

Detection of changes in luminance distributions

Thomas Y. Lee

Department of Psychology, University of Pennsylvania, USA



David H. Brainard

Department of Psychology, University of Pennsylvania, USA



How well can observers detect the presence of a change in luminance distributions? Performance was measured in three experiments. Observers viewed pairs of grayscale images on a calibrated CRT display. Each image was a checkerboard. All luminances in one image of each pair consisted of random draws from a single probability distribution. For the other image, some patch luminances consisted of random draws from that same distribution, while the rest of the patch luminances (*test patches*) consisted of random draws from a second distribution. The observers' task was to pick the image with luminances drawn from two distributions. The parameters of the second distribution that led to 75% correct performance were determined across manipulations of (1) the number of test patches, (2) the observers' certainty about test patch location, and (3) the geometric structure of the images. Performance improved with number of test patches and location certainty. The geometric manipulations did not affect performance. An ideal observer model with high efficiency fit the data well and a classification image analysis showed a similar use of information by the ideal and human observers, indicating that observers can make effective use of photometric information in our distribution discrimination task.

Keywords: lightness/brightness perception, computational modeling, detection/discrimination, perceptual organization

Citation: Lee, T. Y., & Brainard, D. H. (2011). Detection of changes in luminance distributions. *Journal of Vision*, 11(13):14, 1–16, <http://www.journalofvision.org/content/11/13/14>, doi:10.1167/11.13.14.

Introduction

The visual system's representation of objects includes percepts that correlate with object surface reflectance. In general, these include color as well as perceptual correlates of object material properties, such as glossiness (Maloney & Brainard, 2010). The retinal image, however, does not provide an explicit representation of object reflectance. Rather, image intensities depend both on object reflectance and the properties of the illumination. To produce stable perceptual representations of object surface reflectance, the visual system must process the retinal image to minimize effects of variation in the illumination. Figure 1 shows an achromatic image where there is large spatial variation in the illumination.

A number of theorists have postulated that the stabilization of object appearance occurs in two stages (Adelson, 2000; Gilchrist, 1977; Gilchrist et al., 1999; Kardos, 1934; Koffka, 1935). The first stage segments the image into regions that each have roughly constant illumination. The second stage then, in effect, estimates the illuminant within each region and uses the estimate in its conversion between luminance and lightness for that region.

What information could the visual system use to segment the image according to illumination? Photometric cues provide one source of information that can indicate illumination changes. Surface albedo is typically thought to vary over about a 30 to 1 range in natural scenes (see, for example, reflectance data summarized in Wyszecki & Stiles, 1982). Thus, if two grayscale image regions vary in

luminance by a factor much larger than 30, they are unlikely to share a common illuminant. In the image shown in Figure 1, it is easy to imagine that such a difference in image intensity helps mediate the impression that the floor is lit by two distinct illuminants.

On the other hand, a number of geometric factors may also correlate with illumination changes. One, for example, is distance: The further apart two surface patches are in a scene, the less likely it seems that they will share a common illuminant. Accordingly, experiments have found a decreasing influence of contextual surfaces on target surface appearance with increasing distance (Kurki, Peromaa, Hyvärinen, & Saarinen, 2009; Reid & Shapley, 1988; Shimozaki, Eckstein, & Abbey, 2005; Spehar, Debonet, & Zaidi, 1996). Closely related is the idea that coplanar surfaces are more likely to share a common illuminant than surfaces oriented differently within a scene (Gilchrist, 1980). Various cues are available to indicate surface orientation in a scene (e.g., binocular disparity), as well as changes in orientation of groups of surfaces (e.g., Ψ -junctions, Sinha & Adelson, 1993). Finally, the luminance relations across certain geometric configurations may signal illumination boundaries (e.g., X- and T-junctions, Todorović, 1997).

Despite the centrality of segmentation in theories of lightness, little is known about how well observers can use the type of photometric information induced by changes of illumination to segregate scenes. For achromatic images, changing the illumination changes the statistical distribution of the luminances reaching the observer, because the luminance distribution arises as the product



Figure 1. Image containing regions with different illumination. The parts of the garden seen through the windowpanes in direct sunlight are adjacent to shadowed walls inside the room. However, the two lighting environments are very different. Image taken from <http://www.flickr.com/photos/molinaris/3585205048/> and used with permission of the photographer.

of the illuminant intensity and the underlying distribution of surface albedos. In the present paper, then, we step back from the specifics of illuminant-based segmentation and ask the more basic question of how well observers can detect within-image changes in the distribution of image luminances. That is, we sought to study fundamental aspects of this ability, using simple stimuli that did not evoke percepts of illuminated surfaces. We used checkerboard stimuli and asked observers to judge which of two images contained a region where the luminance statistics differed from those in the rest of the scene. We also compared the data to predictions from an ideal observer model. Finally, we asked whether manipulating the geometric structure of the images affected performance on our segregation task. The measurements provide baseline information that can be exploited in future experiments that study illumination segmentation in more complex scenes to determine the role of additional information sources.

Overview

Observers viewed displays consisting of two side-by-side grayscale checkerboards (e.g., Figure 2). The luminances of the patches in one of the checkerboards were drawn from a single probability distribution (a truncated Gaussian); for the other checkerboard, two distributions were used. Observers indicated which of the two checkerboards contained patches with luminances drawn from two distributions. The experiment embodies an abstracted version of the illuminant segmentation task that observers confront in natural viewing.

Performance was measured in three experiments. In Experiment 1, we varied the number of patches drawn

from the second distribution, as well as observers' uncertainty about the spatial locations of these patches. Performance was compared to that of an ideal observer model, as well as a number of alternative simpler models.

In Experiments 2 and 3, geometric manipulations were introduced. In Experiment 2, these consisted of (a) varying the contiguity of the patches drawn from the second distribution and (b) changing the spatial arrangement of the images to introduce Ψ -junctions. This was done to test the idea that the geometric cues would limit integration of photometric information to patches grouped together by those cues. In Experiment 3, binocular depth cues were used to separate patches into two different depth planes, again with the idea that this might lead to processing grouped by depth.

Experiment 1

Methods

Observers

Four observers (one male, three females, mean age = 23) participated in this experiment. Each observer came to the laboratory for two sessions and was compensated for his or her time. The observers all had Snellen acuity of at least 20/40 (corrected) and scored at least 36/38 correct on the Ishihara (1998) color plates.

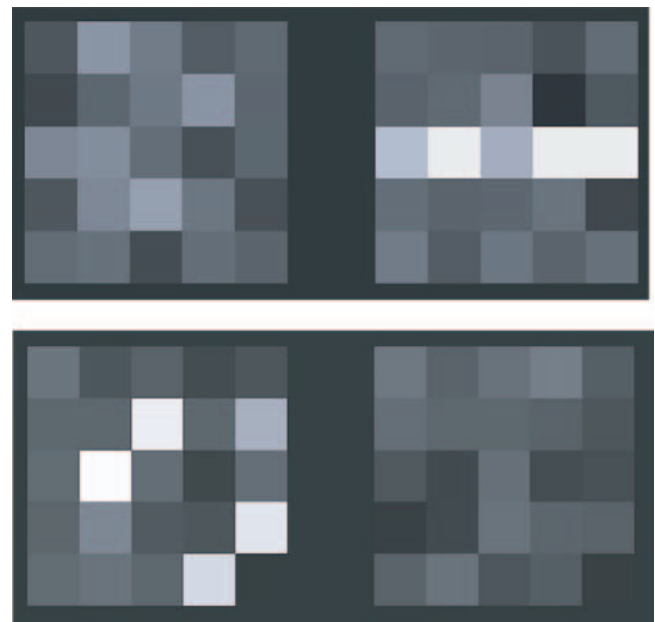


Figure 2. Examples of Experiment 1 stimuli. (Top) Location-known condition. (Bottom) Location-unknown condition. In both examples, there are 5 test patches in one of the two checkerboards. For the top panel, they are in the center row on the right; for the bottom panel, they are scattered in the left checkerboard.

Stimuli and setup

The stimuli on each trial consisted of two side-by-side grayscale checkerboards (Figure 2). These were presented on a calibrated ViewSonic G220fb CRT monitor. Observers viewed the monitor from a distance of 560 mm, with viewing position stabilized by a headrest-chinrest assembly. Each checkerboard consisted of five rows of five patches, with each patch 29×30 mm ($2.97^\circ \times 3.07^\circ$).¹ The overall size of the 5×5 checkerboards was thus 145×150 mm ($14.75^\circ \times 15.26^\circ$). The two checkerboards were presented against a dark gray background (4.4 cd/m²) and were separated horizontally by 48 mm (4.91°). The CIE 1931 xy chromaticity of the background and checkerboard patches was held fixed at $[0.30, 0.30]$.

