# Ideal cue combination for localizing texture-defined edges

**Michael S. Landy and Haruyuki Kojima**

*Department of Psychology and Center for Neural Science, New York University, 6 Washington Place, New York, New York 10003*

Received December 12, 2000; accepted April 13, 2001; revised manuscript received April 24, 2001

Many visual tasks can be carried out by using several sources of information. The most accurate estimates of scene properties require the observer to utilize all available information and to combine the information sources in an optimal manner. Two experiments are described that required the observers to judge the relative locations of two texture-defined edges (a vernier task). The edges were signaled by a change across the edge of two texture properties [either frequency and orientation (Experiment 1) or contrast and orientation (Experiment 2)]. The reliability of each cue was controlled by varying the distance over which the change (in frequency, orientation, or contrast) occurred—a kind of "texture blur." In some conditions, the position of the edge signaled by one cue was shifted relative to the other ("perturbation analysis"). An ideal-observer model, previously used in studies of depth perception and color constancy, was fitted to the data. Although the fit can be rejected relative to some more elaborate models, especially given the large quantity of data, this model does account for most trends in the data. A second, suboptimal model that switches between the available cues from trial to trial does a poor job of accounting for the data. 

*OCIS codes:* 330.4060, 330.5000, 330.5510, 330.6100.

## 1. INTRODUCTION

Many visual estimation problems require the observer to combine information from multiple sources. The obvious example of this is the perception of depth, where the sources of information are regarded as independent depth cues, and more than a dozen such cues have been described.[1] As a prerequisite to the perception of metric depth, many depth cues also require an estimate of the fixation distance, and this may also be estimated by using a number of different retinal and extraretinal cues (Ref. 2, for example). The segregation of figure and ground and both the detection and localization of object boundaries may also be viewed as problems of combining information from different sources. Object boundaries typically arise between surfaces differing in one or more surface qualities or differences in object geometry or motion. These manifest themselves in the retinal images as differences across the boundary in luminance, color, texture, disparity, and local motion. In this paper we will examine whether localization of texture edges may be successfully modeled as a problem of cue integration and whether, as in the depth cue combination case,[3] the cue integration strategy shares characteristics of the ideal observer.

### A. Theories of Cue Combination in Depth Perception

In computer vision, one is often concerned with how information is combined from multiple sources and, in particular, when those sources of information involve different sensors of the environment. In that field, the combination problem is referred to as data or sensor fusion. Clark and Yuille[4] discuss the data fusion problem in great detail and describe a hierarchy of solutions to it. The simplest scheme is weak fusion, where independent data sources are each used to make a separate estimate of a property of the environment, and then these estimates are combined, e.g., by some form of average. At the opposite extreme is strong fusion, where information sources can interact in arbitrarily complex ways to achieve an estimate. The weak fusion scheme corresponds, to some extent, with the ideas of Marr,[5] who suggested that different depth modules (structure-from-motion, shape-from-shading, binocular stereopsis, etc.) provide independent estimates of depth, shape, and distance, and these are combined at the level of the $2\frac{1}{2}$-D sketch. The strong fusion scheme has been used by, e.g., Aloimonos and Shulman,[6] who describe a number of algorithms that combine two or more depth cues in idiosyncratic ways to make up for the deficiencies in the use of the individual cues, leading to nonmodular algorithms, for example, shape from stereo-and-motion.

There are certainly good reasons for the visual system to treat cues in a modular fashion. Using depth perception again as an example, different depth cues provide different kinds of information. Some yield only ordinal assertions about depth, some give relative depth, and some give fully metric depth. Some lead naturally to estimates of depth while others provide local surface orientation. Furthermore, the reliability and even the availability of individual cues varies from scene to scene and even at different locations within a scene. All of these observations make it logically simpler to organize the system modularly, a method in which cues, when available, are used to make estimates of scene characteristics. These separate estimates can then be combined, taking into account viewing conditions, which cues are available in the scene, and the cues' reliabilities.

Considerations such as these led Landy and colleagues[3,7] to propose a modified weak fusion (MWF)

 

model for depth cue combination. It begins with an assumption of modularity and a linear cue combination rule or weighted average (i.e., weak fusion). But then, using normative considerations, the authors point out that the cue weights should depend on relative cue reliability (more reliable cues should get higher weight) and on the discrepancy between estimates from different cues (to be statistically robust, outlier estimates should be downweighted). Finally, cues do not provide commensurate information and must be promoted to a common format (e.g., scaled by the fixation distance) before averaging is even a meaningful operation. This scaling may involve additional retinal or extraretinal information or may involve interactions between depth cues.

In our previous work, we have seen demonstrations of all of these behaviors including linear cue combination with cue weights dependent on relative reliability,[8,9] statistical robustness,[10] and cue interactions for scaling.[2,9,11] One might be concerned with how a model such as MWF is implemented in the brain. Fine and Jacobs[12] have tried to implement it using a neural network. The study described in the current paper will be an application of the MWF model to judgments of spatial localization.

The modified weak fusion model is certainly not the only theory that has been put forward to explain how cues are combined. In its essence, it treats the observers as if they were statisticians confronted with a problem in estimation. A recent trend in perceptual psychology is instead to describe behavior in the terms of Bayesian estimation (Ref. 13, for example). If the cues are treated as statistically independent and the individual distributions are Gaussian (Ref. 14, in which this is termed a weakly coupled system), the resulting Bayesian calculation is indistinguishable from the MWF calculation. With different distributions and without independence assumptions, however, Bayesian calculations can be arbitrarily complex (i.e., result in forms of strong fusion).

Another model for cue combination is the fuzzy logical model of perception of Massaro and Friedman.[15] This model begins with the language of fuzzy logic for describing support for individual propositions. Propositions are combined by multiplication (rather than by the weighted sum of MWF). In most of its applications, the competing alternatives are discrete and finite (e.g., alternative perceptual organizations, competing phonemic interpretations of a speech utterance), and the multiplication operation coincides with the Bayesian rule of combination. Thus one can cast this as yet another version of the same combination scheme, although that depends on how it is applied, and there has been an ongoing debate on the relative efficacy of the fuzzy logical model of perception versus linear cue combination rules.[16–18]

### B. Cue Combination for Other Estimation Problems

We suggest that information fusion schemes such as MWF and Bayesian methods have general applicability in vision beyond the calculation of depth and scene geometry. We will briefly review some recent studies of this sort in color and spatial vision.

#### 1. Color Constancy

Recently both MWF and the empirical techniques used to estimate cue weights (perturbation analysis; see Section 2) were applied to the problem of estimating the spectral power distribution of the illuminant as a step toward achieving color constancy.[19,20] In that work, different illuminant cues (the background and various cues involving specular highlights or shadow boundaries) are used independently to estimate the illuminant, and MWF appears to be applicable as a model of how cues to the illuminant are combined.

#### 2. Motion

The detection and analysis of motion involves combining both first-order (luminance and color) and second-order (contrast and other drift-balanced texture) information, and many studies suggest that they are detected by independent mechanisms.[21–23] Thus these separate motion components may be treated as individual cues to motion. These cues may be pooled for judging the direction of motion[24] and for computing depth from motion parallax.[25]

#### 3. Detection of Spatial Features

Borders and other spatial features may be defined by multiple visual features including luminance, color, texture content, and so on. The detection of such features may involve integration over these features. If the features are detected independently, then combination may be via probability summation, where a feature is detected if any individual cue signals the feature. For example, Frome *et al.*[26] examined the fusion of color and luminance information for the visibility of borders and found evidence for independent contributions of the two cues to border visibility.

#### 4. Orientation Discrimination

Once a border is detected, features of that border such as orientation may then be computed. Rivest *et al.*[27] studied orientation discrimination with texture bars defined by motion, color, and/or luminance. When subjects were trained to discriminate the orientation of borders defined by one of the three attributes (e.g., motion), orientation discrimination performance improved for motion-defined borders but also improved for color-defined and for luminance-defined borders. This indicated that the training affected a site at which the three cues had already been combined. They found that discriminability with multiple cues combined the individual $d'$ values in Euclidean fashion, consistent with cue independence.

#### 5. Localization

Multiple cues may also be combined for localization judgments, which is the topic of the research described in this paper. In a preliminary version of the research described here,[28] stimuli included two cues to a texture-defined edge. Subjects performed a Vernier task, judging whether the uppermost of two such edges lay to the right or left of the lower edge. The data were consistent with the notion that subjects estimated edge location from each cue individually and then averaged the two estimates for each edge. In addition, lowered cue reliability in the stimulus resulted in a lower weight for the corresponding cue. Both of these results are, of course, consistent with MWF.

Rivest and Cavanagh[29] combined multiple edge cues in dynamic texture, including luminance, color, motion, and scale information. When the two cued edges were reasonably far apart (and hence discriminable as different edges), the two edges interacted, leading to both attraction and repulsion (Ref. 30, where this was found for luminance edges). When the multiple cues signaled identical locations, localization accuracy improved as if two noisy estimates were averaged with equal weights. Note that this paper did not consider the possibility of cue combination with different weights for the individual cues.

Gray and Regan[31] used texture stimuli similar to those of Experiment 2 below. They also found that localization accuracy improved when two cues (texture orientation and contrast) were available as compared with accuracy for only a single stimulus cue. They did not model the improvement in accuracy quantitatively.

McGraw *et al.*[32] used a combination of a smooth luminance blob and a second-order texture blob (black and white dots windowed by a smooth blob) in a Vernier task. Different, asymmetric blobs were used for the two cues, so that the perceived location of the luminance blob alone was shifted relative to that for the second-order texture blob. The perceived location of the two-cue stimulus lay in between the locations for the individual cues, indicating that cue combination had taken place. The perceived location of the two-cue stimulus lay closer to the location of the cue that was more reliable (i.e., that had higher stimulus contrast). As the discrepancy between the locations signaled by the two cues increased, a loss of accuracy was found. This result is not expected if cue weights are constant in their task, independent of the position signaled by each edge, and it may be an indication of a statistically robust combination scheme that changes cue weights as cue conflict increases.

