Depth and luminance edges attract

Alan E. Robinson

Department of Psychology and Department of Cognitive Science, University of California, San Diego, La Jolla, CA, USA

Donald I. A. MacLeod

Department of Psychology, University of California, San Diego, La Jolla, CA, USA

The spatial resolution of disparity perception is poor compared to luminance perception, yet we do not notice that depth edges are more blurry than luminance edges. Is this because the two cues are combined by the visual system? Subjects judged the locations of depth-defined or luminance-defined edges, which were separated by up to 5.6 min of arc. The perceived edge location was a function of the depth-defined edge and the luminance-defined edge, with the luminance edge tending to play a larger role. Our data are compatible with but not completely explained by an optimal cue-combination model that gives more reliable cues a heavier weight. Both edge cues (depth and luminance) contribute to the final percept, with an adaptive weighting depending on the task and the acuity with which each cue is perceived.

Introduction

Depth perception relies on many cues. Of these, disparity is perhaps the most compelling. Just by itself it can impart otherwise random dot patterns with clear and distinct depth, as real as any natural scene. And without it, a natural scene seems much flatter, even though depth can be inferred from a multitude of other cues.

While disparity conveys a powerful effect, our sensitivity to it is limited. In particular, our spatial resolution for disparity is poor. The highest perceivable spatial frequency of a disparity defined grating is 4 cycles/° (Tyler, 1974), which is much lower than for luminance. This low-pass characteristic predicts that disparity transitions should appear blurry. A near object against a far background should appear to warp toward the background around the edges, due to the loss of high spatial frequencies, just as blurring a luminance pattern creates intermediate values between light and dark transitions. Alternatively, the precise location of the depth transition might appear vague and hard to localize. Casual inspection suggests that neither of these occur in real scenes. Recently Kane, Guan, and Banks (in press) studied the appearance of depth transitions in random dot stereograms and found that step transitions always appeared sharp, as did smooth transitions when the width of the transition was narrower than the spatial sensitivity of disparity mechanisms. Thus, it seems the visual system defaults to seeing sharp depth transitions when the input is ambiguous. This default assumption, however, does nothing to help the visual system determine the exact location of that step transition.

This problem could be resolved by combining disparity with other features perceivable at a higher spatial resolution, thus constraining the perceived location of depth transitions. Here we investigate if luminance, which has a much higher spatial resolution, acts in this way. Formally, we propose that when two edges occur near each other, one defined by depth, and one defined by luminance, the perceived depth-edge location is shifted toward (attracted to) the location of the luminance edge by some significant amount. One model of how this could occur is cue combination, where two or more noisy sources of information (cues) are combined to produce a more accurate percept.


doi: 10.1167/13.11.3

Received December 22, 2012; published September 6, 2013

ISSN 1534-7362 © 2013 ARVO

Downloaded From: https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932807/ on 12/17/2018
Rivest and Cavanagh (1996) measured the attraction between pairs of edges defined by luminance, color, motion, and texture, and in all cases they found that the target edge was shifted toward the flanking edge by 1–2 arcmin. Measureable attraction was found even when edges were separated by up to 10 arcmin for almost all pairs of edge types, and for a few rare cases up to 30–40 arcmin. They also measured vernier acuity for a single edge defined by one, two, or three features and found modest improvements in localization with more features.

There is also single unit neurophysiology showing that cells in V2 respond to edges defined by luminance or stereoscopic depth (von der Heydt, Zhou, & Friedman, 2000). Interestingly, these cells respond less strongly when the cues suggest conflicting (opposite) depth polarities (Qiu & von der Heydt, 2005) raising the possibility that they are involved in integrating these cues.

The closest ancestor to our work is Morgan (1986). Morgan constructed a vernier alignment task where the target was a white bar offset in depth (disparity) against a random dot stereogram background. The edge of the luminance bar could be camouflaged by randomly flipping elements of the bar from white to black, thus making it match the background. This may have made the disparity of the target slightly more difficult to see, but had a far larger effect on the detectability of the luminance edge, and in the limit the target was invisible when viewed monocularly. Vernier acuity was always better when both disparity and luminance cues were available, which is suggestive of cue combination.