On each trial, one of the two checkerboards (left or right) was randomly designated the *test checkerboard* and the other the *standard checkerboard*. The luminances for the 25 patches in the standard checkerboard and almost all of the patches in the test checkerboard were randomly drawn from a truncated Gaussian distribution with a mean of 15.0 cd/m², a standard deviation of 5.0 cd/m², and a truncation range of $[5.0, 50.0]$ cd/m². For the test checkerboard, the luminances of the remaining patches, which we refer to as the *test patches*, were drawn from a different distribution. The test patch distribution was a truncated Gaussian whose mean and standard deviation were larger than the standard distribution by a multiplicative constant. Across trials, this constant varied between 1 (minimum) and 2.5 (maximum). These corresponded to test patch distributions with a mean of 15.0 cd/m², a standard deviation of 5.0 cd/m², and a truncation range of $[5.0, 50.0]$ cd/m² (minimum) and with a mean of 37.5 cd/m², a standard deviation of 12.5 cd/m², and a truncation range of $[12.5, 125.0]$ cd/m² (maximum).

Procedure

The observer's task on each trial was to indicate via a button press which of the two checkerboards was the test checkerboard. We found during pilot experiments that the most effective instructions were to ask observers to identify the checkerboard that contained some patches drawn "from a larger range of luminances than the rest of the patches." These instructions were used. The full instructions are provided in the Supplementary material available at <http://color.psych.upenn.edu/supplements/distribdiscrim/>. Feedback was provided by a tone whenever the observer made an error.

The multiplicative constant for the test distribution parameters was adjusted trial by trial based on whether the observer's response was correct, using a 2-down 1-up staircase procedure (Levitt, 1971): The test distribution parameters decreased after every two correct responses and increased after every incorrect response. The threshold test distribution, parameterized by its mean, at which observers were correct on 75% of the trials was estimated by fitting a Weibull psychometric function to all of the

data, using the psignifit toolbox version 2.5.6 (Wichmann & Hill, 2001; see <http://bootstrap-software.org/psignifit>). The psignifit software implements a maximum-likelihood fit of the Weibull parameters along with a lapse rate parameter.

Two variables were manipulated across conditions: the number of test patches in the signal checkerboard and the observers' certainty about the location of these patches. The number of test patches varied between 1 and 5. The test patches were either at fixed locations known to the observer (in the center row of the signal checkerboard; *location-known condition*; top panel of Figure 2) or were randomly selected on each trial from the 25 checkerboard patches (*location-unknown condition*; bottom panel of Figure 2). This resulted in a total of 10 conditions (5 patch numbers \times 2 levels of certainty). Conditions were blocked and observers were informed beforehand which condition was being tested. The order of conditions was randomized for each observer. For each condition, observers ran five blocks of 100 trials before moving onto the next condition.

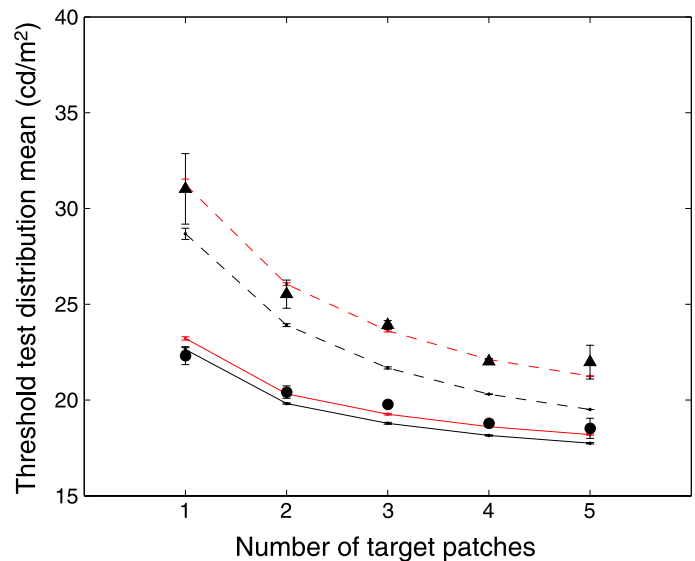


Figure 3. Average (across observers, $n = 4$) threshold plotted as a function of number of test patches, for the location-known (solid circles) and location-unknown (solid triangles) conditions. Error bars show ± 1 SEM. Ideal observer simulation data (dots connected by solid black lines) are also shown (solid line, location-known; dashed line, location-unknown). Error bars show ± 1 SEM over multiple simulation runs, except for the 4 and 5 test patch points for the location-unknown condition where only a single run was done. Red lines/red solid dots show the ideal observer data scaled by a multiplicative constant to best fit the experimental data. Individual observer data for this experiment, as well as for Experiments 2 and 3, are provided in the Supplementary material available at <http://color.psych.upenn.edu/supplements/distribdiscrim/>.

Results

Figure 3 plots average thresholds as a function of number of test patches. Two broad effects are apparent. First, in both the location-known and location-unknown conditions, thresholds fall with increasing number of test patches, with the largest drop occurring between one and two test patches. Second, knowledge of test patch location decreased thresholds for all test patch numbers. On average, the location-known condition thresholds were 4.9 cd/m^2 lower than the location-unknown thresholds.

Ideal observer thresholds are also plotted in Figure 3. The ideal observer calculations are described in Appendix A. The broad patterns visible in the experimental data are also apparent for the ideal observer: Thresholds decrease with increasing numbers of test patches and thresholds are greater for the location-unknown case. In addition, the ideal observer thresholds are slightly lower than the human observer thresholds. We scaled the ideal observer data to fit the experimental data (red lines in Figure 3, separate scaling for location-known and location-unknown conditions). In each case, the ideal observer predicts the dependence of threshold on number of test patches. The scale factor required was slightly larger for the location-unknown case (1.03 for location known; 1.09 for location unknown). Thus, human performance is close to ideal in the location-known case, but uncertainty in test patch location adds an additional cost, beyond what would be experienced by an ideal observer faced with the same uncertainty.

Intermediate discussion 1

Although the data are consistent with an ideal observer model that efficiently integrates information from the test patch locations to judge which image contained the distributional change, it is possible that similar performance could be obtained for our stimuli using simpler strategies. We thus considered models based on three such strategies: (a) a *mean luminance model* that chooses the checkerboard with the larger mean luminance on each trial, (b) a *highest luminance model* that chooses the checkerboard containing the highest luminance patch on each trial, and (c) a *highest range model* that chooses the checkerboard with the highest luminance range on each trial. We constructed variants of these models for the location-known and location-unknown conditions. For the location-known condition, the mean luminance and highest luminance were evaluated only over the known test patch locations in each checkerboard, while the range was obtained by subtracting the lowest non-test patch location luminance from the highest test patch location luminance. For the location-unknown condition, the mean, highest luminance, and luminance range were computed over all the locations in each checkerboard.

Figure 4 replots the data from Experiment 1 along with the predictions from each of these models. The data from the location-unknown condition clearly falsify the mean luminance model, as that model's dependence on test patch number has a very different form from the measurements. Both the highest luminance model and luminance range models, however, make predictions quite similar to those of the ideal observer model and are not ruled out by the data. Of note is that these models require only the most primitive form of integrating information across test patch locations: identifying which patches in each checkerboard image have the highest (and lowest) luminances.

To investigate further, we conducted a classification image analysis of the relationship between the trial-by-trial variation in the stimulus and the trial-by-trial responses, both for our human observers and simulations of performance based on the models. Such analyses can reveal the relative weighting of various sources of stimulus information that lead to the same level of overall performance (Abbey & Eckstein, 2006; Ahumada & Lovell, 1971; Gold, Murray, Bennett, & Sekuler, 2000; Murray, Bennett, & Sekuler, 2002). The intuition behind this kind of analysis is that the correlations between aspects of the stimulus that contribute to the decision and the response will be high, while the corresponding correlations for aspects of the stimulus that are irrelevant will be low.