### C. Preview

In this paper we expand on the preliminary results of Landy[28] on cue combination for texture edge localization. We report two experiments involving Vernier alignment judgments of texture-defined edges. The first experiment uses filtered-noise textures in which an edge is signaled by a change in local spatial frequency and/or orientation. As found previously, perceived edge location was a weighted average of the locations indicated by the individual cues, with weights dependent on relative cue reliability, consistent with the MWF model. Experiment 2 is a replication using the same logic but different stimuli. The textures for that experiment contain randomly positioned texture elements (texels) varying in contrast and/or orientation across the edge, and similar results were obtained. The results of these two experiments are then analyzed on the basis of a version of the MWF model that assumes that observers accurately estimate individual cue reliabilities. This is an ideal-observer model that we refer to as the optimal linear cue combination model.

## 2. EXPERIMENT 1: SPATIAL FREQUENCY AND ORIENTATION

In this experiment, observers made a forced-choice Vernier judgment about two texture-defined edges, one above the other. The textures consisted of noise processed by a filter that was bandpass in spatial frequency and orientation. The texture-defined edge was cued by a change in peak spatial frequency and/or orientation across the edge. The reliability of the individual cues was manipulated by varying the texture blur, that is, the distance in the stimulus over which the change in spatial frequency or orientation occurred.[33] Vernier judgments were made with single-cue stimuli (spatial frequency *or* orientation) to estimate the reliability (that is, accuracy) of the individual cues. Then, judgments were made with two-cue stimuli (spatial frequency *and* orientation) with perturbation analysis.[3]

Perturbation analysis was first used for relative depth judgments[3,8,9] and has also been used to study interpolation of sampled contours.[34] Here it is easily adapted to localization.

In perturbation analysis, two kinds of two-cue stimuli are used: consistent-cues stimuli, in which the two cues signal the same edge location, and inconsistent-cues stimuli, in which the position of the edge signaled by one cue is perturbed (that is, translated laterally) relative to that signaled by the other cue. If the perturbation is large enough, then two edges are perceived and the task of aligning with *the* edge in the stimulus becomes ill-defined. Rather, the experiments are performed by using a range of cue perturbations for which observers do not perceive a double edge without scrutiny. Using a Vernier task, we measured the perceived location of an inconsistent-cues edge relative to a consistent-cues edge as a function of the degree of perturbation $\Delta$cue (the amount that one cued edge is translated relative to the other in the inconsistent-cues stimulus).

Suppose edge location $x$ is estimated by using a weighted average of the estimated locations of the individually cued edges $x_1$ and $x_2 = x_1 + \Delta\text{cue}$, with weights $w_1$ and $w_2 = 1 - w_1$. The perceived location of the inconsistent-cues edge, $f(\Delta\text{cue}) = w_1x_1 + w_2x_2 = x_1 + w_2\Delta\text{cue}$. This is a linear function of $\Delta$cue. The slope of that line is the weight applied to the perturbed cue $w_2$ and should be lower if that cue has lower reliability. As we shall see, these predictions are, in fact, borne out by the data of the following experiments.

### A. Methods

#### 1. Stimuli

Stimuli consisted of filtered noise. An edge was cued by a change in spatial frequency [Figs. 1(a) and 1(b)], orientation [Figs. 1(c) and 1(d)], or both [Fig. 1(e)]. When spatial frequency cued an edge, spatial frequency increased from 3 to 6 cycles per degree (cpd) from left to right. In the single-cue spatial frequency edges, two noise orientations were used in the experiments: vertical [Fig. 1(a)] and horizontal [Fig. 1(b)]. When change in orientation was used to cue an edge, orientation was horizontal at the left and rotated smoothly to vertical at the right. In the single-cue orientation-defined edges, two noise spatial frequencies were used: 3 cpd [Fig. 1(c)] or 6 cpd [Fig. 1(d)]. For two-cue stimuli [Fig. 1(e)], spatial frequency and orientation were varied across their respective edge locations as in the single-cue stimuli.

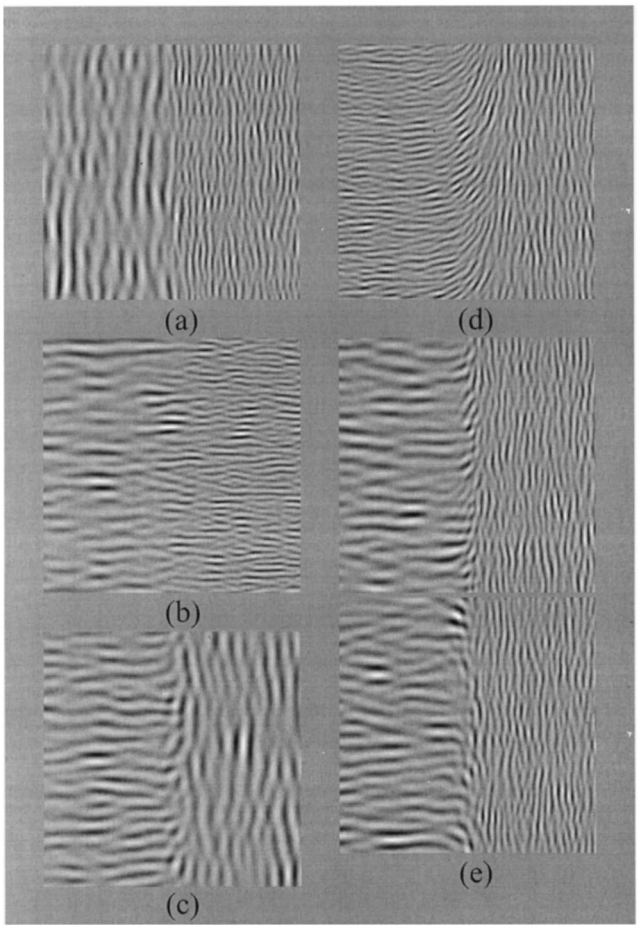


Fig. 1. Example stimuli from Experiment 1. (a) An edge cued by a change in spatial frequency from 3 to 6 cpd with texture blur $\sigma_f = 9$ min and vertically oriented filtered noise. (b) $\sigma_f = 36$ min, horizontal noise. (c) Orientation-defined edge: $f = 3$ cpd, $\sigma_o = 9$ min. (d) $f = 6$ cpd, $\sigma_o = 36$ min. (e) A typical stimulus from a two-cue experiment. Upper texture, a consistent-cues stimulus; lower texture, $\Delta$cue = 13.5 min (i.e., the edge cued by spatial frequency is 13.5 min to the right of the orientation-defined edge). Both stimuli have $\sigma_f = \sigma_o = 9$ min. Observers judged whether the upper edge was to the right or left of the lower edge.

Single-cue stimuli were generated as follows. First, a noise stimulus of size 2048 × 600 pixels was created, with each pixel a random and independent draw from a Gaussian distribution. Next, an edge profile was created. For example, consider the case of an orientation-defined edge with spatial frequency 6 cpd. An image of the same size as the noise was created with value 0 deg in the left half of the image (indicating "horizontal") and 90 deg in the right half (indicating "vertical"). This edge image was then blurred by a Gaussian with standard deviation $\sigma_o$. As we shall see, the value of $\sigma_o$ determined the reliability of the orientation cue; it acted as a kind of texture blur. Next, the Gaussian noise image was filtered nonisoplanatically. That is, a different filter was used at each pixel as guided by the blurred edge image. For the case that we are first considering, the filter used was a Gabor patch:

$$G(x,y) = \frac{1}{f}\exp(-f^2r^2)\sin\{2\pi f[x\cos(\theta) + y\sin(\theta)]\}, \quad (1)$$

where $(x,y)$ is the position in the Gabor patch, $r = \sqrt{x^2 + y^2}$ is the distance from the center of the patch, $f$ is the spatial frequency, and $\theta$ is the orientation. The contrast of the patch is scaled by $1/f$ so that the rms contrast of the filtered noise is independent of spatial frequency. The Gaussian width is also scaled by $1/f$ so that the filters have a constant bandwidth in octaves. The filter masks were clipped at ±2.8 standard deviations of the Gaussian window. For the orientation-defined edge that we are considering, $f = 6$ cpd, and $\theta$ for a given pixel is given by the corresponding pixel in the blurred edge image. Each pixel of the Gaussian-noise image was used to scale the contrast of the corresponding Gabor patch (centered on that pixel), and these Gabor patches were then summed to produce a filtered-noise image. This algorithm ensures that no artifactual luminance edges in the stimulus result from the rapid changes in filter across the edge.[33,35] The filtered-noise images were then linearly scaled so that the mean luminance corresponded to a pixel value of 128, and the full contrast range was utilized (in fact, fewer than 0.01% of the pixels were clipped at pixel values of 0 and 255). Finally, the central 2048 × 512 pixels were saved as a stimulus strip to be used during the experiments.

All other stimuli were generated analogously. For edges cued by spatial frequency, the orientation $\theta$ remained constant across the stimulus (with a value of 0 or 90 deg), and spatial frequency was determined by a blurred edge image with a value of 3 cpd at the left, 6 cpd on the right, and a blur standard deviation of $\sigma_f$. Two-cue stimuli were generated by using a separate blurred edge image to determine the local spatial frequency and orientation. The orientation edge was always located at the center of the stimulus strip. The spatial frequency edge, for inconsistent-cues stimuli, was shifted laterally by a distance $\Delta$cue [Fig. 1(e)]. The nominal position of the two-cue stimuli will be the position of the orientation edge in the data plots that follow.

On each trial, two stimuli were shown, one above the other. These stimuli were 400 × 400 pixel squares taken from the corresponding stimulus strip. These were shown 7.2 min apart on the monitor, and each had a size of 6 × 6 deg as viewed from 1 m. The stimuli were computed with the HIPS image processing software[36] and displayed with a Pixar II Image Computer on a Sony GDM-1950 monitor with 1280 × 1024 resolution at 60 Hz. The background luminance was 35 cd/m$^2$, and luminance was linearized with a Minolta CS-100 chromometer.