Morgan argues, however, that the data can be explained just as well by assuming “probability summation” (hereafter, cue switching) rather than cue combination. With an intermediate amount of camouflage one or both cues might not be detected, and with two cues there is a greater chance that at least one cue will be detected accurately on any given trial. Thus, acuity could be higher even with absolutely no integration of cues if subjects simply switch to using whichever cue is most visible on any given trial.

In Morgan’s experiment the cues (visible or not) always indicated the same target location, so changes in acuity are the only potential measure of cue combination. Here we ask subjects to judge the location of an edge defined by disparity and/or by luminance (luminance). When both edges are present they can either coincide (as in Morgan) or are offset from each other. This allows us to differentiate between cue combination and cue switching. Cue combination predicts that the apparent edge will be at an intermediate location between cues, even though subjects are explicitly told to attend only to the task-relevant cue. Furthermore, when the cues overlap localization should be improved, relative to the cues in isolation. In cue switching, subjects will base their judgment on one of the two cues; typically this will be the task-relevant cue, but when that is particularly difficult to see they might use the task-irrelevant cue instead. This could lead to more accurate localization when the edge cues indicate the same location, as in Morgan’s experiment, but could lead to drastically worse localization when they are in different locations.

Experiment

Our stimuli contained a depth-defined edge and a luminance-defined edge, which were separated vertically by 0.2, 2.8, or 5.6 arcmin. Depending on the condition, subjects were asked to judge the location of the depth edge or the luminance edge, ignoring the other edge. Would their judgments be a weighted function of both the task-relevant edge and task-irrelevant edge? Furthermore, we manipulated the depth-edge visibility (by changing the magnitude the change in binocular disparity at the edge) to see if a more visible depth edge would receive relatively more weight, as predicted by optimal models of cue combination (e.g., Jacobs, 1999).

Methods

We ran three subjects: One was the first author and the others were naive to the purpose of the study. All were experienced psychophysical observers with normal stereopsis and corrected to 20/20 vision in both eyes.

Apparatus

Stimuli were presented on a 22 in. LaCie electron22blueIV Diamondtron CRT driven by an NVIDIA GeForce GT 545 GT video card at a refresh rate of 120 Hz in an otherwise unlit room. A photometer (UDT, model 247, UDT Instruments, San Diego, CA) was used to select the appropriate lookup table values for gamma compensation. A chin rest was used to maintain a viewing distance of 6.4 ft. A mirror stereoscope presented a separate image to each eye. Each image subtended 6° by 9.6° (W × H) with a resolution of 640 × 1024 pixels (107 pixels/°). Maximum luminance was 71 cd/m².

Stimuli were generated and displayed using Matlab running the Psychophysics Toolbox, version 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) on a Windows XP computer.

Stimuli

Our stimuli consisted of a random texture divided into two different (stereoscopic) depths and two
different intensities (Figure 1). The nearer and brighter parts of the texture were on the lower portion of the CRT.

The depth edge was added to the texture by shifting the lower portion of the texture in opposite directions in each eye, generating a total disparity of 16.8 arcmin or 5.6 arcmin, depending on condition. We refer to these as large and small depth step conditions, respectively. This difference was intended to modulate the detectability of the depth edge.

The luminance edge was added to the texture as a uniform increment (30\% of maximum luminance) from the bottom of the screen up to the depth edge (the coincident condition), or beyond (by 2.8 or 5.6 arcmin, referred to as the noncoincident conditions). In pilot work, we found that the luminance increment (with the associated slight reduction in contrast) decreased texture visibility slightly, so in order to minimize differences in visibility of the depth discontinuity the luminance step was always above the depth edge so that pattern contrast at the location of the depth edge was unaffected by the step. The luminance step was inset from the vertical edges of the texture by 5.6 arcmin on each side so that it would be somewhat distant from the reference line, thus making it slightly harder to localize.

We also included a depth-edge only condition (the luminance increment was added to the whole texture), and a luminance-edge only condition (the texture was all at the same depth).

A single pixel (0.5 arcmin) horizontal line was shown halfway down the screen on the right side as a reference edge; all judgments of edge position were made relative to this line.