To implement the analysis, we fit the coefficients of a logistic regression model (Alexander & Lutfi, 2004) to

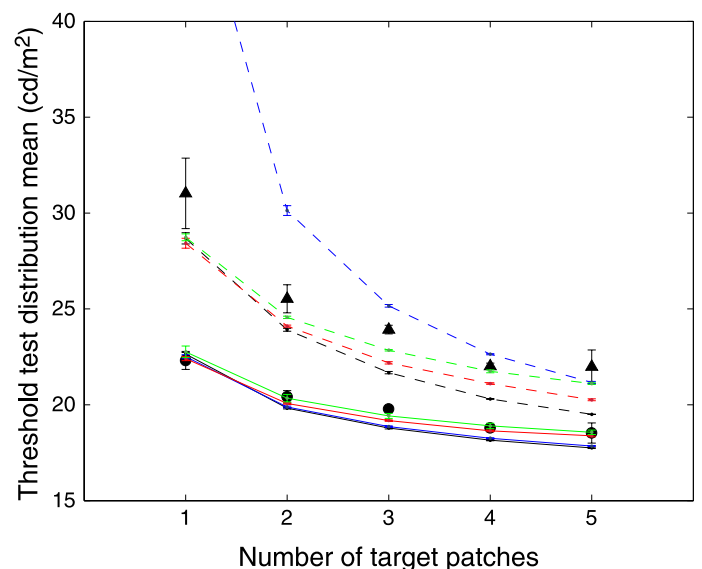


Figure 4. Data from Experiment 1, plotted with simulated thresholds from ideal observer model (black), mean luminance model (blue), highest luminance model (red), and highest range model (green). Circles and triangles are replotted from Figure 3. Solid lines connect points from location-known conditions; dashed lines connect points from location-unknown conditions. Error bars show $\pm 1 \text{ SEM}$.

estimate the weights placed on patch luminance differences from trial-by-trial responses:

$$\mathbf{r} = \frac{1}{1 + e^{-\mathbf{P}\mathbf{w}}} \quad (1)$$

In Equation 1, \mathbf{r} is a binary column vector coding responses (left/right) as 1 s and 0 s, \mathbf{w} is a row vector with the weights found by the regression, and \mathbf{P} is a matrix whose rows were per-trial vectors of luminances obtained from the stimulus. The values in the rows of \mathbf{P} were obtained as follows: on each trial, for each checkerboard (left and right), we took the luminances from the test patch locations, the two most luminous non-test patches, and the two least luminous non-test patches. We sorted the luminances for the test and non-test patches in descending order separately, for each checkerboard. Then, we took the differences between the corresponding luminance-ranked patches (i.e., the most intense test patch on the left minus the most intense test patch on the right, the second most intense test patch on the left minus the second most intense test patch on the right, etc.). The regression thus tells us how much weight is assigned to

the luminance differences between corresponding rank-ordered test patches across the two checkerboards and to the luminance differences of corresponding rank-ordered non-test patches at the high and low luminance ends of the non-test patch range. We analyzed only the data from the location-known conditions in this way, as we found that our data set did not have sufficient power to provide reliable estimates of the weights when all checkerboard squares were considered.

The estimated weights are plotted in Figure 5 for the human observers, the ideal observer model simulation, and the highest luminance model simulation. Because our interest is in the relative importance of the luminance-ordered test patch differences, we normalized the weights within model/condition by the weight assigned to the most luminous test patch.

For both model simulations, the weights for the non-test patches cluster around zero, indicating minimal influence on the decision, consistent with the fact that these test patches have no influence on the model's decision. The scatter of the weights around zero provides a visual sense of how precisely the weights are determined, given the number of simulated trials. For the highest luminance

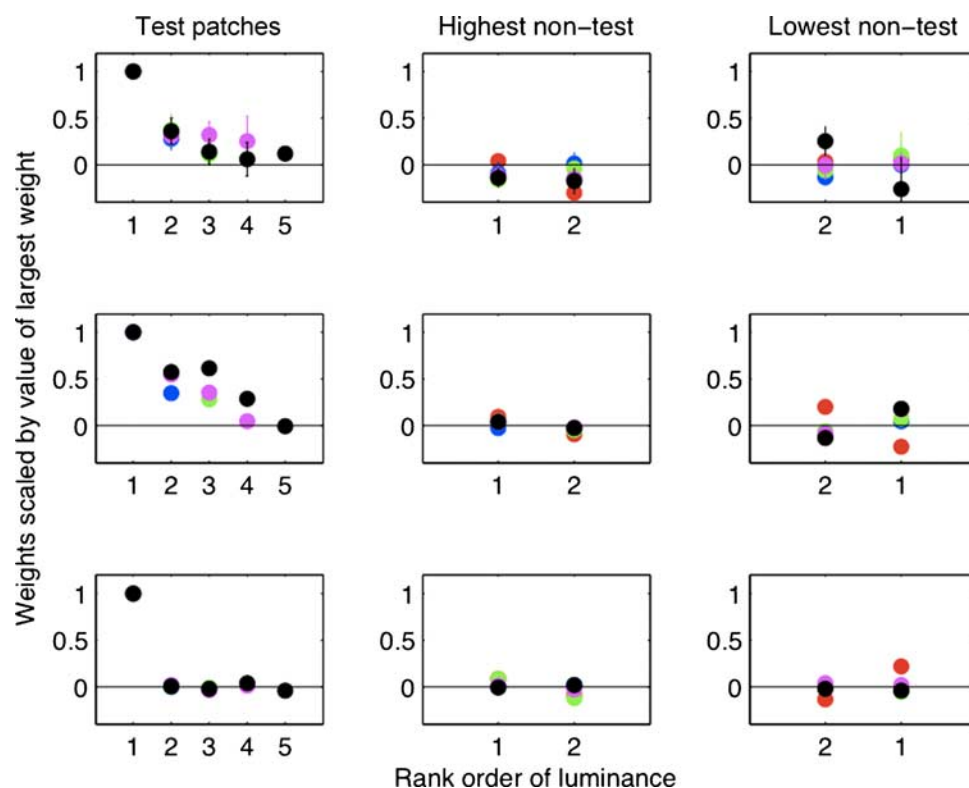


Figure 5. Estimated relative weights from classification analysis for (top row) human observers, (middle row) ideal observer model, and (bottom row) highest luminance model. The leftmost column graphs show the weights for the luminance rank-ordered test patches, the middle column graphs show the weights for the two most luminous non-test patches, and the rightmost column graphs show the weights for the two least luminous non-test patches. Only the weights for the location-known conditions are plotted. For 1 to 5 test patches, the color code is: red, blue, green, purple, black. The weights for human observers are the mean of four observers, and error bars are one standard error of the mean. For the model simulations, we matched the number of trials to that used in the human experiments and ran the simulations four times to match the number of observers. The weights shown for the simulations are the mean of these four runs.

model, only the most luminous test patch receives a high weight; the weights for the remaining test patches resemble those of the non-test patches. For the ideal observer, which integrates information from all test patches, the weights fall as a function of number of test patches. We also estimated weights for the other two models (plots provided in the Supplementary material available at <http://color.psych.upenn.edu/supplements/distribdiscrim/>). The test patch weights for the mean luminance model are equally high, as all test patches contribute equally to the average. The test patch weights for the highest range model resemble those of the highest luminance model but with an equally large negative weight on the least luminous non-test patch, as both these patches are necessary in the calculation of the luminance range.

If the human observers were using one of the strategies implemented in the models, then the pattern of their estimated weights should resemble the pattern from that model. The human observers' average weights more closely resemble those of the ideal observer than the other models. In particular, the human observers assign positive weight to multiple test patches. As with the weights from the ideal observer model, the human weights for the test patches decrease with test patch luminance. This analysis suggests that humans integrate photometric information over multiple test patches. The fact that overall the weights obtained for the human observer decrease more rapidly with ordinal test patch luminance than for the ideal observer is consistent with the fact that human efficiency with respect to the ideal observer is less than 1.

Experiment 2

Methods

Purpose

Experiment 1 established that observers can perform the luminance distribution discrimination task and the classification image analysis showed they are able to integrate information across patch locations. In **Experiment 2**, we explored the effect of other display manipulations on performance. Geometric cues were introduced in an attempt to negatively affect performance, by inducing segregation of the checkerboards into regions that were spatially incongruent with the different luminance distribution regions. The hypothesis we sought to test was that such geometric cues impose a mandatory segregation on the scene and prevent the use of photometric information from both sides of the geometrically cued boundary.

Suppose, for example, that a geometric cue grouped one of two test patch locations separately from the other test patch location. If the observer were unable to use information from the two regions, performance for the

two test patch cases with the geometric cue would then resemble performance for one test patch without the geometric cue. To put it another way, thresholds would be predicted to increase across the geometric manipulation.

Two types of geometric cues were tested: separation and the presence of Ψ -junctions. For the separation manipulation, the test patches were located non-contiguously with each other. For the Ψ -junctions, the shapes of the checkerboards were manipulated in a way consistent with folding part of the checkerboard in depth.

Observers

Three observers participated in this experiment. Each observer came to the laboratory for one session and was compensated for his or her time. All three had participated in **Experiment 1**.