*2. Procedure*

The subject's task was a Vernier discrimination. Each trial consisted of a 1-s fixation interval, a 0.5-s blank interval, and a 200-ms presentation of the Vernier stimulus. The subject indicated by a key press whether the upper texture-defined edge appeared to lie to the right or the left of the lower edge. The screen remained blank until the subject's response, which also initiated the subsequent trial. The lower stimulus was mirror reflected about a horizontal axis so that orientation change was counter-

clockwise from left to right in the upper stimulus and clockwise below [Fig. 1(e)]. This reflection was intended to minimize a pronounced alignment bias that otherwise occurred owing to the apparent texture flow from the lower to the upper patch, although no such bias was found in another study using similar stimuli.[37] Interleaved staircases were used to determine the position $\Delta x$ of the upper edge relative to the lower one (to a resolution of 4 pixels or 3.6 min). Staircases began with an increment of 14.4 min, which was halved with staircase reversals until it reached the 3.6 min level. The midpoint between the location of the two edges was jittered from trial to trial $\pm 10.8$ min about the center of the patch. A $400 \times 400$ patch of texture was extracted from the appropriate stimulus strip for each of the two images used in each trial. The horizontal position was determined by the value of $\Delta x$ and the edge location jitter. The vertical position of the patch in the strip was selected randomly.

A block of trials in a single-cue experiment consisted of a single value of texture blur ($\sigma_o$ or $\sigma_f$). Each trial required a Vernier discrimination between two different, single-cue edge stimuli from the same condition (sharing the same value of $\sigma_o$ or $\sigma_f$ and of $f$ or $\theta$), which were different, randomly chosen $400 \times 400$ patches from the same stimulus strip. For the orientation cue blocks, there were four interleaved staircase combining two types of staircase (a 1-up–2-down and a 2-up–1-down, converging to probabilities of 0.707 and 0.293 of saying the upper edge was to the right, respectively) and two values of $f$ (3 and 6 cpd). Two stimulus strips were used in one block (one for each value of $f$). The spatial frequency cue blocks were analogous, except that two values of $\theta$ (vertical and horizontal) were interleaved. Each staircase ran for 50 trials, for a total of 200 trials per block.

A block of trials in a two-cue experiment consisted of a single pair of values of texture blur ($\sigma_o$ and $\sigma_f$). Each block of trials had six or eight interleaved staircases, combining the two types of staircases with either three ($-9$, 0, or 9 min) or four ($-13.5$, $-4.5$, 4.5, or 13.5 min) values of $\Delta$cue. This resulted in blocks of 300 or 400 trials. Each trial required a vernier discrimination between a consistent-cues stimulus ($\Delta\text{cue} = 0$) and an inconsistent-cues stimulus ($\Delta$cue determined by the staircase from which the trial was drawn). Whether the inconsistent-cues stimulus was in the upper or lower patch was determined randomly. The staircase level determined the relative position of the orientation-defined edges in the two stimuli; the edge cued by spatial frequency was displaced a distance $\Delta$cue away from that position in the inconsistent-cues stimulus. Again, the midpoint between the locations of the two orientation-defined edges was jittered $\pm 10.8$ min about the center of the patch. One stimulus strip was used for each value of $\Delta$cue in the block, plus one for the consistent cues stimuli ($\Delta\text{cue} = 0$) if needed.

The blocks of trials were kept relatively short, resulting in the complex blocking of the inconsistent-cues stimuli. With longer blocks, the task grew difficult for observers. These stimuli are potent adapting stimuli as they are high contrast and narrow band in both spatial frequency and orientation. With accurate fixation, each patch of retina, except those near the position of the edge, is repeatedly stimulated with the same bands of orientation and spatial frequency. The stimuli are high contrast, so the adaptation does not cause them to become difficult to detect as such. Rather, after a large number of trials it appears increasingly difficult to detect the texture-defined edge itself. The block lengths used here were intended to alleviate this difficulty.

Each subject ran forty blocks of trials. First, eight blocks of single-cue trials were run combining two cues (orientation and spatial frequency) and two values of texture blur ($\sigma_o$ or $\sigma_f$ equal to 9 or 36 min), with two blocks for each combination. These first blocks were run with feedback to familiarize the subjects with the materials and to minimize any response biases. These eight single-cue blocks were then rerun with no feedback to the subjects. Only the results from the no-feedback blocks were analyzed. Finally, the two-cue blocks were run. First, eight blocks were run with $\sigma_o = \sigma_f = 9$ min (four blocks with three, and four blocks with four values of $\Delta$cue). Next, eight blocks were run with $\sigma_o = 36$ min and $\sigma_f = 9$ min. Finally, eight blocks were run with $\sigma_o = 9$ min and $\sigma_f = 36$ min.

*3. Subjects*

Two subjects participated in the experiment. Both had normal or corrected-to-normal vision.

## B. Results

Figure 2 shows the data for the four single-cue conditions for observer MJY. Plotting raw data can be a problem when a staircase method is used, as some levels may provide only a small number of trials (and hence yield highly variable data). Here, we indicate this by using larger symbols for data points corresponding to a greater number of trials. In Fig. 2(a) cumulative Gaussian functions are fitted to the data from each condition with a maximum-likelihood (ML) criterion. When texture blur was increased in the stimulus, localization accuracy suf-

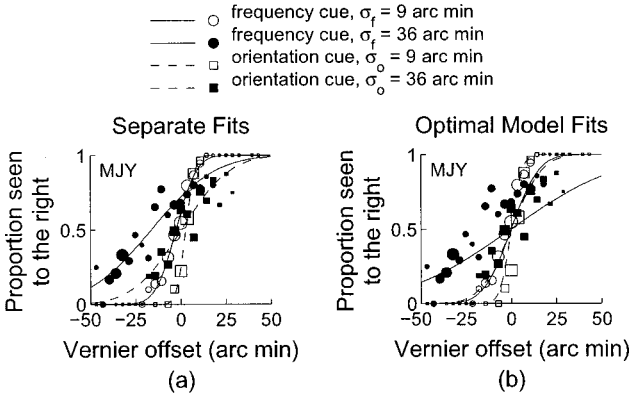


Fig. 2. Results from single-cue trials for Experiment 1 for observer MJY. Data from each of the four single-cue conditions are plotted. The proportion of times the upper edge was seen as lying to the right of the lower edge is plotted as a function of the distance of actual vernier shift. Symbol area is proportional to the number of trials contributing to each point. The same data are plotted in the two panels. (a) The solid curves are ML fits of cumulative Gaussian functions to individual psychometric functions. (b) The solid curves are the ML fits of the optimal linear cue combination model to the entire Experiment 1 data set for this observer.

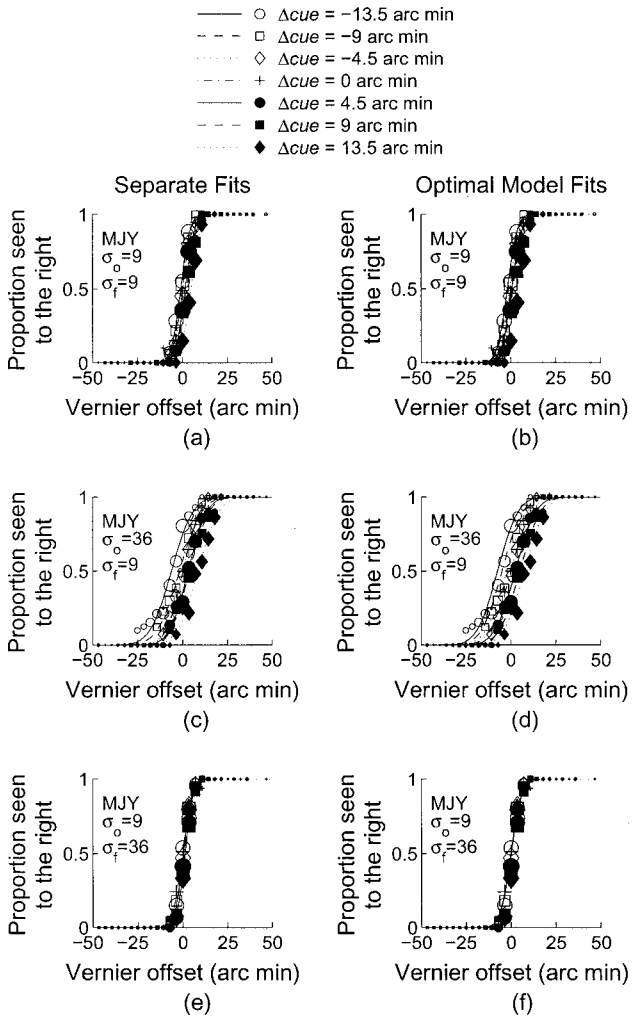


Fig. 3. Results from two-cue trials for Experiment 1 for observer MJY. Data from each of the 21 two-cue conditions are plotted. The proportion of times the consistent-cues edge was seen as lying to the right of the inconsistent-cues edge is plotted as a function of the distance of actual vernier shift between the orientation-defined edges. The separate data sets in each plot are for different values of Δcue: the shift of the frequency-defined edge relative to the orientation-defined edge in the inconsistent-cues stimulus. The symbol area is proportional to the number of trials contributing to each point. The same data are plotted in each pair of panels. Each row of panels corresponds to a different cue reliability condition (indicated in each figure). (a), (c), (e) The curves are ML fits of cumulative Gaussian functions to individual psychometric functions. (b), (d), (f) The curves are ML fits of the optimal linear cue combination model to the entire Experiment 1 data set for this observer.

fered (compare the slopes of the two solid curves or of the two dashed curves), so the localization reliability manipulations $\sigma_f$ and $\sigma_o$ were effective. This observer evidently had difficulty with the $\sigma_f = 36$ min condition; the data points are highly variable, and there was a substantial bias to indicate that the upper edge was located to the right of the lower edge. At both levels of texture blur, this observer has better localization accuracy with the orientation cue than with the frequency cue. This and subsequent data figures also include fits of a model that will not be discussed until Subsection 4.A.