The random texture was designed to have highly visible vertical edges of varied lengths and widths, much like a natural image. It was generated by drawing 2,000 randomly placed rectangles, with intensities ranging from 0\% to 70\% of maximum luminance, starting with a black background. Rectangles were up to 0.30\ in width and height (selected independently). Note that this texture could provide a monocular cue to the location of the depth edge because that edge would break the alignment of all rectangles that cross it. To hide this, we introduced random horizontal texture breaks that spanned the entire width of the texture, thus mimicking the monocular depth cue. We randomly selected 0.28\ tall horizontal slices of the texture, and randomly shifted them by up to 0.19\ to the left or right. This did not introduce any disparity, however, since the offset was equal in both eyes. Only the central portion of the final texture was used, so that there were no artifacts along the texture perimeter. Four different textures were generated and one was selected at random on each trial, to prevent local adaptation and motion cues between trials.

**Procedure**

Subjects judged the location of the depth edge or the luminance edge in different sessions, with breaks between sessions ranging from minutes to days. On each trial they indicated if the task-relevant edge was above or below the reference line on the right side of the screen. There was no time limit and no feedback.

A variable step-size staircase was used to adjust the location of the task-relevant edge on each trial, first making large changes in position, and then progressively smaller changes. After 20 trials we switched to a probability tracking staircase that uniformly sampled from the current estimate of the 10\%, 40\%, 60\%, and 90\% “edge appears above the reference” points on the psychometric function for that condition.

The psychometric function relates the physical offset between task-relevant edge and reference to the fraction of trials on which the edge was judged to be above the reference. We approximated the psychometric function by a cumulative Gaussian (normal ogive). The point of subjective equality (PSE) is the 50\% point on the psychometric curve and corresponds to the position at which the task-relevant edge is equally likely to be judged as above or below the reference edge.

A second parameter of the cumulative Gaussian is the width (or slope, which is inversely proportional to width; a narrower curve with steeper slope indicates greater sensitivity because a smaller positional change produces...
a larger response change). We characterize the width by the difference between the 50% point and the 84% point on the psychometric function, in arcmin. This measure is $\sigma$, the standard deviation of the cumulative Gaussian. It is also the standard deviation of the implied distribution of the perceived locations of the edge, if we neglect variability in the perceived location of the reference (an assumption we relax below).

The best-fitting PSE and $\sigma$ were found by determining the likelihood of the observed data as a function of both these parameter values (densely sampled over a wide range), assuming binomial variability in the counts. Marginal likelihood distributions were then obtained by evaluating the likelihood of the data as a function of each model parameter in turn while integrating over values of the other parameter; assuming a flat prior, these marginal likelihoods are proportional to the posterior probability density functions of the model parameter given the data. The adopted best-fitting estimates for PSE and $\sigma$ were the means of the marginal likelihood distributions, and the standard deviations of the marginal likelihoods defined the associated standard errors.

**Results**

We had two kinds of conditions: coincident, where task-relevant and irrelevant edges were at the same location, and noncoincident, where the distance between edges differed by a fixed offset. To determine the effect of the offset task-irrelevant edge in the noncoincident conditions we subtracted the coincident PSEs from the noncoincident PSEs (thus eliminating response bias). A positive value indicates that the task-relevant edge’s perceived position is shifted toward the task-irrelevant edge. These values are plotted for our three subjects in Figures 2–4. We manipulated the visibility of the depth edge by changing the magnitude of the step (large and small). Surprisingly, one of our subjects (JM) actually found the small depth edge more detectable. For ease of interpretation, we hereafter adopt the terms weak and strong depth edge, with weak referring to the large depth edge for JM and the small depth edge for the other two subjects.

First we will consider the conditions where subjects judged the location of the depth edge. For all these conditions, all subjects showed a shift in the perceived location of the depth edge toward the nearby luminance edge (left-hand side of Figures 2–4). The magnitude of this shift varied by subject and condition, but was at least half the distance to the luminance edge, and often much more. Averaging across subjects and depth-edge strengths, the shift was 64% of the distance to the task-irrelevant edge when the edges were separated by 2.8 arcmin. When the edges were 5.6 arcmin apart the average shift was larger but was only 53% of the distance between them.