Stimuli and procedure

The stimuli were presented using the same apparatus as for **Experiment 1**, and the luminance statistics were also the same.

Observers were tested in three main conditions, all using two test patches. The first was a replication of the location-known condition of **Experiment 1** using two test patches ("center"). The test patches were the center patch and the patch to its right within each checkerboard.

The second condition presented two test patches in a spatially separated configuration ("sep"). In this condition, the two test patches were the center patch and the lower right-hand corner patch of each checkerboard (top panel of **Figure 6**).

The third condition ("psi") presented two test patches on a square + parallelogram checkerboard (bottom panel of **Figure 6**). To generate this checkerboard, the rightmost vertices of the square checkerboard were shifted down a distance equal to two side lengths of each square (5.94°) and the next rightmost vertices down by one side length (2.97°). This manipulation introduced Ψ -junctions between the third and fourth columns of each checkerboard. Interpreted as a three-dimensional object, this checkerboard would appear to have its rightmost two columns folded in depth. Note, however, that this was a purely monocular manipulation: No binocular depth cues were used. The test patches were again the center patch and the patch immediately to its right.

Three control conditions were also run, all with one test patch. The first was also a replication of the location-known condition of **Experiment 1** but with only one test patch ("center"). The other two used the Ψ -junction checkerboard, with the test patch in the center ("psiC") or immediately to the right of the center ("psiR"). These control conditions tested the effect of introducing parallelograms on thresholds for a single square patch and for a single parallelogram patch.

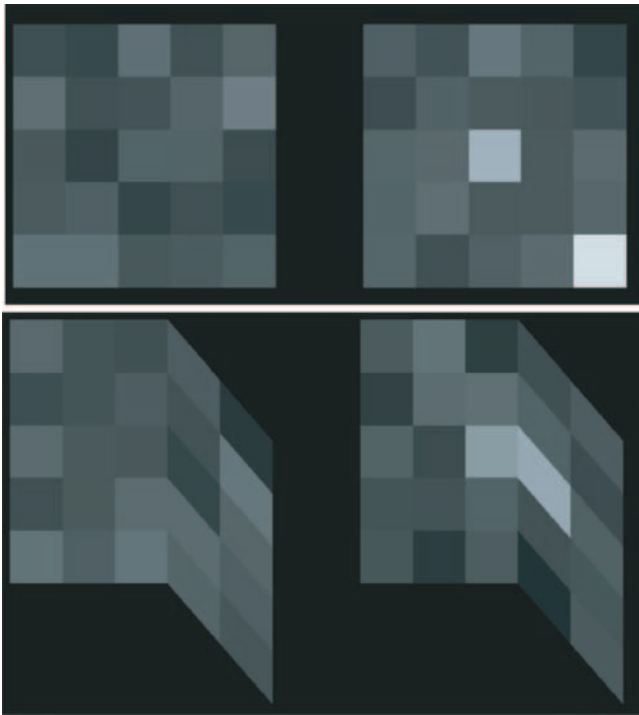


Figure 6. Examples of non-contiguous and Ψ -junction stimuli in Experiment 2. (Top) The target patches lie in the right checkerboard, center patch, and lower right-hand corner. (Bottom) The target patches lie in the right checkerboard, center patch, and parallelogram contiguous with it to the right.

In all conditions of Experiment 2, observers knew the locations where the test patch or patches could appear (location-known). The observer's task was the same as in Experiment 1—to pick the test checkerboard containing the test patches. Here, however, observers ran one block of trials from each of the conditions in random order, then repeated all the conditions in a different order, until five blocks of each condition were obtained. As in Experiment 1, observers were informed before each block of trials as to what condition was being run for that block. Psychometric functions were again fit to the data, and the threshold value was estimated as a measure of performance.

Results

Mean thresholds are plotted in Figure 7 for the various conditions tested. A repeated measures ANOVA with observer as a random factor indicated a significant effect of condition, $F(5,10) = 10.4$, $p \leq 0.001$. Examination of the plot suggests that this result was driven by the effect of test patch number: Conditions with two test patches led to lower thresholds than conditions with one test patch. This difference is comparable to the difference between the 1- and 2-test patch location-known conditions in Experiment 1, replotted in Figure 7 with X symbols.

According to the ideal observer model, only the number of test patches and knowledge of their locations should

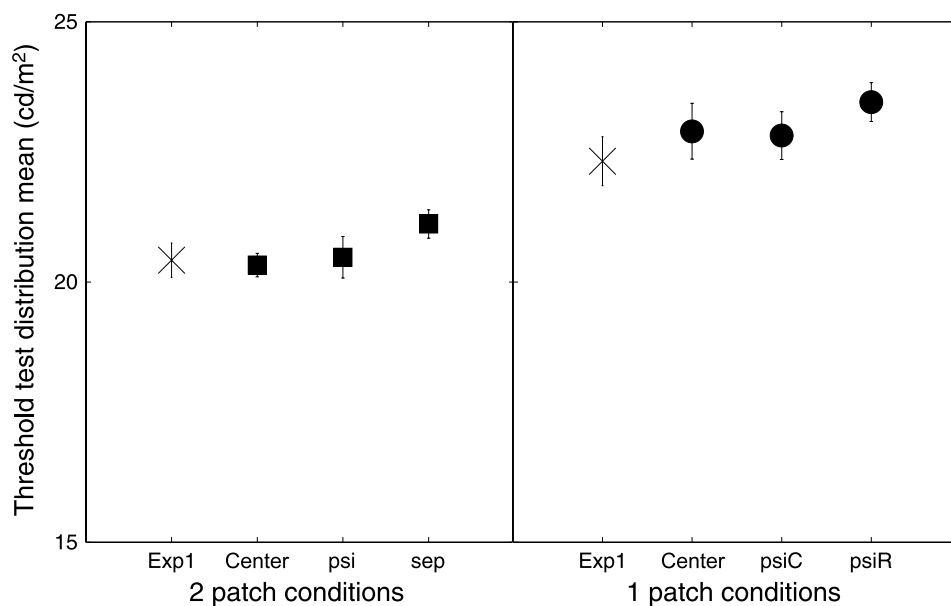


Figure 7. Mean thresholds from three observers in Experiment 2. (Left) Two-test-patch conditions: center row, Ψ -junction checkerboard, separated. (Right) One-test-patch conditions: center row and two control conditions with Ψ -junction checkerboards. The 1- and 2-test-patch means from the location-known condition in Experiment 1 are plotted with the X symbols for comparison. Error bars are one standard error of the mean.

Experiment 2: Test patch 2 weights

	<i>center</i>	<i>psi</i>	<i>sep</i>
Mean weight	0.57	0.46	0.17
SEM	0.32	0.21	0.13

Experiment 3: Test patch 2 weights

	<i>bp</i>	<i>fp</i>	<i>m</i>	<i>rp</i>
Mean weight	0.61	0.30	0.43	0.27
SEM	0.20	0.14	0.16	0.06

Table 1. Mean estimated weights ($n = 3$) for second test patch in 2-patch conditions, [Experiments 2](#) and [3](#). Conditions are labeled as described in the text.

affect performance. If thresholds for the *psi/sep* two-patch conditions are greater than those for the *center* two-patch condition, then those manipulations have a detrimental effect on human performance not accounted for in the model. However, the data in the left panel of [Figure 7](#) show a minimal effect on threshold for those manipulations: A repeated measures ANOVA with observer as a random factor was not significant, $F(2,4) = 1.27$, $p = 0.37$. In addition, note that the non-significant trend toward an elevated threshold for the *sep* condition is small relative to the effect of test patch number.

The thresholds from the control conditions using one test patch were also not different from each other, $F(2,4) = 4.08$, $p = 0.68$. Simply introducing parallelograms into the image does not appear to affect performance for a single test patch, regardless of its shape.

The weights for the second test patch in the three 2-patch conditions were also estimated, and the means across observers are listed in [Table 1](#). The weights were all positive and not different from each other, $F(2,6) = 0.80$, $p = 0.49$. This pattern suggests that similar strategies were used regardless of scene geometry and that both test patch luminances affected the decision.

Intermediate discussion 2

We did not find evidence for an effect of the geometric manipulations in [Experiment 2](#). Interestingly, this suggests that for the location-unknown condition of [Experiment 1](#), the performance decrease may have been driven primarily by uncertainty and not by the non-contiguity of the target patches. That is, the fact that a random draw of patches within a checkerboard in the location-unknown condition of [Experiment 1](#) often resulted a non-contiguous configuration of test patches may not have affected performance directly, at least if the results from the single separation manipulation of [Experiment 2](#) generalize to more test patches.