The results of the two-cue conditions for MJY are shown in Fig. 3. The abscissa indicates the relative location of the orientation edges in the two texture stimuli; in the inconsistent-cues stimulus the frequency edge was displaced Δcue away from the corresponding orientation edge. This observer was more accurate in general when using the unperturbed, orientation cue. As a result, the perturbation had only a modest effect on the psychometric functions. Although it is difficult to see in this figure, the order of the psychometric functions reflects that of the cue perturbations, consistent with the idea that the observer perceived an edge location for the inconsistent-cues stimulus intermediate between the location of the orientation-defined and the frequency-defined edges (i.e., as if the edge locations signaled by the individual cues were averaged).

The psychometric functions are closest when orientation is the more reliable cue in the stimulus [Fig. 3(e)], most separated when frequency is the more reliable cue allowing the perturbation to have an effect [Fig. 3(c)], and intermediate for the case when $\sigma_f = \sigma_o$ [Fig. 3(a)]. This is consistent with a weighted average of cues with more weight given to the more reliable cue. In addition, the psychometric functions in this dataset are, by and large, parallel. This would be expected if the observer were us-

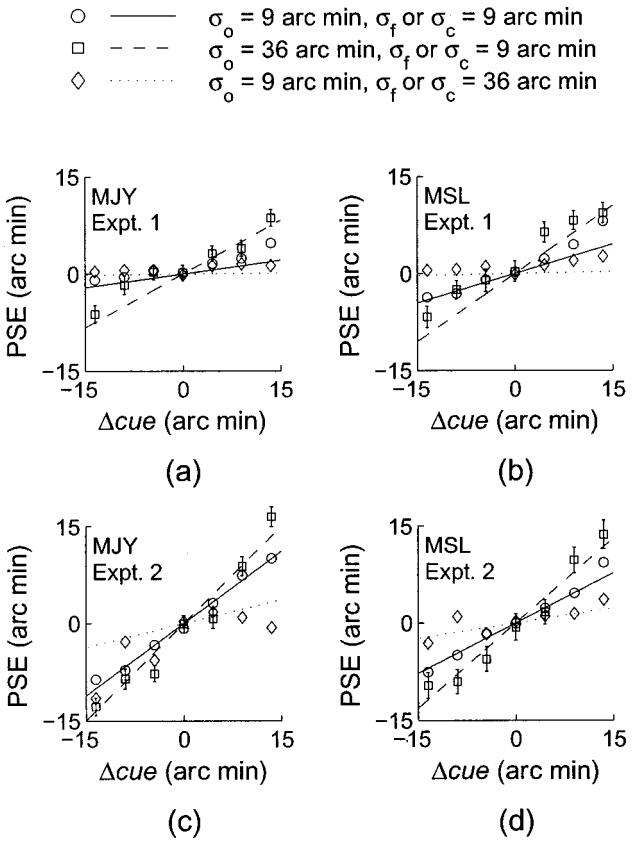


Fig. 4. PSEs as a function of Δcue. PSEs are estimated as the 50% points of individual cumulative Gaussian fits to the 21 two-cue conditions in each experiment (e.g., left-hand panels of Fig. 3). Error bars shown on a subset of the data are ±2 standard error of the mean, computed via the matrix of second derivatives of the log-likelihood function with respect to the fit parameters.[40,43] Lines are from the fits of the optimal linear cue combination model to the entire data set for each observer and experiment.

ing the same combination rule and the same cue weights across all values of Δcue within a given cue reliability condition.

These shifts of the psychometric functions with different amounts of cue perturbation may be seen more clearly in Fig. 4. Here, the points of subjective equality (PSEs) estimated by the individual psychometric function fits (the 50% points of each curve in the left-hand panels of Fig. 3) are plotted as a function of the cue perturbation. Figures 4(a) and 4(b) show the results for the two observers in Experiment 1. The PSE is approximately a linear function of the cue perturbation. The linearity of the PSEs is noisy, and the deviations from a straight line show no discernible pattern. The slope of each line is the weight predicted for the perturbed cue and varies as one would expect given the relative cue reliabilities. The slopes of the PSEs for the condition in which $\sigma_f = 9$ min and $\sigma_o = 36$ min (open squares) are steepest, indicating that the perturbation of the frequency-defined edge location had maximal effect when that cue was most reliable.

## 3. EXPERIMENT 2: CONTRAST AND ORIENTATION

A second experiment was carried out as an attempt to replicate the results with a different class of stimuli and cues. The stimuli consisted of randomly placed short line segments. The edges were cued by changes in line orientation and contrast. Such textures with line segment texture elements and orientation modulation have been much studied in our lab and others.[38,39] Otherwise, this experiment was identical in conditions and procedure to Experiment 1.

### A. Methods

*1. Stimuli*

As in Experiment 1, stimuli were $400 \times 400$ pixels extracted from a $2048 \times 512$ pixel stimulus strip. The stimulus strips consisted of 7,550 blurred, white line segments varying in orientation and contrast on a midgray background (Fig. 5). The position of each segment was chosen randomly and uniformly over the strip. Line segments were computed by using a grid with five times the resolution in both dimensions as the stimulus image for the purposes of antialiasing and subpixel location resolution. Segments were 9 min long and blurred by a Gaussian with a standard deviation of 0.9 min.

For an orientation-defined edge [Fig. 5(a)], line segment orientation was determined with the same edge image as in Experiment 1: a scaled cumulative Gaussian with standard deviation $\sigma_o$ changing from vertical on the left side of the image and rotating smoothly to horizontal on the right. The contrast-defined edges were analogous: with an edge image with standard deviation $\sigma_c$, with line segments of 50% contrast on the left and 100% contrast on the right [Fig. 5(b)]. As before, the lower of the two stimuli on each trial was reflected about a horizontal axis [Fig. 5(c)]. All other aspects of the stimuli (hardware, viewing distance, stimulus size and geometry, display timing and calibration, etc.) were identical to those of Experiment 1.

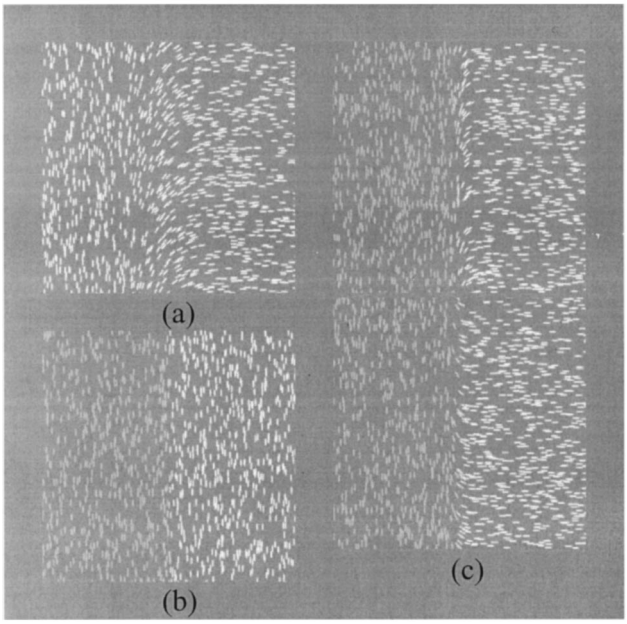


Fig. 5. Example stimuli from Experiment 2. (a) Orientation-defined edge with 100% contrast and $\sigma_o = 36$ min. (b) Contrast-defined edge with vertical line segments and $\sigma_c = 9$ min. (c) A typical stimulus from a two-cue experiment. Lower texture, consistent-cues stimulus; upper texture, Δcue $= -13.5$ min (i.e., the edge cued by contrast is 13.5 min to the left of the orientation-defined edge). Both stimuli have $\sigma_c = \sigma_o = 9$ min.

*2. Procedure*

The procedure was identical to that in Experiment 1, including the task and the order and blocking of practice trials, single-cue trials, and two-cue blocks (with the contrast cue substituted for the frequency cue of Experiment 1). For two-cue trials, as in Experiment 1, the position of the orientation-defined edge was controlled by the staircase, and the position of the edge defined by the other cue (here, the contrast-defined edge) was perturbed by Δcue from that position in the inconsistent-cues stimulus. $\sigma_o$ and $\sigma_c$ took on the same values and value pairs as $\sigma_o$ and $\sigma_f$ in Experiment 1.

*3. Subjects*

The same two subjects were used as in Experiment 1.

### B. Results

Figure 6 shows the results for the single-cue conditions for observer MSL. As in Experiment 1, the manipulations of cue reliability in the stimulus were effective in varying localization accuracy (compare the slopes of the two solid curves or the two dashed curves). The relative slopes indicate that the localization reliability manipulations $\sigma_o$ and $\sigma_c$ were about equally effective.

The two-cue condition results for MSL are shown in Fig. 7. As before, the abscissa indicates the relative location of the orientation edges in the two texture stimuli; in the inconsistent-cues stimulus the contrast edge was displaced Δcue away from the corresponding orientation edge. As in Experiment 1, all aspects of the data are reasonably consistent with the idea that the observer is localizing each two-cue edge by calculating a weighted av-

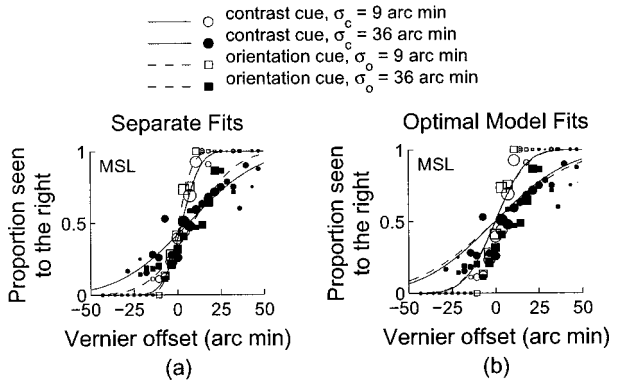


Fig. 6. Results from single-cue trials for Experiment 2 for observer MSL. Data from each of the four single-cue conditions are plotted. Plotting conventions and fits are identical to those of Fig. 2. (a) Independent cumulative Gaussian fits. (b) Fits of the optimal linear cue combination model to the entire Experiment 2 data set for this observer.