Next we consider the effect of the depth edge on the perceived location of the luminance edge (right-hand side of Figures 2–4). Here the perceived edge location was again shifted toward the task-irrelevant edge, though the effect was smaller, and the magnitude varied somewhat inconsistently between subjects and conditions. For the 2.8 arcmin separation the average shift was 19% of the distance to the task-irrelevant edge; for the 5.6 arcmin separation the average shift was just 10%. Though all subjects showed a trend in the same direction for all conditions (except subject JM in one of the weak depth-edge conditions), the effect was significant only for the strong depth-cue conditions (where it would be expected to be largest), and only one case remained significant ($p < 0.05$) after multiplying the significance level by 12 to correct for multiple comparisons. The consistent direction of effect (11 out of 12 conditions), however, does suggest this effect is real, though much smaller than the effect of the luminance edge. If we consider the perceived shifts in the 12 cases as independent measures, a two-tailed $t$ test rejects the null hypothesis of zero shift at $p < 0.001$.

This analysis implicitly neglects subject variation as well as variation with offset and depth strength, and these factors were indeed not significant in analysis of variance on the luminance edge judgments.

Together, these results show that depth and luminance edges can interact, giving rise to a shift in apparent location. Next we ask if there is also a change in the precision with which the edge locations are detected. The same psychometric functions that gave us the PSEs also indicate the overall precision of those judgments: the more steep the slope (i.e., the more narrow the function) the more precisely an edge is perceived (Figure 5).

To see if acuity improved we compared the sigma values for depth-edge judgments made in isolation and in the presence of a coincident luminance edge (where there was no cue conflict). The sigma values are not very tightly constrained by the data. But in the condition where the depth edge was perceived with least acuity (small for AR and MS and large for JM) the improvement in acuity from adding a luminance edge was significant. In the condition where the depth edge was more detectable, (and thus the luminance edge would have relatively less to contribute), the same trend was observed, but was only significant for subject MS.

Figure 5 also shows the sigmas for the conditions where both depth and luminance edges were shown, but with an offset between them. Here too the acuity is on average higher than in the depth in isolation condition: the addition of a noncoincident luminance edge improved the precision of depth-edge localization in 10 of the 12 cases (3 Subjects $\times$ 2 Depth/Luminance
Edge Separations × 2 Depth Differences), reducing sigma to 78% of its depth-only value on the average, a statistically significant difference if the sigma estimates are treated as independent observations ($t = 2.94$, df = 11, $p < 0.01$). No effects of subject, offset, or depth difference were statistically significant.

We also looked for the same effect for luminance-edge acuity. Here, much as with the position data, the effect was much weaker, with some trends in the same direction, but no statistical significance.

**Discussion**

**Cue combination and cue switching**

Our results indicate that the perceived position of a depth edge is influenced by (attracted toward) nearby luminance edges. We also found that the perceived position of the luminance edge can be influenced by the depth edge, although this was clearly a smaller effect. We also found evidence of increased acuity for perceiving depth edges in the presence of supporting luminance edges, and a hint of the symmetrical effect. These data go beyond those of Morgan (1986) by showing the edges influence each other even when not overlapping. The question remains, however, if our data specifically supports cue combination, or if like Morgan’s data, it can also be explained by cue switching.

We consider first the widely adopted optimal cue-combination model. If the location estimates implied by each cue are contaminated by independent random errors, it is advantageous to combine the two signals by taking a weighted average that favors the more reliable cue (e.g., Jacobs, 1999; Yuille & Bulthoff, 1996). The weight is set by the error variances of the perceived edge locations signaled by each cue (when shown in isolation). If the weights are inversely proportional to
the variances, the lower variance cue gets higher weight. This weighting is optimal in the sense that it minimizes the variance of the resultant estimate when the individual estimates are contaminated by Gaussian noise.

We contrast this with a cue-switching model, where the subject chooses to attend to one cue or the other whenever both are present. Ideally they would use the task-relevant cue exclusively, but if the availability of the cues fluctuates from trial to trial, they might switch to using the irrelevant cue on some percentage of the trials. Whatever the reason, if they do alternate between cues, the final psychometric function will be an additive mixture of the two functions for the cue in isolation. Since the functions have inflection points at different locations, the mixture function will as well.

Figure 6 illustrates the predictions of these two models for a mixed cue case in which the luminance edge was located 5.6 arcmin above the depth edge. The horizontal axis shows the height of the reference line relative to the depth edge in arcmin; the vertical axis shows the fraction of trials on which the reference was judged on some percentage of the trials. Whatever the reason, if they do alternate between cues, the final psychometric function will be an additive mixture of the two functions for the cue in isolation. Since the functions have inflection points at different locations, the mixture function will as well.