One possibility for our failure to find an effect of geometry in [Experiment 2](#) is that the geometric manipulations we

used were either ineffective at introducing segmentation, or at least not effective enough to overcome a larger effect of the provided photometric cues. In [Experiment 3](#), we introduced binocular disparity as a different cue to geometric segregation. The logic was the same as in [Experiment 2](#)—would performance for a 2-test-patch condition resemble performance for a 1-test-patch condition in the presence of disparity cues that segregated the test patches into separate depth planes?

Experiment 3

Methods

Observers

Three observers (1 male, 2 females, mean age = 20) participated in this experiment. None had participated in [Experiment 1](#) or [2](#). Each was screened using the same screening procedure and criteria as in [Experiment 1](#). An additional screening test for stereopsis was also used. This test was performed using the same apparatus as for the experiment (see below). A 2.62° square patch appeared either in front or at the back of a fixed background plane due to binocular disparity, and observers were asked to indicate where the patch appeared (“front” or “back”). The background plane was rendered to be 764 mm away from the observer. The simulated depth (i.e., amount of disparity) was adjusted via a 2-down 1-up staircase procedure (Levitt, 1971) to estimate 71% accuracy on this task. All observers had thresholds below a simulated depth change of 10 mm.

Setup and stimuli

The stimuli were presented on a different apparatus from the previous two experiments. This stereo apparatus, illustrated in [Figure 8](#), consisted of two calibrated hp p1230 CRTs controlled by a single computer. Observers sat with their heads stabilized via a chinrest in front of a black felt-covered faceplate. They viewed the stimuli through two 30 mm \times 30 mm square openings in the plate. The distance between the centers of the two openings was 40 mm. A black cardboard divider sat perpendicular to the faceplate, preventing overlap between the visual inputs to the two eyes.

Each eye received input from a CRT whose light was reflected off an angled mirror before reaching the eye. The optical distance of the CRTs to the eyes was approximately 764 mm. The apparatus was aligned by replacing the mirrors with beam splitters and aligning a grid image on each monitor to a physical grid located 764 mm from the eyes.

The stimuli closely resembled the square checkerboards from the location-known condition in [Experiment 1](#), with the following changes. The checkerboards were rendered to be 764 mm away from the observer ($11.21^\circ \times 11.21^\circ$),

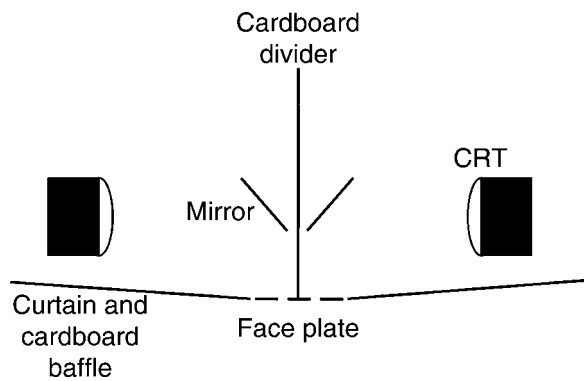


Figure 8. Bird's-eye-view schematic of stereo apparatus. Two CRTs and two mirrors were separated by a black cardboard divider. Each mirror reflected the light from one CRT to one eye of the observer, who was seated in front of the faceplate. Curtains hid the CRTs from the observer's view.

and the space between checkerboards was 28 mm (2.10°). Owing to a smaller monitor gamut, the standard distribution had a mean of 6.0 cd/m^2 and a standard deviation of 3.0 cd/m^2 , truncated on the interval $[1.5, 15] \text{ cd/m}^2$. The test distribution had minimum parameters equal to the standard distribution parameters. Its maximum parameters were a mean of 17.7 cd/m^2 , a standard deviation of 8.8 cd/m^2 , and a truncation range of $[4.4, 44.2] \text{ cd/m}^2$.

The number of test patches was always one or two, in different conditions. For the 1-test-patch conditions, the test patch was the center patch of the checkerboard; for

the 2-test-patch conditions, the test patches were the center patch and the patch immediately to its right.

The experimental condition was a 2-test-patch condition ("Mixed"). However, the center patch was rendered with a binocular disparity, so that it appeared to float 100 mm in front of the rest of the checkerboard. The size of the floating patch was increased to $34.5 \text{ mm} \times 34.5 \text{ mm}$ ($2.59^\circ \times 2.59^\circ$) so that with the disparity it appeared to be approximately the same size as it was while coplanar with the checkerboard, $30 \text{ mm} \times 30 \text{ mm}$ ($1.12^\circ \times 1.12^\circ$). The test patch to the right remained in the same plane as the checkerboard.

A series of control conditions were also tested. The 1- and 2-test-patch cases from the location-known condition of [Experiment 1](#) were replicated on this stereo apparatus ("BackPlane," for 1 and 2 test patches). Additionally, two control conditions using 1 and 2 test patches brought forward in depth using binocular disparity were also run ("FrontPlane," for 1 and 2 test patches). These conditions were used to check for changes in performance due to the presence of binocular disparity in the stimulus, without the segregation of test patches into two depth planes. Finally, a random depth plane condition ("RandomPlanes") was tested. This was a 2-test-patch condition where the potential test patches on both checkerboards could each randomly appear at either depth plane. The depth arrangement of the two checkerboards on every trial was yoked, e.g., if both potential test patches on one were closer in depth, so were the corresponding test patches on the other.

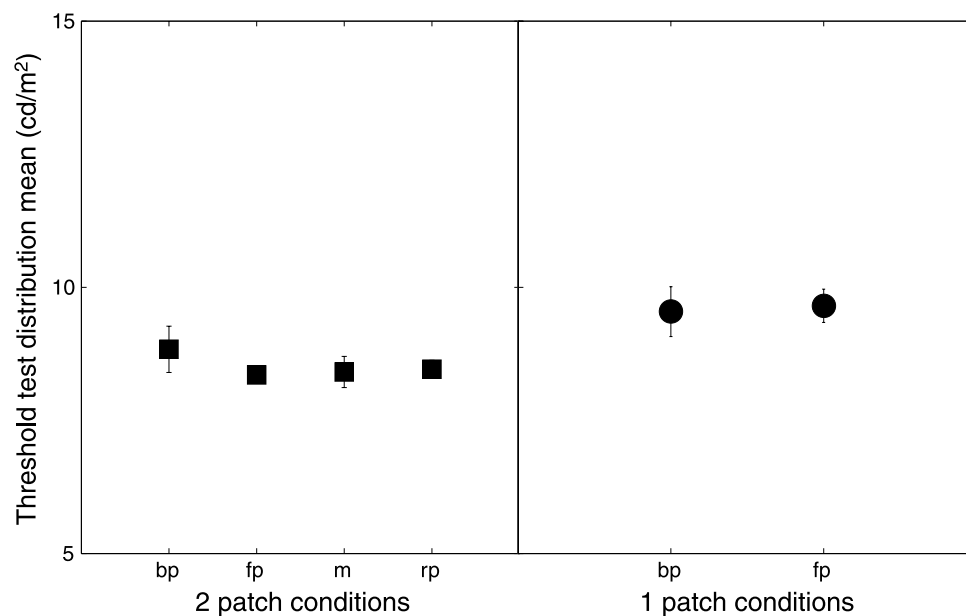


Figure 9. Mean thresholds for three observers in [Experiment 3](#). (Left) Two-test-patch conditions. (Right) One-test-patch conditions. Error bars are one standard error of the mean. "BackPlane" (bp) refers to the test patches being in the same plane as the checkerboard; "FrontPlane" (fp) refers to the test patches in a different plane from the checkerboard. The "Mixed" (m) condition has one test patch in the same plane and one in a different plane from the checkerboard. The depth planes for "RandomPlanes" (rp) were randomized across trials as described in the text.

Procedure

The procedure was similar to the one used in [Experiment 2](#). Observers completed five blocks of all six conditions in random order and knew in advance of each block which condition was being run. Each block consisted of 100 trials.

Results

The mean data for three observers are plotted in [Figure 9](#) as a function of condition. A repeated measures ANOVA with observer as a random factor indicated a significant main effect of condition, $F(5,10) = 4.4$, $p \leq 0.02$. As with [Experiments 1](#) and [2](#), thresholds for the 1-test-patch conditions are higher than those for the 2-test-patch conditions.

The critical comparisons are for the two-patch conditions. Here, thresholds did not differ across the conditions we ran. A repeated measures ANOVA with observer as a random factor revealed no effect of condition, $F(3,6) = 0.6$, $p = 0.66$. That is, separating the two test patches in depth did not affect threshold nor was threshold different in the condition where the depth of the two test patches was randomized on every trial.