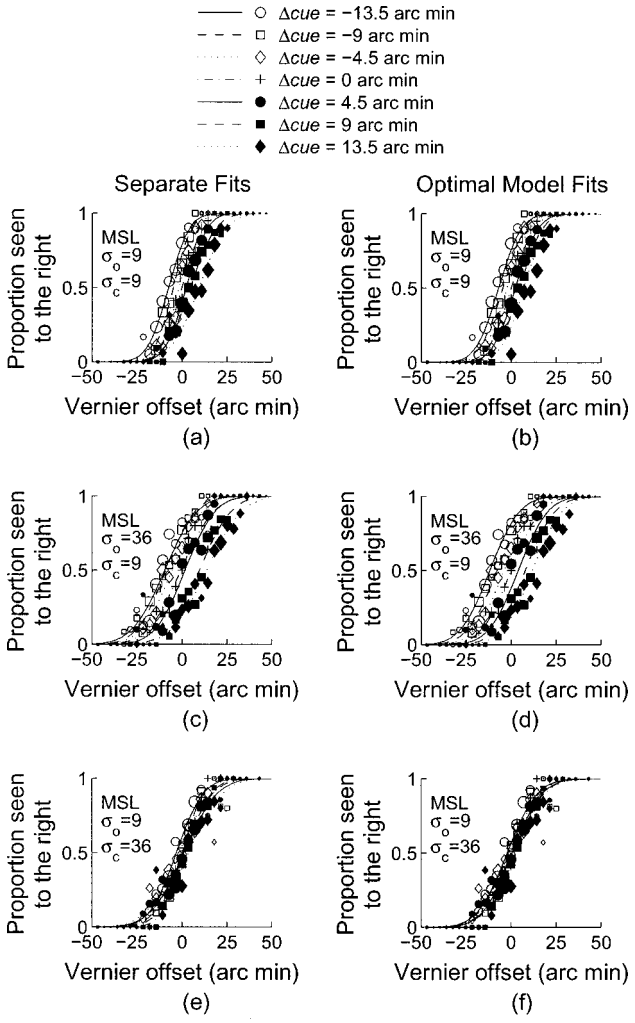


Fig. 7. Results from two-cue trials in Experiment 2 for observer MSL. Data from each of the 21 two-cue conditions are plotted. Plotting conventions and fit curves are as in Fig. 3. (a), (c), (e) The curves are ML fits of cumulative Gaussian functions to individual psychometric functions. (b), (d), (f) The curves are ML fits of the optimal linear cue combination model to the entire Experiment 2 data set for this observer.

erage of the locations signaled by each cue individually, with cue weights related to relative cue reliability. First, the psychometric functions shift with changes in the value of Δcue. Second, these shifts are ordered by the amount of Δcue. The more the edge cued by contrast was shifted to the right of the edge signaled by orientation, the more the perceived edge location of the inconsistent cues edge was shifted rightward, and hence the less often the consistent-cues edge was perceived as being to the right of the inconsistent-cues edge. Third, these psychometric functions are reasonably equally spaced, as would be expected since the values of Δcue are equally spaced. Fourth, the amount of shift is greater when the reliability of the perturbed cue is greatest [Fig. 7(c)] and least when it is lowest [Fig. 7(e)]. Finally, the psychometric functions have approximately equal slopes, consistent with the use of the same cue weights for all values of Δcue.

The plots of PSE as a function of Δcue for Experiment 2 for both observers may be seen in Figs. 4(c) and 4(d). The balanced reliabilities for the two single cues for observer MSL resulted in a contrast cue weight for the $\sigma_c = \sigma_o = 9$ min condition (the slope of the PSEs as a function of Δcue) intermediate between the other two conditions. Otherwise, these data are entirely analogous to those from Experiment 1. In short, the basic predictions of a linear cue combination are, at least qualitatively, apparent in the data of Experiment 2, as they were in Experiment 1, showing that these effects are not dependent on the particular choice of stimuli or edge cues.

## 4. MODELING

In this section we describe several models that were fitted to the data of Experiments 1 and 2. The bulk of the section is concerned with a fit of an optimal linear cue combination model, that is, modified weak fusion as applied to this texture localization task. After thoroughly exploring various aspects of the fits of this model, we will compare it with several alternative models.

### A. Optimal Linear Cue Combination Model

The modified weak fusion model suggests that observers combine cues in a linear fashion except under certain specific conditions. In our task, it suggests that observers determine an edge location independently for each cue and that these two location estimates are combined with a weighted average. Nonlinearities would arise, for example, if the cue weights depended on the locations signaled by the individual cues. This would be expected if, for example, there was such a strong conflict between the locations signaled by the individual cues that the observer decided to increase the weight of the more reliable of the two cues in an attempt to be more statistically robust. In cases of extreme conflict, of course, one would expect the observer to perceive two separate edges. In most of the analysis that follows, few traces of such nonlinear behavior are seen. In one case we will point out a hint of robust combination.

The optimal linear cue combination model, which we now describe, is a version of MWF that assumes that the individual location estimates determined by each cue are unbiased estimates, normally distributed and indepen-

dent, and have an accuracy that depends on the texture blur associated with that cue in the stimulus. Thus, just as our stimulus conditions have four parameters ($\sigma_o$ = 9 or 36 min, $\sigma_f$ = 9 or 36 min), the model has four parameters corresponding to the localization accuracy of a single-cue stimulus in each of these cue conditions. We designate these localization accuracies, which are the standard deviations of the corresponding Gaussian noises that perturb observer location estimates, as $\sigma_{\mathrm{cue}}$. A fit of this model to one observer's data in one experiment will result in four such noise estimates. For example, in Experiment 1, these are $\sigma_{\mathrm{cue}}(o, 9)$, $\sigma_{\mathrm{cue}}(o, 36)$, $\sigma_{\mathrm{cue}}(f, 9)$, and $\sigma_{\mathrm{cue}}(f, 36)$.

For any given choice of these four parameters, the optimal linear cue combination model has a prediction for all data points for that observer in that experiment. For the single-cue experiments the predictions are quite simple. A single-cue data point (see Figs. 2 and 6) is specified by a cue ($o$, $f$, or $c$), its texture blur (9 or 36 min), and the Vernier shift $\Delta x$ of the upper edge relative to the lower one. For example, the predicted probability $p(o, 9\ \mathrm{min}, 7.2\ \mathrm{min})$ that an orientation-cued edge with texture blur of 9 min and a rightward Vernier shift of the upper edge by 7.2 min is actually perceived as lying to the right of the lower edge will be

$$p(o, 9, 7, 2) = \Pr\{N[7.2, \sigma^2_{\mathrm{cue}}(o, 9)] > N[0, \sigma^2_{\mathrm{cue}}(o, 9)]\}, \tag{2}$$

where the $N(\mu, \sigma^2)$ are independent, normally distributed random variables with the specified mean and variance.

For the two-cue case, the logic is only slightly more complicated. This model is referred to as "optimal" because we assume that the observer uses cue weights that are optimal given the underlying localization reliabilities of the two cues. For example, if the localization accuracies of the individual cues are $\sigma_{\mathrm{cue}}(o, 9)$ and $\sigma_{\mathrm{cue}}(f, 9)$, then overall localization accuracy will be maximal if the weights used for each cue satisfy

$$w_f = \frac{1/\sigma^2_{\mathrm{cue}}(f, 9)}{[1/\sigma^2_{\mathrm{cue}}(f, 9)] + (1/\sigma^2_{\mathrm{cue}}(o, 9))],}$$

$$w_o = 1 - w_f. \tag{3}$$

The location estimator that uses these weights has an accuracy

$$\sigma^2_{\mathrm{opt}}(f, 9, o, 9) = \frac{\sigma^2_{\mathrm{cue}}(f, 9)\sigma^2_{\mathrm{cue}}(o, 9)}{\sigma^2_{\mathrm{cue}}(f, 9) + \sigma^2_{\mathrm{cue}}(o, 9)}. \tag{4}$$

A particular two-cue data point is specified by the condition [e.g. $(f, 9, o, 9)$], the value of $\Delta$cue (how far the frequency-defined edge is shifted relative to the orientation-defined edge in the inconsistent-cues stimulus), and the value of $\Delta x$ (how far the orientation-defined edge in the consistent-cues stimulus is shifted relative to the inconsistent-cues edge). The predicted probability that the consistent-cues edge is seen to the right of the inconsistent-cues edge is

$$p(f, 9, o, 9, \Delta\mathrm{cue}, \Delta x) = \Pr[N(\Delta x, \sigma^2_{\mathrm{opt}}) > N(w_f\Delta_{\mathrm{cue}}, \sigma^2_{\mathrm{opt}})]. \tag{5}$$

Note that the predicted mean location of the inconsistent-cues edge is between that specified by orientation (0) and that specified by frequency ($\Delta$cue). The precise value depends on the weight $w_f$ of the perturbed cue.

In summary, for a choice of the four localization-accuracy parameters, the optimal linear cue combination model predicts every data point in each of the 25 psychometric functions resulting from one observer's trials in one experiment (four single-cue conditions and three two-cue conditions each of which was run at seven levels of $\Delta$cue) by use of Eqs. (2)–(5). This model was fitted to the data by ML. That is, a set of parameters was chosen so as to maximize Pr(data|parameters).

The resulting predictions of the single-cue data from Eq. (2) are shown as the curves in Figs. 2(b) and 6(b). The model produces psychometric functions that contain no response bias (e.g., no tendency to respond more often that the upper edge lies to the right of the lower edge). Thus all of the prediction curves intersect at the point (0, 0.5). For observer MJY in Experiment 1 (Fig. 2) this results in a substantial difference in the predictions from the individual psychometric function fits. Observer MSL in Experiment 2 does not show substantial bias. However, here the model fits appear to be notably shallower than the individual psychometric function fits in Fig. 6(a). We will discover the reason for this shortly.