Figure 6 illustrates the predictions of these two models for a mixed cue case in which the luminance edge was located 5.6 arcmin above the depth edge. The horizontal axis shows the height of the reference line relative to the depth edge in arcmin; the vertical axis shows the fraction of trials on which the reference was judged above the test stimulus edge. The red curve and asterisks are the cumulative Gaussian that best fit the data of subject JM. If the subject relied entirely on the depth signal, the psychometric function would have its inflection point at zero on the horizontal axis, as shown by the left-hand dashed curve; the reference at \( x = 0 \) is neither above nor below the test depth edge, but aligned with it. If the subject relied entirely on the luminance signal, the psychometric function would have its inflection point at \( +5.6 \) arcmin on the horizontal axis, as shown by the right-hand dashed curve; at that location the reference is aligned with the test-luminance edge. The data show an inflection at an intermediate location, where the reference is 3.7 min of visual angle above the depth edge, hence 1.9 min below the luminance edge. The cue-switching model can approximate this behavior by assuming that the luminance cue is consulted on 61% of trials, and the depth cue on the remaining 39%. The dashed and dotted curves in Figure 6 are scaled accordingly. The open circles show the total fraction of “above” judgments, obtained by summing the dashed and dotted functions, that is by summing over both kinds of trials in the 61:39 ratio shown. The open circles are the cue-switching model’s maximum likelihood fit to the data of JM by optimizing the proportion of luminance-based judgments.
In these model evaluations, the values of sigma assumed for each separate cue were always the ones obtained experimentally in the single-cue conditions (with a very small correction, discussed below, for variance in the perceived location of the reference line). Thus in Figure 6 the dashed and dotted curves have the sigma values, or widths, measured in those conditions for subject JM. Also as noted in the methods section, we discounted small subject-dependent biases in relative location of reference and test by setting \( x = 0 \) at the PSE measured for the coincident condition, rather than the point of physical alignment.

Although the cue-switching model appropriately predicts the PSE, the open circles show a conspicuous double inflection reflecting the dual origin of the judgments. This is typical of the cue-switching model’s predictions for large misalignments between depth and luminance edges. The red curve fit to the data is constrained to be singly inflected, and cannot in itself reject the model. But we show below that the raw data are likewise inconsistent with the cue-switching model. Notably, the psychometric functions with noncoincident cues are steeper (the sigmas are smaller) than predicted. The data underlying Figure 6 were typical in this respect, and the best-fitting curve (red asterisks) is clearly steeper than the cue-switching prediction (a more formal statistical analysis is given below).

Under the cue-combination model the two-cue psychometric slope is steeper (smaller sigma) than when either cue is shown in isolation because the (independent) noise for each cue is averaged out. The filled circles in Figure 6 show the cue-combination prediction generated by optimizing the relative weighting of the luminance and depth cues. Just as the proportion of luminance-based judgments dictates both the PSE and form of the psychometric function in the cue-switching model, the relative cue-weighting parameter dictates the PSE and sigma for the (cumulative Gaussian) prediction of the cue-combination model. The best single-parameter fit for cue combination is indeed slightly steeper than the best two-parameter fit to the data, but it does come closer than the competing cue-switching model. This is a typical outcome as we show below.

To apply the optimal cue-combination model we calculated optimal weights from the single-cue sigma values. These optimal weights define the PSE and sigma for the two-cue case. It is tempting, and common, to identify the single-cue sigma values with the values for the single-cue psychometric functions. This is a
Figure 5. The precision of judgments for each condition and subject. Sigma is the psychometric curve width (the distance between 50% and 84% on the psychometric function), measured in arcmin. Thus, smaller values denote greater precision. Error bars denote ±1 SEM.
mistake, because in a comparison of single-cue test and reference locations, the variance in relative location that underlies the psychometric function is due in part to variance in the perceived location of the reference. To address this, we ran a control experiment in which the test edge was replaced by a second high contrast line similar to our reference. The resulting psychometric function was fit with sigma roughly fivefold smaller than the single-cue test sigmas, implying that only about 2% of the single-cue test variance was attributable to the reference line. To obtain single-cue sigma values, we subtracted the implied reference variance from the square of each experimental single-cue psychometric function sigma. This small correction made no substantial difference to the estimated single-cue sigmas or to the derived optimal (inverse variance) weights. Likewise, the two-cue sigmas had to be augmented by the addition of reference variance when we generated the predicted psychometric functions. This correction too was practically negligible in effect.