As in [Experiment 2](#), the weights for the second test patch in the four 2-patch conditions were also estimated and listed in [Table 1](#). Again, the weights were not found to be different across condition, $F(3,8) = 1.08$, $p = 0.41$.

General discussion

Summary

In [Experiment 1](#), we investigated how well observers use photometric information to detect changes in luminance distributions. We found that observers perform this task with high efficiency, relative to an ideal observer. This was true both when the test patch locations were known and when there was uncertainty about these locations. Efficiency was somewhat higher, however, in the location-known case. In both cases, the dependence of threshold on the number of test patches was also well modeled by the ideal observer. Our classification image analysis of the trial-by-trial responses showed that although high efficiency for our stimuli could be achieved with a simple strategy that only relied on the highest luminance in the two checkerboard images, observers appeared to follow the ideal observer in that they integrated photometric information from multiple locations. In [Experiments 2](#) and [3](#), we showed that simple geometric manipulations did not affect performance on our task.

Taken together, our findings suggest that the visual system is quite sensitive to the sorts of changes in luminance distributions that might indicate changes of illumination within a scene in the presence of uncertainty about surface albedo. That is, the visual system is fairly efficient at using such photometric cues to perform a task that models illuminant segregation. Although for low-level perceptual tasks human efficiency is generally low compared to that of ideal observers ($\sim 5\%$; Banks, Geisler, & Bennett, 1987), higher efficiencies have been reported previously for tasks such as symmetry detection (Barlow, 1980).

Overall, thresholds in our experiments changed only with test patch number and uncertainty about test patch location but not as a result of manipulations of test patch separation or geometric cues that might segment the image. Our data did not reveal a significant effect of introducing geometric segregation cues on performance. A caveat, of course, is that conclusions in this regard hold only up to the power of our data. Our data did, however, contain sufficient power to reveal changes in performance between presentation of one and two test patches.

The lack of geometric effects in our experiments is perhaps not surprising, given that our task was structured so that photometric cues provided the only information available to perform the task. What our results do show is that when photometric information is available, the visual system can integrate this information across spatial boundaries created by geometric factors. An interesting question, but one different from that we studied, is how geometric and photometric information interact when both types of information are task-relevant.

Relation to other studies

Our work makes contact with a number of related threads in the literature. We touch on these below.

Illumination perception

The literature on illumination perception is much smaller than that on surface perception, but there are a number of studies that relate to our current work. Koenderink, Pont, van Doorn, Kappers, and Todd (2007) asked observers to adjust the illuminant impinging on one object in a scene so that it matched the illumination field of the scene as a whole. The fact that observers could do this with reasonable precision indicates that they could discriminate local changes in illumination within a single image, broadly consistent with our results. Earlier studies (e.g., Beck, 1959, 1961; Kozaki & Noguchi, 1976; Noguchi & Kozaki, 1985; Noguchi & Masuda, 1971; Oyama, 1968; Rutherford & Brainard, 2002) also employed illuminant matching or explicit judgments of the illuminant but considered between- rather than within-scene variation.

Gerhard and Maloney (2010a) showed that observers can differentiate between illumination changes common to all surfaces in a scene and illumination changes that vary from one scene location to another. In a second paper, they found that observer performance in a task that required estimating the motion of a collimated light source was well predicted by an ideal observer model that interpreted photometric changes in the context of noisy knowledge about surface geometry (Gerhard & Maloney, 2010b). In their case, however, the focus was on geometric changes in the three-dimensional location of an illumination source rather than on the efficacy of photometric cues for illumination discrimination. Similarly, Khang, Koenderink, and Kappers (2006) found that observers could match illuminant source directions across scenes. In the color domain, Craven and Foster's (1992) work has a similar flavor, although they cast their measurements in terms of discriminations between a spatially global illumination change and changes in the surface reflectances of the objects within a scene (see also Nascimento & Foster, 2000).

Role of geometric cues

A number of studies show that changes in perceived geometry affect perceived surface lightness (Boyaci, Maloney, & Hersh, 2003; Gilchrist, 1977, 1980; Hochberg & Beck, 1954; Knill & Kersten, 1991; Radonjić, Todorović, & Gilchrist, 2010; Ripamonti et al., 2004). These effects are often interpreted as resulting from an effect of geometry on the (perhaps implicitly) perceived illumination (see, e.g., Brainard & Maloney, 2011). A recent result in this tradition, however, suggests that when strong photometric cues are available, they can dominate geometric information (Gilchrist & Radonjić, 2010).

In related work, a number of lightness illusions (Adelson, 1993, 2000; Anderson & Winawer, 2005; Todorović, 1997) also implicate a key role for geometric cues in the perception of surface lightness. In much of this work, the emphasis has been on understanding how the luminance relationships across junctions support particular scene interpretations. Similar themes are found in the literature on transparency (Anderson, 1997; Anderson & Winawer, 2005; Beck, Prazdny, & Ivry, 1984; Metelli, 1985; Singh & Anderson, 2002).

Our results do not contradict the conclusion that geometric factors play a role in the perception of scene illumination. They do, however, emphasize the need to understand in detail how information carried by photometric and geometric cues interact.

Formal connections

The formal structure of our task is that the observers had to identify which of the two images contained patches drawn from a mixture of two luminance probability

distributions and which contained luminance patches drawn from a single distribution. At this formal level, our task is thus closely related to work on the perception of texture (for a review, see Landy & Graham, 2004), where textures are defined in terms of the statistical properties of their luminance distributions. Despite the formal similarity, however, there are important content differences between most texture experiments and our experiments. For example, texture work often holds the mean luminance and variance (contrast) constant across distributional manipulations, so as to allow investigation of the structure carried by higher order statistical regularities. In addition, these textures are typically generated using small, spatially contiguous micropatterns rather than the more macroscopic spatial structure of interest when one considers illumination discrimination.

Our task also shares formal features with a contour integration task introduced by Field, Hayes, and Hess (1993), where observers were asked to detect the presence of a coherently oriented contour of Gabor patches embedded in a field of randomly oriented Gabor patches.

Future directions

Our work represents an initial foray toward understanding how photometric information enables illuminant segregation. To make progress, we employed simple stimuli and studied performance using a simple psychophysical task. These simplifications allowed us to observe a number of clear regularities within the laboratory model we studied. Nonetheless, it is worth keeping in mind some of the limitations of this model. First, the test checkerboards did not produce a strong perceptual sense of a collection of surfaces seen under two illuminants and, thus, may not have engaged all of the mechanisms that normally subserve illuminant segregation. Second, our use of a two-alternative forced-choice procedure simplified the task demands, relative to the case of viewing a single image that might contain multiple regions of illumination. Third, the only task-relevant information in our stimuli was photometric and this may have weakened any potential geometric effects. Expanding the research to richer stimuli and tasks, so as to overcome these limitations, is a clear direction for future research.

Appendix A

Ideal observer calculations

An ideal observer was developed for each condition in [Experiment 1](#) and its performance was characterized

through simulation. For each condition, 100 trials each of 38 multiples of the test distribution parameters were simulated, and for each simulated trial, the ideal observer calculation indicated which checkerboard contained the test patches, based on the luminance of all 50 checkerboard patches. Each condition was repeated three times in a run of the simulation; the results were analyzed in the same way as the human data. For each simulation, the mean threshold from the three repetitions was taken as the measure of performance.

In the location-known conditions, means from ten simulations were averaged together for each value of test patch number. This was also done for the 1-, 2-, and 3-test-patch conditions for the location-unknown conditions. The computations for the 4- and 5-test-patch location-unknown conditions were lengthy, and for these, only a single simulation mean (of three repetitions) was obtained.

The ideal observer's choice was based on the log likelihood ratio:

$$\ell(\mathbf{x}) = \log\left(\frac{p(\mathbf{x}|\text{TestOnLeft})}{p(\mathbf{x}|\text{TestOnRight})}\right), \quad (\text{A1})$$

where the vector \mathbf{x} represents the luminances of the 50 checkerboard patches presented on a particular trial. If $\ell(\mathbf{x})$ was greater than 0, the ideal observer indicated that the test was on the left; if $\ell(\mathbf{x})$ was less than or equal to 0, the ideal observer indicated that the test was on the right. For a given trial, the vector \mathbf{x} can be thought of as the concatenation of the vectors \mathbf{x}^{Left} , representing the luminances of the 25 checkerboard patches on the left checkerboard, and $\mathbf{x}^{\text{Right}}$, the luminances of the 25 checkerboard patches on the right checkerboard.