Figure 8 shows the fit values of $\sigma_{\mathrm{cue}}$ for the two experiments. In every case, increasing the stimulus texture blur resulted in a large decrease in localization accuracy. Note the logarithmic scale on the ordinate of these plots. This is to accommodate the data for observer MJY in Ex-

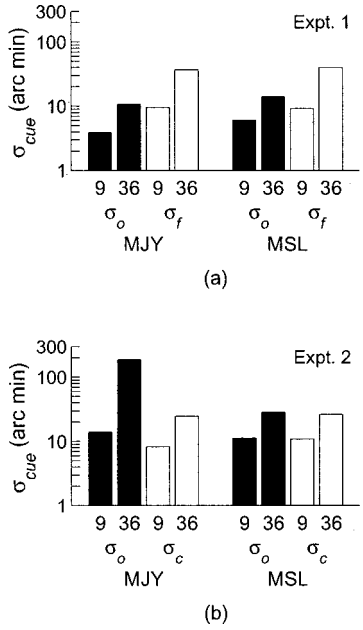


Fig. 8. Values of the four $\sigma_{\mathrm{cue}}$ parameters from fits of the optimal linear cue combination model to the data for each observer in each experiment. The fit values of $\sigma_{\mathrm{cue}}$ increased with increasing texture blur. A logarithmic scale was employed to accommodate the huge value of $\sigma_{\mathrm{cue}}(o, 36)$ for observer MJY in Experiment 2.

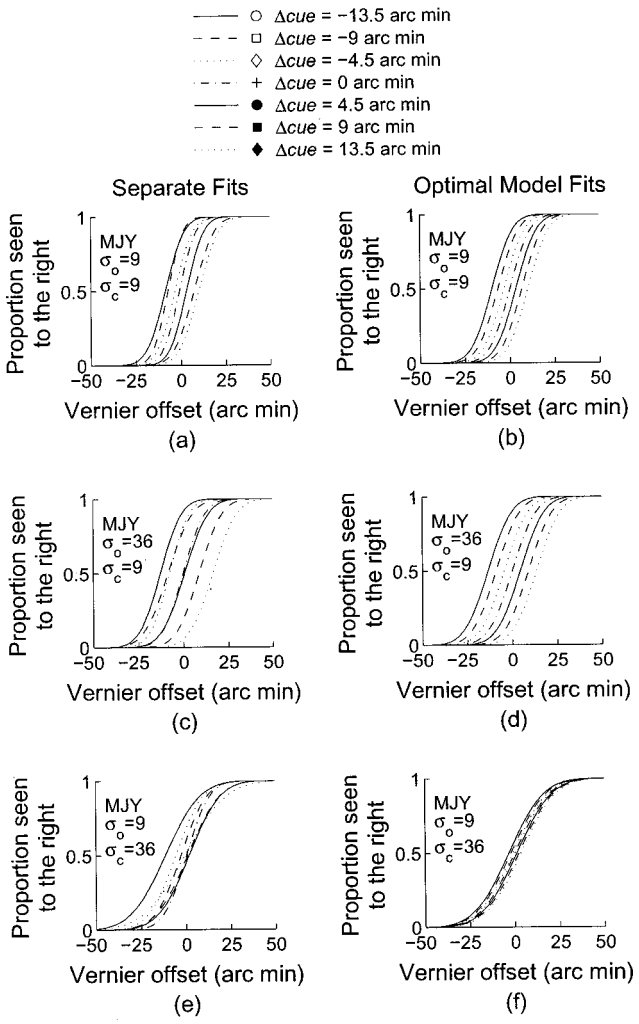


Fig. 9. Fits to the two-cue data sets from Experiment 2 for observer MJY. Fits to each of the 21 two-cue conditions are plotted. Plotting conventions and curves are as in Fig. 3, but with the data points omitted. (a), (c), (e) The curves are fits of cumulative Gaussian functions to individual psychometric functions. (b), (d), (f) The curves are fits of the optimal linear cue combination model to the entire Experiment 2 data set for this observer.

periment 2, who was nearly unable to use the orientation cue in the $\sigma_c = 36$ min condition.

The predictions of the optimal linear cue combination model for the two-cue conditions may be seen as the curves in the right-hand panels of Figs. 3 and 7. It can be seen that this model captures much of what is going on in these very regular data sets. The predictions are even easier to examine in Fig. 9, which shows the individual psychometric function fits as well as the optimal linear cue combination model fits for observer MJY in Experiment 2, omitting the raw data points. In each case, the predictions are a series of parallel psychometric functions, because the noise in Eq. (5) does not depend on Δcue. The shifts of successive psychometric functions are predicted to be the difference in the corresponding Δcue values multiplied by the weight on the perturbed cue ($w_f$ or $w_c$). Large shifts are predicted when this is the more reliable cue, and smaller shifts are predicted when it is the less reliable cue. In comparing these predictions with the individual psychometric function fits, we can see that they predict the main aspects of the data, although they are far more regularly spaced and the data do not always show identical slopes.

There are a couple of interesting things to note in Fig. 9. First, recall that this is the experiment in which subject MJY had difficulty with the orientation cue, especially in the $\sigma_o = 36$ min condition (viz., Fig. 8). Thus in Fig. 9(d) the predicted curves are maximally displaced, so that a value of Δcue = 13.5 min displaces the curve by almost 13.5 min. Second, there is a hint of shallower slopes for the conditions with the largest cue perturbations, especially in Fig. 9(e). This is interesting, as it is what one would expect if a subject were to use a robust estimator, increasing the weight of one cue relative to another when cue conflict grew large (McGraw *et al.*[32] had a similar result). For this condition, given the relative reliabilities of the cues shown in Fig. 8, we would have predicted that the observer would have down-weighted the less reliable contrast cue, and thus the mean would have regressed back toward zero (the location of the orientation-defined edge). This seems to be the case for the right-hand Δcue = 13.5 min curve but not for the left-hand Δcue = −13.5 min curve.

The PSEs predicted by the model are shown as the solid curves in Fig. 4. A weighted average of cues results in a curve of PSEs as a function of Δcue that is a straight line through the origin with a slope equal to the weight of the perturbed cue. These lines are consistent with the major trends in the data but do not fit the data particularly well. The residual errors, however, do not show any clear, explicable trends. One apparent trend is in the $\sigma_o = 9$ min/$\sigma_f = 36$ min data for MSL in Experiment 1 [open diamonds in Fig. 4(b)]. These data points lie mostly above the prediction line, indicating a bias to indicate that the inconsistent-cues stimulus lay to the right of the consistent-cues stimulus, independent of the value of Δcue. This makes little sense, since the inconsistent-

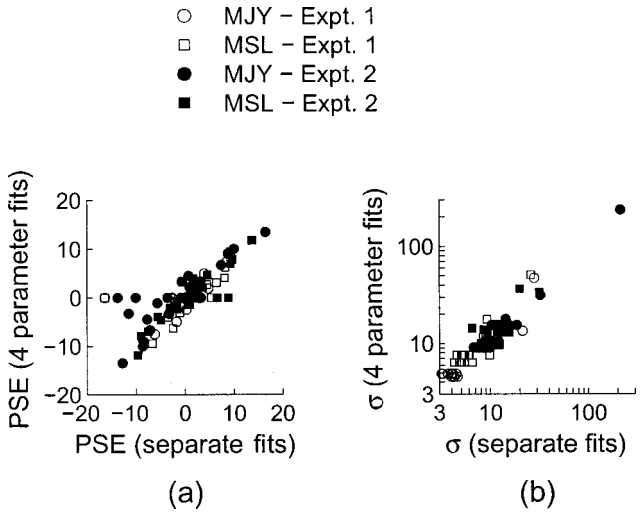


Fig. 10. Scatterplots of the parameters of the individual psychometric function fits to each of the 100 psychometric functions (25 per experiment, 2 experiments, 2 observers) versus those predicted by the four fits of the optimal linear cue combination model to the data from each experiment and observer. (a) PSEs, (b) standard deviations.

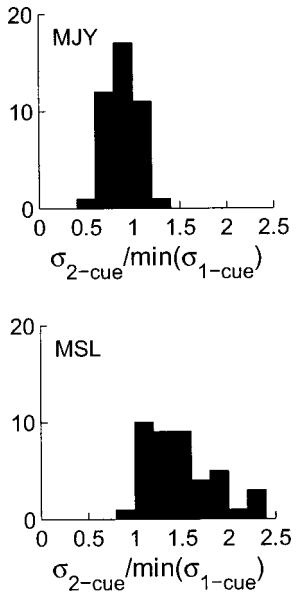


Fig. 11. Histograms of the ratios of the estimated standard deviations from the individual fits for the two-cue conditions to the smaller of the two standard deviations from the corresponding one-cue conditions.

cues stimulus was randomly chosen to be the upper or lower stimulus from trial to trial, and it was very difficult to tell which of the two stimuli was the inconsistent-cues stimulus in the brief presentations we employed. We conclude that the poor fit to the predicted PSEs is simply noisy data.

Figure 10 shows scatterplots of PSEs and $\sigma_{\text{cue}}$'s for the entire set of data collected here (100 psychometric functions: 25 each for each observer and experiment). This is simply another way to represent the degree to which the model was able to capture the bulk of the regularities in the data. The horizontal row of data points in Fig. 10(a) at an ordinate value of zero corresponds to the single-cue conditions, many of which are biased in the separate fits but all of which are predicted to have a PSE of zero by the optimal cue combination model. The correlation in Fig. 10(a) is 0.78 (and rises to 0.94 if we omit the data from the single-cue conditions). The correlation in Fig. 10(b) is 0.99 (0.93 if calculated between the log-transformed values).

The optimal cue combination model implies that the localization accuracy achieved in a two-cue condition should always be as good as or better than the accuracy achieved with either of the single-cue conditions of which it is composed. In Fig. 11 we show a histogram for each of the 42 two-cue conditions run by each subject, of the ratio of the $\sigma$ value for that condition (from the individual psychometric function fits) divided by the minimum of the two $\sigma$ values from the corresponding single-cue conditions. For observer MJY these values are generally near or slightly lower than 1. However, for observer MSL these values are almost all above 1, indicating that MSL was *less* accurate in the two-cue conditions than in the single-cue conditions even though his data [Figs. 4(b) and 4(d)] showed clear effects of cue combination, linear cue weighting, and cue weights dependent on relative cue reliability. This explains why the slopes in the fits of the optimal cue combination model to MSL's single-cue data are shallower than the slopes of the individual fits (Fig. 6). These poor fits were required for the slopes to better match the 21 two-cue psychometric functions. We have no explanation for how a subject can satisfy so many predictions of linear, weighted cue combination and still suffer a loss of localization accuracy with additional localization cues.