Next we consider how well cue combination explains our data on perceived location, and in particular the idea that the perceived location is the mean of the values suggested by the two cues, weighted inversely by their variance. Recall that we found large effects of luminance on the PSE for judgments of a depth edge, and weak effects in the reverse direction. To explain those results on the basis of optimal cue combination, the greater weighting of the luminance cue should be traceable to a correspondingly smaller variance. The variance of perceived edge location based on each cue in isolation is simply the square of the sigma value for the condition where that edge type was shown in isolation (this is the variance over trials of the point of subjective equality for that edge type, with the above noted minor correction for variance in the perceived location of the reference line). The location variance for the depth-only edge was indeed greater than for the luminance-only edge, qualitatively justifying the greater weight for the luminance cue that was reflected in behavior of the experimental PSEs (Figure 5).

Yet in Figures 2–4 the experimental PSEs are not quantitatively well predicted by the optimally weighted cue-combination model. To examine this point more carefully we also fit each two-cue psychometric function using an unconstrained combination model, in which the relative weighting is a single free parameter, varied for a best fit to the two-cue data using the experimentally obtained single-cue sigmas. The predicted two-cue psychometric function depends in a very simple way on the weight parameter. The psychometric functions are always cumulative Gaussian; the PSE given by the weighted average of the two cued locations divides the interval between the

![Diagram](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/932807/)
single-cue PSEs in the ratio of the weights. Sigma is also determined by the weight parameter.

Figure 7 compares the best-fitting weights for the unconstrained cue-combination model with the optimal weights. The horizontal axis shows the optimal weight for the depth cue, and the vertical one shows the best-fitting depth cue weights with 95% confidence intervals, for all 12 conditions in which the luminance and depth edges were noncoincident. Each circle represents one condition for one subject. Open and filled circles are for the depth-edge and luminance-edge location tasks, respectively. The observed weights have a positive central tendency, which is reliable even for the filled circles: for 8 of the 12 cases the 95% confidence intervals exclude zero, and a t test yields \( p < 0.001 \) if the individual weights are treated as independent observations; in support of this admittedly questionable idealizing assumption of independence, analysis of variance showed no significant main effects or interactions for the subject or experimental condition factors.

The optimal weighting model predicts that the depth edges associated with greater single-cue precision should be influenced less by luminance edges. This prediction finds limited support. The predicted correlation between observed and optimal weight is indeed positive but not statistically significant. Effects of edge misalignment and depth difference magnitude were also not significant. But curiously, in subject JM, the larger depth difference unexpectedly gave less precise localization than the small one, and for that subject the weight of the depth cue was reduced correspondingly (and significantly, with \( p < 0.05 \) for the interaction between subject and depth difference in analysis of variance on the best-fitting weights).

The filled circles clearly fall below the diagonal and below the open circles, implying a less than optimal weight for the depth cue in the luminance-edge location task (\( p < 0.001 \)). The open circles show an opposite but statistically insignificant tendency, suggesting a greater than optimal weight for the depth cue when judging depth-edge location.

Model comparison: Mountains in a probabilistic landscape

The cue-combination model can generate any desired PSE by suitable variation of the cue weight. The cue-switching model can similarly generate any PSE intermediate between the luminance and depth PSEs by varying the probability of using the depth cue. But the models can be distinguished using the shape of the predicted psychometric functions. In particular, the model parameter dictates not only the PSE but the psychometric function sigma, which we consider next.