Location-known condition

For the location-known conditions, the log likelihood of the data given that the test was on the left is

$$\log(p(\mathbf{x}|\text{TestOnLeft})) = \sum_{i \in \{t\}} \log(p_t(x_i^{\text{Left}})) + \sum_{i \in \{s\}} \log(p_s(x_i^{\text{Left}})) + \sum_{i=1}^{25} \log(p_s(x_i^{\text{Right}})). \quad (\text{A2})$$

In this expression, i indexes patch location within a single checkerboard (left or right), $\{t\}$ represents the indices of the test patches within the test checkerboard, and $\{s\}$ represents the remaining indices within the test checkerboard. The probability $p_t(x)$ is the probability of observing luminance x at a single patch under the test distribution, and $p_s(x)$ is the probability of observing luminance x at a

single patch under the standard distribution. The corresponding expression when the test is on the right is

$$\log(p(\mathbf{x}|\text{TestOnRight})) = \sum_{i \in \{t\}} \log(p_t(x_i^{\text{Right}})) + \sum_{i \in \{s\}} \log(p_s(x_i^{\text{Right}})) + \sum_{i=1}^{25} \log(p_s(x_i^{\text{Left}})). \quad (\text{A3})$$

The probability $p_s(x)$ was evaluated using the probability density function of a truncated Gaussian distribution with mean $\mu = 15 \text{ cd/m}^2$ and standard deviation $\sigma = 5 \text{ cd/m}^2$ (the parameters of the luminance distribution for stimulus patches under the standard illuminant). The Gaussian density function was truncated between the range $[5, 50]$ and renormalized so that the total probability was 1, to match how the stimuli were generated for the experiments.

For the case where the parameters of the test distribution are known to the observer, the probability $p_t(x)$ would be evaluated using the same basic method as described for the standard distribution but with the truncated Gaussian having a mean, variance, and truncation range computed for the test distribution.

Because the experiments were run using a staircase method, there was uncertainty about the parameters of the test distribution. To model this uncertainty, in a separate simulation, $p_t(x)$ was evaluated as a weighted sum of the likelihood for a given set of test distribution parameters, with the weights given by the probability of those parameters. The distribution of the parameters, $p(\text{TestDistParam})$, was estimated by creating a histogram of the parameters used in the observers' experimental runs corresponding to the condition being simulated. With this,

$$p_t(x) = \int_{\text{TestDistParam} \mu=15, \sigma=5}^{\text{TestDistParam} \mu=37.5, \sigma=12.5} p_t(x|\text{TestDistParam}) p(\text{TestDistParam}). \quad (\text{A4})$$

We found that adding this uncertainty had little effect on the predictions and, in the interest of computational efficiency, ran our main simulations with the test distribution parameters known.

Location-unknown condition

Before providing the general equations for the location-unknown conditions, we first develop the ideas for a

simplified example. Suppose that there are only four patches in each checkerboard and that there are two signal patches to be detected. Let \mathbf{x} be the vector of eight luminances concatenated from the vectors \mathbf{x}^{Left} , the luminances of patches $x_1^{\text{Left}}, \dots, x_4^{\text{Left}}$ belonging to the left checkerboard, and $\mathbf{x}^{\text{Right}}$, the luminances of the four patches $x_1^{\text{Right}}, \dots, x_4^{\text{Right}}$ belonging to the right checkerboard.

Suppose that the test is on the left. In this case, there are $\binom{4}{2} = 6$ possible combinations of the standard illuminant and test illuminant patch locations in the left checkerboard, each equally likely. The log likelihood of the observed luminance vector can be written as

$$\log(p(\mathbf{x}|\text{TestOnLeft})) = \log(p(\mathbf{x}^{\text{Left}}|\text{TestOnLeft})) + \log(p(\mathbf{x}^{\text{Right}}|\text{TestOnLeft})). \quad (\text{A5})$$

Here,

$$\begin{aligned} p(\mathbf{x}^{\text{Left}}|\text{TestOnLeft}) &= (1/6)p_t(x_1^{\text{Left}})p_t(x_2^{\text{Left}})p_s(x_3^{\text{Left}})p_s(x_4^{\text{Left}}) \\ &+ (1/6)p_t(x_1^{\text{Left}})p_s(x_2^{\text{Left}})p_t(x_3^{\text{Left}})p_s(x_4^{\text{Left}}) \\ &+ (1/6)p_t(x_1^{\text{Left}})p_s(x_2^{\text{Left}})p_s(x_3^{\text{Left}})p_t(x_4^{\text{Left}}) \\ &+ (1/6)p_s(x_1^{\text{Left}})p_t(x_2^{\text{Left}})p_t(x_3^{\text{Left}})p_s(x_4^{\text{Left}}) \\ &+ (1/6)p_s(x_1^{\text{Left}})p_t(x_2^{\text{Left}})p_s(x_3^{\text{Left}})p_t(x_4^{\text{Left}}) \\ &+ (1/6)p_s(x_1^{\text{Left}})p_s(x_2^{\text{Left}})p_t(x_3^{\text{Left}})p_t(x_4^{\text{Left}}), \end{aligned} \quad (\text{A6})$$

is the weighted sum of the likelihoods for all six equally likely possible combinations of standard and test distribution locations. The expression

$$p(\mathbf{x}^{\text{Right}}|\text{TestOnLeft}) = p_s(x_1^{\text{Right}})p_s(x_2^{\text{Right}})p_s(x_3^{\text{Right}})p_s(x_4^{\text{Right}}), \quad (\text{A7})$$

provides the likelihood of the observed luminances under the standard distribution.

For notational simplicity, the combinations of locations in each term on the right-hand side of Equation A6 can be represented as a set of combinations of two t and two s characters, where t in the i th location indicates that the i th patch on the left is drawn from the test distribution, $p_t(x_i^{\text{Left}})$, and s indicates that the i th patch on the left is drawn from the standard distribution, $p_s(x_i^{\text{Left}})$. Let C be a matrix of the set of combinations of t and s in Equation A6 whose rows are C_a for the six combinations $a = 1, 2, \dots, 6$.

Within a given row C_a , the columns are indexed by $i = 1, 2, 3, 4$. We write

$$C = \begin{Bmatrix} t & t & s & s \\ t & s & t & s \\ t & s & s & t \\ s & t & t & s \\ s & t & s & t \\ s & s & t & t \end{Bmatrix}, \quad (\text{A8})$$

where the individual entries may be denoted by C_{ai} . Equation A6 can then be represented as

$$p(\mathbf{x}^{\text{Left}}|\text{TestOnLeft}) = (1/6) \sum_{C_a \in C} \prod_{i=1}^4 p_{C_{ai}}(x_i^{\text{Left}}). \quad (\text{A9})$$

Combining Equations A7 and A9 into Equation A5, we obtain

$$\begin{aligned} \log(p(\mathbf{x}|\text{TestOnLeft})) &= \log(1/6) + \log\left(\sum_{C_a \in C} \prod_{i=1}^4 p_{C_{ai}}(x_i^{\text{Left}})\right) \\ &+ \sum_{i=1}^4 \log(p_s(x_i^{\text{Right}})). \end{aligned} \quad (\text{A10})$$

We can generalize Equation A10 to the case of k test distribution patches displayed at n possible patch locations. In this case, there are $N = \binom{n}{k}$ possible test patch arrangements and we obtain

$$\begin{aligned} \log(p(\mathbf{x}|\text{TestOnLeft})) &= \log(1/N) + \log\left(\sum_{C_a \in C} \prod_{i=1}^n p_{C_{ai}}(x_i^{\text{Left}})\right) \\ &+ \sum_{i=1}^n \log(p_s(x_i^{\text{Right}})). \end{aligned} \quad (\text{A11})$$

A similar equation can be written for $\log(p(\mathbf{x}|\text{TestOnRight}))$. The log likelihoods are then compared as in Equation A1.

In a separate simulation, Equation A11 was modified by using Equation A4 for every instance of $p_t(x)$ to model observer uncertainty in the parameters of the test distribution. The weights for signal level were again taken from observers' empirical test distribution parameter probability histograms from the corresponding experimental

condition. Because of computational limitations, this simulation was done only for the 1-, 2-, and 3-test-patch cases. As with the corresponding simulations for the location-known case, we found little effect of test level uncertainty and we report results for the case where there was no uncertainty about test illuminant level.

Acknowledgments

This work was supported by NIH RO1 EY10016, NIH P30 EY001583, NIH T32 EY007035, and NIH T90 DA022763. We thank C. Broussard for technical assistance and two anonymous reviewers for commenting on an earlier version of this manuscript. A preliminary version of this work was presented as a poster at the 2010 Optical Society of America Fall Vision Meeting.

Commercial relationships: none.

Corresponding author: Thomas Y. Lee.

