Oxford University Press.

Muller, C. M., Brenner, E., & Smeets, J. B. (2009). Maybe they are all circles: Clues and cues. *Journal of Vision, 9*(9):10, 1–5, http://www.journalofvision.org/content/9/9/10, doi:10.1167/9.9.10. [PubMed] [Article]

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

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

Rosas, P., Wagemans, J., Ernst, M. O., & Wichmann, F. A. (2005). Texture and haptic cues in slant discrimination: Reliability-based cue weighting without statistically optimal cue combination. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 22,* 801–809.

Saunders, J. A., & Knill, D. C. (2003). Humans use continuous visual feedback from the hand to control fast reaching movements. *Experimental Brain Research, 152,* 341–352.

Saunders, J. A., & Knill, D. C. (2004). Visual feedback control of hand movements. *Journal of Neuroscience, 24,* 3223–3234.

Saunders, J. A., & Knill, D. C. (2005). Humans use continuous visual feedback from the hand to control both the direction and distance of pointing movements. *Experimental Brain Research, 162,* 458–473.

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

Scarfe, P., & Hibbard, P. B. (2006). Disparity-defined objects moving in depth do not elicit three-dimensional shape constancy. *Vision Research, 46,* 1599–1610.

Seydell, A., Knill, D. C., & Trommershauser, J. (2010). Adapting internal statistical models for interpreting visual cues to depth. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Journal of Vision, 10*(4):1, 1–27, http://www.journalofvision.org/content/10/4/1, doi:10.1167/10.4.1. [PubMed] [Article]

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. [Research Support, Non-U.S. Gov't]. *Trends in Cognitive Sciences, 14,* 425–432.

Smeets, J. B., & Brenner, E. (2008). Why we don't mind to be inconsistent. In P. Cavo & T. Gomila (Eds.), *Handbook of cognitive science—An embodied approach* (pp. 207–217). Amsterdam: Elsevier.

Tassinari, H., & Domini, F. (2008). The intrinsic constraint model for stereo–motion integration. *Perception, 37,* 79–95.

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived 3-dimensional structure-from-motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 663–678.

Todd, J. T. (1998). Theoretical and biological limitations on the visual perception of three-dimensional structure from motion. In T. Watanabe (Ed.), *High-level motion processing—Computational, neurophysiological and psychophysical perspectives* (pp. 359–380). Cambridge, MA: MIT Press.

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

Todd, J. T., Chen, L., & Norman, J. F. (1998). On the relative salience of Euclidean, affine, and topological structure for 3-D form discrimination. *Perception, 27,* 273–282.

Todd, J. T., Christensen, J. T., & Guckes, K. C. (2010). Are discrimination thresholds a valid measure of variance for judgments of slant from texture? *Journal of Vision, 10*(2):20, 1–18, http://www.journalofvision.org/content/10/2/20, doi:10.1167/10.2.20. [PubMed] [Article]

Todd, J. T., & Norman, J. F. (2003). The visual perception of 3-D shape from multiple cues: Are observers

capable of perceiving metric structure? *Perception & Psychophysics, 65,* 31–47.

Todd, J. T., Tittle, J. S., & Norman, J. F. (1995). Distortions of 3-dimensional space in the perceptual analysis of motion and stereo. *Perception, 24,* 75–86.

van Ee, R., van Dam, L. C., & Erkelens, C. J. (2002). Bi-stability in perceived slant when binocular disparity and monocular perspective specify different slants. [Research Support, Non-U.S. Gov't]. *Journal of Vision, 2*(9):2, 597–607, http://www.journalofvision.org/content/2/9/2, doi:10.1167/2.9.2. [PubMed] [Article]

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

Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision, 5*(10):7, 834–862, http://www.journalofvision.org/content/5/10/7, doi:10.1167/5.10.7. [PubMed] [Article]

Watt, S. J., Akeley, K., Girshick, A. R., & Banks, M. S. (2005). Achieving near-correct focus cues in a 3-D display using multiple image planes. *Proceedings of SPIE: Human Vision and Electronic Imaging, 5666,* 393–401.

Wichmann, F. A., & Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63,* 1293–1313.

Wichmann, F. A., & Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics, 63,* 1314–1329.