### B. Other Models

Up to this point we have discussed two models, each fitted to each data set. The first consisted simply of separate fits of cumulative Gaussian distributions to each of the 25 individual psychometric functions, with two parameters each ($\mu$ and $\sigma$). This is a 50-parameter model. It is merely descriptive; it has no particular theoretical motivation. On the other hand, we have discussed extensively the optimal linear cue combination model with its 4 $\sigma_{\text{cue}}$ parameters. We now examine the results of seven different models fitted to each of the four data sets:

1. Separate fits. This is the fit of separate cumulative Gaussian functions, each with a $\mu$ and $\sigma$ parameter, to the 25 psychometric functions in each data set.
2. No bias. This is the same as Model 1, except that the $\mu$ parameters for the single-cue conditions are fixed at zero. That is, this model assumes no observer bias to respond that the upper edge is, e.g., to the right of the lower edge.
3. Biased, parallel. This is the same as Model 1, except that for each two-cue condition, the seven psychometric functions corresponding to different values of $\Delta$cue are assumed to be parallel and hence share a single $\sigma$ parameter.
4. No bias, parallel. This is the same as Model 3, except that the single-cue $\mu$'s are again fixed at zero.
5. Detecting. This is an extension of the optimal linear cue combination that includes parameters that govern how likely the observer will detect each of the cued edges. Thus, in addition to the four $\sigma_{\text{cue}}$ parameters, there are corresponding $p_{\text{cue}}$ parameters (e.g., in Experiment 1 there are $p_{\text{cue}}(o, 9)$, $p_{\text{cue}}(o, 36)$, $p_{\text{cue}}(f, 9)$ and $p_{\text{cue}}(f, 36)$). In a single-cue trial, if both the upper and lower edges are detected, their noisy location signals are compared, leading to a response. If either or both are missed, then the observer is forced to guess. With a two-cue edge there are two cued edges in the upper stimulus and two in the lower one. Any may be either detected or missed (controlled independently by the corresponding values of $p_{\text{cue}}$), leading to $2^4 = 16$ possible scenarios. If all four edges are detected, optimal cue combination proceeds. If either upper or lower edge has only a single cue detected, then that cue's noisy location signal is used to estimate that edge's location. If both cues are missed in at least one of the stimuli, then the observer is forced to guess.
6. Switching. This is a cue-switching model. It has four $\sigma_{\text{cue}}$ parameters that are used to predict the single-cue data in the same way as in the optimal linear cue combination model. In addition, it has three parameters, one for each two-cue condition, that give the probability $p_o(\sigma_o, \sigma_f)$ that the orientation cue will be used. Thus, on

each two-cue trial, with probability $p_o$ the orientation cue is the only cue used on that trial, and with probability $1 - p_o$ the frequency cue is the only cue that is used (and analogously for Experiment 2, substituting the contrast for the frequency cue).

7. Optimal. This is the optimal linear cue combination model that we have discussed extensively.

Each of these models was fitted, by using a ML criterion, to each of the four data sets. Table 1 gives the log-likelihood at the maximum relative to that of Model 1, which is the model with the greatest number of free parameters. When two models are fitted to the same data set with use of ML, one can test whether the improvement in fit in the more complex model justifies the number of extra parameters it has by performing a "nested hypothesis test."[40] In brief, the quantity $2[\log(L_{\text{complex}}) - \log(L_{\text{simple}})]$ is distributed as $\chi^2$ with degrees of freedom equal to the number of additional parameters in the more complex model. In the table, $p$ values are supplied for a number of such tests, uncorrected for the 24 tests that have been performed (across Tables 1 and 2). Most conclusions are unchanged if a Bonferroni correction is applied (multiplying each $p$ value by 24).

The choice of models tested was motivated by several considerations. First, we were interested in a parsimonious account of the data. The optimal model is an extremely efficient description of the data, as it uses only 4 parameters to account for 25 psychometric functions, comprising over 400 choice probabilities and 10,000 individual psychophysical trials. Second, to the extent that such a model did not fit the data, we were interested in which individual predictions of the model succeeded and which failed. Models of intermediate complexity (Models 2, 3, and 4) were formulated to better understand where the optimal model succeeded and where it failed. Finally, a common alternative model of cue combination is a veto model, where cues are not averaged but rather a single cue is chosen (e.g., the cue perceived to be the most reliable). The detecting and switching models both involve trials in which cues are not combined. The switching model is literally a cue veto model: on each trial one cue vetoes the other. The probability that a particular cue will be chosen is a function of the relative reliability of the cues (i.e., it can be different for each of the three two-cue conditions). On the other hand, the detecting model is more of a generalization of the optimal model. It allows

**Table 1. Relative Log-Likelihood Values and Nested Hypothesis Tests for Various Models Fitted to the Entire Data Set for the Two Experiments and Two Observers[a]**

| Model No. | Model/Test | No. of Parameters | Experiment 1 | | Experiment 2 | |
|---|---|---|---|---|---|---|
| | | | MJY | MSL | MJY | MSL |
| 1 | Separate fits | 50 | 0 | 0 | 0 | 0 |
| | 1 versus 2 | | <0.0001 | <0.0001 | <0.0001 | <0.0001 |
| 2 | No bias | 46 | −53.9 | −81.0 | −21.9 | −28.5 |
| 3 | Biased, parallel | 32 | −9.9 | −26.9 | −16.8 | −17.8 |
| 4 | No bias, parallel | 28 | −63.8 | −107.9 | −38.7 | −46.3 |
| | 4 versus 7 | | <0.0001 | <0.0001 | <0.0001 | <0.0001 |
| 5 | Detecting | 8 | −193.8 | −200.3 | −122.6 | −97.3 |
| | 5 versus 7 | | <0.0001 | <0.0001 | <0.0001 | 0.18 |
| 6 | Switching | 7 | −282.5 | −196.6 | −135.5 | −98.2 |
| 7 | Optimal | 4 | −208.6 | −255.2 | −134.8 | −100.4 |

[a] The log-likelihoods listed are relative to Model 1, which has the greatest number of free parameters. When two models are compared, a nested hypothesis test is performed and the resulting $p$ value (with no correction for multiple tests) is given. See text for details.

**Table 2. Relative Log-Likelihood Values and Nested Hypothesis Tests for Various Models Fitted to the Two-Cue Data for the Two Experiments and Two Observers[a]**

| Model No. | Model/Test | No. of Parameters | Experiment 1 | | Experiment 2 | |
|---|---|---|---|---|---|---|
| | | | MJY | MSL | MJY | MSL |
| 1 = 2 | Separate fits | 42 | 0 | 0 | 0 | 0 |
| | 1 versus 3 | | 0.36 | <0.0001 | 0.014 | 0.008 |
| 3 = 4 | Parallel | 24 | −9.8 | −27.0 | −16.8 | −17.7 |
| | 3 versus 7 | | <0.0001 | <0.0001 | <0.0001 | 0.0006 |
| 5 | Detecting | 8 | −119.6 | −94.6 | −90.0 | −34.5 |
| | 5 versus 7 | | <0.0001 | <0.0001 | <0.0001 | 0.01 |
| 6 | Switching | 7 | −128.3 | −112.9 | −113.2 | −42.4 |
| 7 | Optimal | 4 | −132.2 | −166.2 | −102.5 | −41.1 |

[a] See Note for Table I. Note that with the one-cue data omitted, Models 1 and 2 are identical, as are Models 3 and 4.

for trials in which one (or both) of the two cues goes undetected, forcing the observer to rely completely on the other cue. If no cues result in detection for one of the edges, the observer is forced to guess.

The most obvious summary of Table 1 is this: In almost every case, the more complex model fits the data significantly better than the more parsimonious one. Although this is dismaying, it is not particularly surprising. The full 25-psychometric-function data set consists of 10,000 psychophysical trials. With so much data, an improved fit from a more complex model can be statistically significant without being particularly meaningful. It is for this reason that we spent so much time discussing the degree to which our most parsimonious model (the optimal linear cue combination rule) fitted the important aspects of the data.

In three of four cases in Table 1, the detecting model fits significantly better than the optimal cue combination model. The switching model, on the other hand, provides a worse fit to the data than the optimal model in two of four cases and a minimally improved fit in a third case, despite the additional three parameters. This can happen because this is, in fact, not a nested pair of models (the switching model does not contain the optimal model as a special case). Thus no nested hypothesis test is performed and no $p$ value is computed (although the poor fit would reject the switching model in favor of the optimal model by Akaike's AIC criterion for model identification).[41]

The optimal-model prediction is based on a fixed set of optimal weights used in each two-cue condition, derived from the reliabilities in the single-cue conditions. Even if a different, suboptimal set of fixed weights is used in each two-cue condition, a linear cue combination predicts parallel psychometric functions that are equally spaced [as in, e.g., Fig. 9(b)]. We accept half of this prediction by fitting parallel psychometric functions to the data (Models 3 and 4). The fits of parallel psychometric functions to the two-cue data are significantly better than in the optimal model in all cases. The parallel model fits better than the detecting model as well, but these models are not nested so no test was performed. The separate fits (Models 1 and 2) are significantly better than the parallel ones in three of four cases. (Note that this comparison of Model 1 to 3 and Model 2 to 4, in Table 1, is identical to the comparison of Model 1 to Model 3 where the fit is restricted to the two-cue conditions, as shown in Table 2, and where this test is actually listed.)

The single-cue psychometric functions (e.g., in Fig. 2), which give the proportion of trials in which observers saw the upper stimulus as appearing to the right of the lower one, look distinctly biased. In fact, the separate fits (Model 1) are also significantly better than those that do not allow for response bias (Model 2). A response bias of this kind is inconsistent with the optimal, detecting, and switching models. On the other hand, the two-cue data compare the position of the inconsistent-cues stimulus with that of the consistent-cues stimulus, averaging over occurrences of the inconsistent-cues stimulus in the upper or lower position. Thus a response bias to say that the upper stimulus is to the right of the lower stimulus will lead to a response bias in the single-cue data but will lead to shallower slopes (averaging over two opposing biases) only in the two-cue conditions. Because of this result, we repeated the model fits while omitting the single-cue data (Table 2). (Note that once the single-cue data are omitted from the fit, Models 1 and 2 become identical, as do Models 3 and 4.) It turned out that although the log-likelihood of the fits was then substantially higher, it had no effect on the conclusions of the nested hypothesis tests.