Email: thomyle@sas.upenn.edu.

Address: Department of Psychology, University of Pennsylvania, 3401 Walnut St., Ste. 302C, Philadelphia, Pennsylvania, USA.

Footnote

¹Computation of visual angle does not take into account off-axis effects. Order of all dimensions specified is vertical first and then horizontal.

References

- Abbey, C. K., & Eckstein, M. P. (2006). Classification images for detection, contrast discrimination, and identification tasks with a common ideal observer. *Journal of Vision*, 6(4):4, 335–355, <http://www.journalofvision.org/content/6/4/4>, doi:10.1167/6.4.4. [PubMed] [Article]
- Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, 262, 2042–2044.
- Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 339–351). Cambridge, MA: MIT Press.
- Ahumada, A. J., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, 49, 1751–1756.
- Alexander, J. M., & Lutfi, R. A. (2004). Informational masking in hearing-impaired and normal-hearing listeners: Sensation level and decision weights. *Journal of the Acoustical Society of America*, 116, 2234–2247.
- Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of junctions. *Perception*, 26, 419–453.
- Anderson, B. L., & Winawer, J. (2005). Image segmentation and lightness perception. *Nature*, 434, 79–83.
- Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research*, 27, 1915–1924.
- Barlow, H. B. (1980). The absolute efficiency of perceptual decisions. *Philosophical Transactions of the Royal Society of London B*, 290, 71–82.
- Beck, J. (1959). Stimulus correlates for the judged illumination of a surface. *Journal of Experimental Psychology*, 58, 267–274.
- Beck, J. (1961). Judgments of surface illumination and lightness. *Journal of Experimental Psychology*, 61, 368–373.
- Beck, J., Prazdny, K., & Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception & Psychophysics*, 35, 407–422.
- Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision*, 3(8):2, 541–553, <http://www.journalofvision.org/content/3/8/2>, doi:10.1167/3.8.2. [PubMed] [Article]
- Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illumination models. *Journal of Vision*, 11(5):1, 1–18, <http://www.journalofvision.org/content/11/5/1>, doi:10.1167/11.5.1. [PubMed] [Article]
- Craven, B. J., & Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research*, 32, 1359–1366.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system—Evidence for a local association field. *Vision Research*, 33, 173–193.
- Gerhard, H. E., & Maloney, L. T. (2010a). Detection of light transformations and concomitant changes in surface albedo. *Journal of Vision*, 10(9):1, 1–14, <http://www.journalofvision.org/content/10/9/1>, doi:10.1167/10.9.1. [PubMed] [Article]
- Gerhard, H. E., & Maloney, L. T. (2010b). Estimating changes in lighting direction in binocularly viewed three-dimensional scenes. *Journal of Vision*, 10(9):14, 1–22, <http://www.journalofvision.org/content/10/9/14>, doi:10.1167/10.9.14. [PubMed] [Article]
- Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., et al. (1999). An anchoring

- theory of lightness perception. *Psychological Review*, 106, 795–834.
- Gilchrist, A. L. (1977). Perceived lightness depends on perceived spatial arrangement. *Science*, 195, 185.
- Gilchrist, A. L. (1980). When does perceived lightness depend on perceived spatial arrangement? *Perception & Psychophysics*, 28, 527–538.
- Gilchrist, A. L., & Radonjić, A. (2010). Functional frameworks of illumination revealed by probe disk technique. *Journal of Vision*, 10(5):6, 1–12, <http://www.journalofvision.org/content/10/5/6>, doi:10.1167/10.5.6. [PubMed] [Article]
- Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology*, 10, 663–666.
- Hochberg, J. E., & Beck, J. (1954). Apparent spatial arrangement and perceived brightness. *Journal of Experimental Psychology*, 47, 263–266.
- Ishihara, S. (1998). *Ishihara's tests for colour deficiency*. Tokyo: Kanehara.
- Kardos, L. (1934). Ding und Schatten: Eine experimentelle Untersuchung über die Grundlagen des Fabensehens. *Zeitschrift für Psychologie*, 23, 1–184.
- Khang, B., Koenderink, J. J., & Kappers, A. M. L. (2006). Perception of illumination direction in images of 3-D convex objects: Influence of surface materials and light fields. *Perception*, 35, 624–645.
- Knill, D. C., & Kersten, D. (1991). Apparent surface curvature affects lightness perception. *Nature*, 351, 228–230.
- Koenderink, J. J., Pont, S. C., van Doorn, A. J., Kappers, A. M., & Todd, J. T. (2007). The visual light field. *Perception*, 36, 1595–1610.
- Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace and Company.
- Kozaki, A., & Noguchi, K. (1976). The relationship between perceived surface-lightness and perceived illumination. *Psychological Research*, 39, 1–16.
- Kurki, I., Peromaa, T., Hyvärinen, J., & Saarinen, J. (2009). Visual features underlying perceived brightness as revealed by classification images. *PLoS One*, 4, 1–8.
- Landy, M. S., & Graham, N. (2004). Visual perception of texture. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (pp. 1106–1118). Cambridge, MA: MIT Press.
- Levitt, H. (1971). Transformed up–down methods in psychoacoustics. *Journal of the Acoustic Society of America*, 49, 467–477.
- Maloney, L. T., & Brainard, D. H. (2010). Color and material perception: Achievements and challenges. *Journal of Vision*, 10(9):19, 1–6, <http://www.journalofvision.org/content/10/9/19>, doi:10.1167/10.9.19. [PubMed] [Article]
- Metelli, F. (1985). Stimulation and perception of transparency. *Psychological Research*, 47, 185–202.
- Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision*, 2(1):6, 79–104, <http://www.journalofvision.org/content/2/1/6>, doi:10.1167/2.1.6. [PubMed] [Article]
- Nascimento, S. M. C., & Foster, D. H. (2000). Relational color constancy in achromatic and isoluminant images. *Journal of the Optical Society of America*, 17, 225–231.
- Noguchi, K., & Kozaki, A. (1985). Perceptual scission of surface lightness and illumination: An examination of the Gelb effect. *Psychological Research*, 47, 19–25.
- Noguchi, K., & Masuda, N. (1971). Brightness changes in a complex field with changing illumination: A re-examination of Jameson and Hurvich's study of brightness constancy. *Japanese Psychological Research*, 13, 60–69.
- Oyama, T. (1968). Stimulus determinants of brightness constancy and the perception of illumination. *Japanese Psychological Research*, 10, 146–155.
- Radonjić, A., Todorović, D., & Gilchrist, A. L. (2010). Adjacency and surroundedness in the depth effect on lightness. *Journal of Vision*, 10(9):12, 1–16, <http://www.journalofvision.org/content/10/9/12>, doi:10.1167/10.9.12. [PubMed] [Article]
- Reid, C. R., Jr., & Shapley, R. (1988). Brightness induction by local contrast and the spatial dependence of assimilation. *Vision Research*, 28, 115–132.
- Ripamonti, C., Bloj, M., Hauck, R., Mitha, K., Greenwald, S., Maloney, S. I., et al. (2004). Measurements of the effect of surface slant on perceived lightness. *Journal of Vision*, 4(9):7, 747–763, <http://www.journalofvision.org/content/4/9/7>, doi:10.1167/4.9.7. [PubMed] [Article]
- Rutherford, M. D., & Brainard, D. H. (2002). Lightness constancy: A direct test of the illumination estimation hypothesis. *Psychological Science*, 13, 142–149.
- Shimozaki, S. S., Eckstein, M. P., & Abbey, C. K. (2005). Spatial profiles of local and nonlocal effects upon contrast detection/discrimination from classification images. *Journal of Vision*, 5(1):5, 45–57, <http://www.journalofvision.org/content/5/1/5>, doi:10.1167/5.1.5. [PubMed] [Article]
- Singh, M., & Anderson, B. L. (2002). Toward a perceptual theory of transparency. *Psychological Review*, 109, 492–519.
- Sinha, P., & Adelson, E. H. (1993). Recovering reflectance and illumination in a world of painted polyhedra.

- Proceedings of the 4th International Conference on Computer Vision*, 156–163.
- Spehar, B., Debonet, J. S., & Zaidi, Q. (1996). Brightness induction from uniform and complex surrounds: A general model. *Vision Research*, *36*, 1893–1906.
- Todorović, D. (1997). Lightness and junctions. *Perception*, *26*, 379–394.
- Wichmann, F. W., & Hill, N. J. (2001). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, *63*, 1314–1329.
- Wyszecki, G., & Stiles, W. S. (1982). *Color science—Concepts and methods, quantitative data and formulae* (2nd ed., pp. 60–63). New York: John Wiley & Sons.