As noted in the results section, the addition of a luminance edge sharpened the precision of depth localization, reducing sigma, even when the edges were noncoincident. The cue-combination model, however, predicts that the sigma value when both cues are available will be less than the lower of the two single-cue sigma values (see for instance, Landy & Kojima, 2001), in this case the luminance-only sigma values. Our sigmas in the depth task with conflicting cues did not differ reliably in either direction from the luminance-only sigma values, even if we take the liberty of treating the sigmas for each subject and condition as independent observations. They are, however, significantly greater (\( p < 0.01 \)) than predicted by the optimal cue-combination model. And even if we adopt the best-fitting weights (rather than the optimal ones) for cue combination, the sigmas predicted for the noncoincident conditions remain too small, at about 0.8 times the values observed. Thus the cue-combination model may not be strictly consistent with the data in this respect. Others have found that their data only partially matches this prediction of optimal cue combination, e.g., Landy and Kojima (2001); Rivest and Cavanagh (1996). Nonetheless, optimal cue combination does fit our data much better than the cue-switching model, where the predicted sigmas are on average 43% too large. The reason for that large predicted sigma is apparent in Figure 6—in the cue-switching model, the psychometric functions, rather than just the PSEs, are averaged. The interleaving of the mutually noncoincident psychometric functions for luminance and depth cues creates an average function that is shallower than either one.
As Figure 6 also illustrates, the psychometric functions derived from the two models can differ in shape as well as in position and scale. The overall goodness of fit of the models is best assessed from a comparison of the likelihood of the particular observed sequence of responses over all trials under each of the competing models. Consider first the depth task. If we first optimize each model’s free parameter (depth cue weight or depth selection probability) for each subject and stimulus condition, the probability of the data is greater for the cue-combination model in 11 of 12 cases, on average by about a factor of 10. Consequently when all depth task data are considered together, the balance of probability decisively favors cue combination, with a likelihood ratio of $9.4 \times 10^{10}$. A related but distinct alternative to the likelihood ratio is the Bayes factor, which expresses the relative likelihood of the two models with no prior constraint on the parameter values and no prior difference in likelihood. This requires integrating the probability of the data over all possible values of the free parameters under each model. The Bayes factor is the ratio of the integrals. The Bayes factor is less extreme than the likelihood ratio for optimized parameter values; this is expected because the models become equivalent in the limit where the respective parameters both approach one or zero. But at $3.8 \times 10^{8}$, the Bayes factor still decisively favors the cue-combination model. Because the models become equivalent in the limit of low-depth cue importance, the corresponding results for the luminance task are less decisive, but the Bayes factor of 18.7 for and the likelihood ratio of $2.7 \times 10^{7}$ again favor cue combination in both cases.

Neither model is an acceptable fit to any one condition’s data trial by trial. This is not unexpected, since the models are heavily idealized. Notably, they assume no variation in response bias across conditions and no fluctuation in observer bias or sensitivity over trials (because they assume binomial variability only in the judgments). Other inexact idealizations in our chosen implementation of the models include Gaussian variability (cumulative normal psychometric functions), 0.015 probability for finger error, etc. Since neither model is strictly acceptable, interpretation of the ostensibly conclusive evidence in favor of the cue combination alternative requires caution. Each model represents a point in a high dimensional space of recognized and unrecognized parametric influences on the data, including those few just mentioned and doubtless many others, along with the one freely varied model parameter. The variation of likelihood with these many unknown variables defines a complex many-dimensional probabilistic landscape with a high peak at ground truth. The cue-combination model has much higher likelihood than the switch model, but both are far below the peak, and they may represent distinct subpeaks. Thus it is entirely possible that the peak could be reached more easily by some modification of the switch model rather than the cue-combination model. But given the working assumptions made in implementing the models, the advantage lies overwhelmingly with cue combination. Other researchers using cue conflict paradigms like ours have also found that cue combination fits the data better than cue switching (e.g., Landy & Kojima, 2001).

**Mandatory or voluntary fusion?**

Another interesting implication of these data is that there is not a single fused percept that directly determines the judged location of both the luminance and the depth edge—otherwise we would have observed complementary (but opposite) shifts, (e.g., if the depth judgment was 75% of the offset between edges, the luminance shift should be 25%). Instead, the shifts never added up to 100%, indicating that the perceived edge location always regressed somewhat toward the actual location. Thus, the way that the two cues are combined must depend on the intended judgment. It would be interesting to see if this difference would persist if we made subjects produce both judgments in each trial, e.g., first for depth and then luminance. Would the first task “lock in” the fused percept, such that the same weighting would also be used for the second judgment? Or does the visual system maintain both luminance-edge and depth-edge percepts separately, even after cue combination has shifted their perceived locations? At least one study (Hillis, Ernst, Banks, & Landy, 2002) has evidence against this, instead suggesting that only a single fused cue remains, a phenomenon termed mandatory cue fusion. They showed this for perceived slant defined by texture and disparity for small amounts of cue conflict. It is possible that the cues still exerted influence on each other after they were no longer mandatorily fused, but their paradigm could not test this. Indeed, Knill (2007) showed that cues can still be combined (albeit with small weights) when there is a great degree of conflict between them. This is an interesting area for future research.