## 5. GENERAL DISCUSSION

The results of the experiments are somewhat equivocal regarding the optimal linear combination of cues to location. In support of the model are the findings that perceived edge location is midway between that signaled by each cue in the inconsistent-cues stimulus, that this is reasonably modeled as a linear function of $\Delta$cue, and that the slope of this function (the estimated weight of the perturbed cue) is larger when that cue is relatively more reliable. The model largely captured the variation in the slopes of the various psychometric functions and in their means. The model is also consistent, in spirit, with the independent contributions of texture cues to contour detection and orientation discrimination.[26,29,31] In contrast, the model is not consistent with the larger standard deviations in the two-cue conditions for subject MSL compared with that in the single-cue conditions, and the model was handily rejected in favor of, for example, separate fits for each of the 25 conditions.

It is worth considering how an independent-cues model would be implemented in the nervous system. It is certainly true that the cues investigated here (spatial frequency, orientation, and contrast) are multiplexed at early stages in the visual system. That is, any given neuron in, say, cortical area V1 will modulate its response with changes in any of these three stimulus variables. With such a distributed representation of our stimuli, the attribution of edge location to each cue separately may seem artificial. It is true that we can detect edges signaled by one or the other of these cues alone. For example, observers are able to compute an estimate of local orientation and to detect variations in orientation over space, and models have been suggested as to how this is done.[42] It may well be the case that models can fit our data equally well that do not begin by splitting the stimulus into separate estimates of individual cue properties but rather work with the raw, distributed representation and look for significant spatial changes in the pattern of neural responses, whatever form they may take. Formulating models of this sort and relating them more closely to the underlying neural representation of spatial patterns constitute an important and challenging direction that remains for research in this area to take.

## ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grant EY08266, Air Force Office of Scientific Research grant 93NL366, and a National Research Council–NASA Ames Research Center Associateship. We thank Larry Maloney, Paul Warren, and Erica Alliston for comments

on the manuscript, and Jeff Mulligan, Misha Pavel, Mark Young, and Carlo Tiana for guidance in the early stages of this research.

Michael S. Landy can be reached at the address on the title page or by e-mail, landy@nyu.edu. http://www.cns.nyu.edu/~msl.

## REFERENCES


1. L. Kaufman, *Sight and Mind* (Oxford, New York, 1974).
2. M. S. Landy and E. Brenner, "Motion-disparity interaction and the scaling of stereoscopic disparity," in *Vision and Attention*, L. R. Harris and M. R. M. Jenkin, eds. (Springer-Verlag, New York, 2001), Chap. 7.
3. M. S. Landy, L. T. Maloney, E. B. Johnston, and M. J. Young, "Measurement and modeling of depth cue combination: in defense of weak fusion," Vision Res. **35**, 389–412 (1995).
4. J. J. Clark and A. L. Yuille, *Data Fusion for Sensory Information Processing Systems* (Kluwer Academic, Boston, 1990).
5. D. Marr, *Vision* (Freeman, San Francisco, Calif., 1982).
6. J. Aloimonos and D. A. Shulman, *Integration of Visual Modules: an Extension of the Marr Paradigm* (Academic, New York, 1989).
7. L. T. Maloney and M. S. Landy, "A statistical framework for robust fusion of depth information," in *Visual Communications and Image Processing IV*, W. A. Pearlman, ed., Proc. SPIE **1199**, 1154–1163 (1989).
8. M. J. Young, M. S. Landy, and L. T. Maloney, "A perturbation analysis of depth perception from combinations of texture and motion cues," Vision Res. **33**, 2685–2696 (1993).
9. E. B. Johnston, B. G. Cumming, and M. S. Landy, "Integration of stereopsis and motion shape cues," Vision Res. **34**, 2259–2275 (1994).
10. J. R. Li, L. T. Maloney, and M. S. Landy, "Combination of consistent and inconsistent depth cues," Invest. Ophthalmol. Visual Sci. Suppl. **38**, S903 (1997).
11. E. Brenner and M. S. Landy, "Interaction between the perceived shape of two objects," Vision Res. **39**, 933–945 (1999).
12. I. Fine and R. A. Jacobs, "Modeling the combination of motion, stereo, and vergence angle cues to visual depth," Neural Comput. **11**, 1297–1330 (1999).
13. D. C. Knill and W. Richards, *Perception as Bayesian Inference* (Cambridge U. Press, Cambridge, UK, 1996).
14. A. L. Yuille and H. H. Bülthoff, "Bayesian decision theory and psychophysics," in *Perception as Bayesian Inference*, D. C. Knill and W. Richards, eds. (Cambridge U. Press, Cambridge, UK, 1996), pp. 123–161.
15. D. W. Massaro and D. Friedman, "Models of integration given multiple sources of information," Psychol. Rev. **97**, 225–252 (1990).
16. J. E. Cutting, N. Bruno, N. P. Brady, and C. Moore, "Selectivity, scope, and simplicity of models: a lesson from fitting judgments of perceived depth," J. Exp. Psychol. Gen. **121**, 364–381 (1992).
17. D. W. Massaro, "Ambiguity in perception and experimentation," J. Exp. Psychol. Gen. **117**, 417–421 (1988).
18. D. W. Massaro and M. M. Cohen, "The paradigm and the fuzzy logical model of perception are alive and well," J. Exp. Psychol. Gen. **122**, 115–124 (1993).
19. L. T. Maloney, "Physics-based approaches to modeling surface color perception," in *Color Vision: From Genes to Perception*, K. R. Gegenfurtner and L. T. Sharpe, eds. (Cambridge U. Press, Cambridge, UK, 1999), pp. 387–422.
20. L. T. Maloney and J. N. Yang, "The illuminant estimation hypothesis and surface color perception," in *Colour Vision: From Light to Object*, R. Mausfeld and D. Heyer, eds. (Oxford U. Press, Oxford, UK, to be published).
21. A. M. Derrington and D. R. Badcock, "Separate detectors for simple and complex grating patterns?" Vision Res. **25**, 1869–1878 (1985).
22. T. Ledgeway and A. T. Smith, "Evidence for separate motion-detecting mechanisms for first- and second-order motion in human vision," Vision Res. **34**, 2727–2740 (1994).
23. N. E. Scott-Samuel and M. A. Georgeson, "Does early nonlinearity account for second-order motion?" Vision Res. **39**, 2853–2865 (1999).
24. H. R. Wilson and J. Kim, "A model of motion coherence and transparency," Visual Neurosci. **11**, 1205–1220 (1994).
25. P. Cavanagh, S. Saida, and J. Rivest, "The contribution of color to depth perceived from motion parallax," Vision Res. **35**, 1871–1878 (1995).
26. F. S. Frome, S. L. Buck, and R. M. Boynton, "Visibility of borders: separate and combined effects of color differences, luminance contrast, and luminance level," J. Opt. Soc. Am. **71**, 145–150 (1981).
27. J. Rivest, I. Boutet, and J. Intriligator, "Perceptual learning of orientation discrimination by more than one attribute," Vision Res. **37**, 273–281 (1997).
28. M. S. Landy, "Combining multiple cues for texture edge localization," in *Human Vision, Visual Processing, and Digital Display IV*, J. P. Allebach and B. E. Rogowitz, eds., Proc. SPIE **1913**, 506–517 (1993).
29. J. Rivest and P. Cavanagh, "Localizing contours defined by more than one attribute," Vision Res. **36**, 53–66 (1996).
30. D. R. Badcock and G. Westheimer, "Spatial location and hyperacuity: the centre/surround localization contribution function has two substrates," Vision Res. **25**, 1259–1267 (1985).
31. R. Gray and D. Regan, "Vernier step acuity and bisection acuity for texture-defined form," Vision Res. **37**, 1713–1723 (1997).
32. P. V. McGraw, D. Whitaker, and D. R. Badcock, "Localising conflicting visual attributes," Invest. Ophthalmol. Visual Sci. Suppl. **41**, S804 (2000).
33. M. S. Landy and J. R. Bergen, "Texture segregation and orientation gradient," Vision Res. **31**, 679–691 (1991).
34. A. K. Hon, L. T. Maloney, and M. S. Landy, "The influence function for visual interpolation," in *Human Vision and Electronic Imaging II*, B. E. Rogowitz and T. N. Pappas, eds., Proc. SPIE **3016**, 409–419 (1997).
35. J. R. Bergen and M. S. Landy, "Computational modeling of visual texture segregation," in *Computational Models of Visual Processing*, M. S. Landy and J. A. Movshon, eds. (MIT Press, Cambridge, Mass., 1991), pp. 253–271.
36. M. S. Landy, Y. Cohen, and G. Sperling, "Hips: a unix-based image processing system," Comput. Vision Graph. Image Process. **25**, 331–347 (1984).
37. N. Prins and A. J. Mussap, "Alignment of orientation-modulated textures," Vision Res. **40**, 3567–3573 (2000).
38. S. S. Wolfson and M. S. Landy, "Discrimination of orientation-defined texture edges," Vision Res. **35**, 2863–2877 (1995).
39. F. A. A. Kingdom, D. Keeble, and B. Moulden, "Sensitivity to orientation modulation in micropattern-based texture," Vision Res. **35**, 79–91 (1995).
40. A. M. Mood, F. A. Graybill, and D. C. Boes, *Introduction to the Theory of Statistics*, 3rd ed. (McGraw-Hill, New York, 1974).
41. H. Akaike, "A new look at the statistical model identification," IEEE Trans. Aut. Control **AC-19**, 716–723 (1974).
42. S. C. Dakin, "Orientation variance as a quantifier of structure in texture," Spatial Vision **12**, 1–30 (1999).
43. M. K. Kendall and A. Stuart, *The Advanced Theory of Statistics: Vol. 2. Inference and Relationship*, 4th ed. (Macmillan, New York, 1979).