Our results could also occur through perceptual fusion. But alternatively, the subject might see the depth and luminance edges as separate, and consciously decide on a compromise estimate of the location. Since in our experiments the edge was seen as a single entity, the experience of our subjects supports the preconscious perceptual fusion explanation. The finding that depth weights are lower in the depth task (Figure 7) complicates the perceptual fusion hypothesis, but does not refute it. To account for the result, we need only assume that the preconscious processes leading to perceptual fusion are task contingent; what is excluded
is a strict segregation of preconscious from conscious processing with no feedback from the latter to the former.

But perhaps the competing idealizations of mandatory fusion versus conscious estimation are themselves too simple. Presumably the subject consciously registers a host of relevant impressions originating from the edge and from its nearby context. Thus, instead of one input (a perceptual fusion), or two inputs that are clearly distinct but are reconciled in a conscious best estimate, many introspectively accessible events may affect the decision, and each such event may involve some preconscious processing that allows interaction between the neural traces created by the depth and luminance cues. On this view, both preconscious and voluntary processes can influence the decision made by the subject, and their relative importance could vary across conditions. In particular, the larger misalignments might give more scope to voluntary processes, allowing the depth-cue weight to become more task contingent. Our data give some support to this, with a marginally significant interaction between luminance-depth offset and task ($p = 0.046$ in an analysis of variance on the differences between best-fitting depth weight and optimal weight). While we did not explicitly ask subjects if the two edges could be perceived as separate, the sigma data from the single-cue conditions suggests that subjects might have been able to tell that the edges were separate at least some of the time in the 5.6 offset conditions. Author and subject AR noted that on some trials it did appear that there might be two edges but could not tell with certainty. It would be interesting to extend the distance between edges and also ask on each trial if one or two edges were perceived to see how the attraction falls off as a function of the distinctness of the two edges.

Conclusion

Our data suggest that the perceived edge location is a function of both the depth-defined edge and the luminance-defined edge, with the luminance edge tending to play a larger role for the stimuli parameters tested in this experiment. This was very clear when subjects were asked to judge the depth-edge location. Our data for the luminance-edge judgments is weaker, but entirely consistent with this story nonetheless.

Our data rule out Morgan’s (1986) cue-switching model and are roughly compatible with an optimal cue-combination model, especially for the depth-edge judgments. Thus we can say that part of the apparent spatial resolution of perceived depth is derived from the spatial resolution of the perceived luminance. Low-resolution depth information can be used to assign depth values to regions defined by the higher resolution representation of luminance, much as chromatic signals, also delivered with relatively poor spatial resolution, can assign color to regions demarcated by luminance contrast (Boynton, Hayhoe, & MacLeod, 1977; Williams, MacLeod, & Hayhoe, 1981). This is not an entirely one-way relationship, however, with luminance trumping disparity. Both edge types contribute to the perceived depth-edge location, and furthermore this also occurs for luminance-edge judgments. Thus, a simple filling-in model, where somewhat sparse depth values are simply averaged within regions defined by luminance edges does an imperfect job of explaining our results, though it may still be needed to explain stimuli with sparse textures, etc. Instead, all cues contribute to the final percept, with an adaptive weighting depending on the task and the acuity with which that cue is perceived.

Keywords: sensor fusion, cue combination, edge detection

Acknowledgments

We thank Jody Mac and Matt Scott for their help with data collection. Don MacLeod and Alan Robinson were supported by NIH Grant EY01711 and NSF grant NSF CCF-1065305.

Commercial relationships: none.
Corresponding author: Alan Robinson.
Email: robinson@cogsci.ucsd.edu.
Address: Department of Psychology, University of California, San Diego, La Jolla, CA, USA.

References


