The proximity structure of achromatic surface colors and the impossibility of asymmetric lightness matching

ALEXANDER D. LOGVINENKO
Glasgow Caledonian University, Glasgow, Scotland

and

LAURENCE T. MALONEY
New York University, New York, New York

In asymmetric lightness matching tasks, observers sometimes report that they cannot achieve satisfactory matches between achromatic surfaces under different neutral illuminants. The surfaces appear different, yet no further adjustment of either surface improves the match. There are evident difficulties in interpreting data from a task that the observer cannot always do, and these difficulties likely affect the interpretation of a large number of previous studies. We investigated, as an alternative to asymmetric matching, the direct use of proximity judgments in the study of surface lightness perception. We asked observers to rate the perceived dissimilarity of pairs of achromatic surfaces that were placed in identical scenes and viewed under different neutral illuminants. We develop a parametric model that accurately predicts perceived dissimilarity in terms of physical light intensities and surface albedos. The parameters of this model are readily interpretable. In particular, the ratio of the influence of changes in illuminant intensity and changes in surface albedo is a measure of the extent to which the observer *discounts the illuminant*. Asymmetric lightness matching can be interpreted as an unachievable limiting case of proximity judgment.

When scene lighting changes in intensity, objects that appear white tend to remain white in appearance, black objects remain black, and other achromatic surfaces stay roughly the same shades of gray. Although the intensity of light reflected from an object is proportional to both the intensity of incident light and the object's reflectivity (albedo), the resulting perceptual estimates of surface appearance depend primarily on surface albedo. Explaining how the human visual system achieves this *lightness constancy* is a major challenge for visual science (Brainard, 2003; Gilchrist et al., 1999; Hurlbert, 1998; Maloney, 1999).

The principal method used to study lightness and lightness constancy is asymmetric matching. In an asymmetric matching task, the observer is instructed to match the lightness of two objects under different illumination by, in effect, changing the albedo of one of them. If the observer consistently chooses identical albedos under different illuminants, he or she is judged to be *lightness constant*.

Although asymmetric matching is widely employed to study both lightness and color perception, the method has

a major shortcoming. When observers finally set a match, they often report that the chosen chip is not, in fact, identical in appearance to the target, as the following comment taken from an asymmetric color matching study indicates:

> At this match point, however, the test and the match surfaces looked different, and the observers felt as if further adjustments of the match surface should produce a better correspondence. Yet turning any of the knobs or combinations of knobs only increased the perceptual difference. (Brainard, Brunt, & Speigle, 1997, p. 2098)

In a recent article, Foster (2003) highlights this problem with color matching: Alleged color matches are not always perceptually identical. Our own observers confirm the observations of Foster and of Brainard and colleagues, but in the context of lightness matching. They report that a lightness match is not, in general, achievable when two surfaces are illuminated by different illuminants. Intuitively, a white surface in shadow often does not have the same appearance as a white surface under bright illumination, and no adjustment of the albedo of either surface can produce a perceptual match. So far as we can determine, Katz (1935, pp. 79ff) was the first to report that, when observers make a match in a lightness or color constancy experiment, there is usually a residual difference.

In this article, we recast the asymmetric matching paradigm in terms of the *apparent dissimilarity* of surfaces under different illuminants. We asked observers not to set matches between pairs of surfaces under different illumi-

nants but to rate the dissimilarity of each pair. In doing so, we retain much of the power of asymmetric matching as a method for investigating surface appearance, while eliminating its major drawback. We note that the full range of dissimilarity judgments between all pairs of surfaces under two illuminants can potentially tell us more about lightness perception than can match judgments alone. Particularly, it can shed light on how light and surface properties are represented in biological vision (see Mausfeld, 1998) and what it is that observers actually do when instructed to make asymmetric lightness matches. An evident hypothesis is that observers are engaged in partial matching (as proposed by Wuerger, Maloney, & Krauskopf, 1995; see also Maloney, Wuerger, & Krauskopf, 1995) and that the lightness *match* to a surface under one illuminant is just the *least dissimilar* surface under a second illumination.

We present an experiment in which we first employed standard nonmetric multidimensional scaling (MDS) techniques (Cox & Cox, 2001) to reconstruct a crude spatial representation that parsimoniously summarized the dissimilarity judgments for each observer. Although the resulting representation is highly structured, traditional MDS techniques do not lend themselves to testing explicit hypotheses concerning this structure. Accordingly, we have also developed a parametric model and statistical methods based on previous research in lightness and specifically tailored to the problem at hand. These methods, which we refer to as *maximum likelihood parametric scaling* (MLPS), are described only briefly in the present article (a description of a similar parametric approach to difference scaling can be found in Maloney & Yang, 2003). They allow us to test specific hypotheses concerning lightness perception.

EXPERIMENT 1
Dissimilarity Judgments

Method

Observers

Three experienced observers participated in the experiment. All were unaware of its purpose.

Stimuli

On each trial, the observer saw two identical sets of seven Munsell papers (albedos reported in Table 1) on a random-dot background (60×50 cm), arranged as in Figure 1. We refer to these papers as *chips*. Two stage lamps were used to illuminate each of the two halves of the display independently. The display was 1.7 m in front of the observer. The stage lamps were located slightly behind and to the side of the observer. The intensity of illumination of the Munsell chips was set at one of three levels: 175, 488, or 1,694 lux. Note that the logarithms of the three light intensities are almost exactly uniformly spaced, as are the logarithms of the seven surface albedos (Table 1).

Task

The observer's task was to evaluate the dissimilarity between a designated pair of chips (one in each half), using a 30-point scale. The observers were first instructed in how to use the scale values. They were first shown the display with the two sides of the display illuminated with equal intensity. The experimenter indicated pairs of chips, one on the left side of the display and one on the right, that were physically identical and instructed the subject to assign

Table 1
Luminances (cd/m²) of the Test Objects

	Albedo						
	1	2	3	4	5	6	7
Illuminant	7.7%	10.4%	15.6%	24.6%	36.2%	54.8%	78.7%
1 (175 lux)	4.9	5.6	8.7	13.2	21.6	31.8	46.5
2 (488 lux)	14.6	20.8	25.0	40.4	61.9	92.3	133.3
3 (1,694 lux)	52.2	68.9	89.5	129.0	196.6	283.3	416.5

the smallest scale rank (1) to such pairs. They were then shown a reference display that consisted of a bipartite field containing a copy of the chip with the lowest albedo (7.7%) illuminated by the least intense light and the chip with the highest albedo (79%) illuminated by the most intense light. The observer was told that this extreme difference corresponded to the largest value (30) on the 30-point scale. This reference display was always visible, above the stimuli, during the course of the experiment. It was intended to aid the observers in maintaining a constant criterion for selecting scale values throughout the whole experiment.

There were 6 possible pairings of three illuminant intensities ignoring order (an illuminant could be paired with itself). There were 49 possible pairings of seven surfaces for each light pairing. Each judgment was repeated three times. Each observer completed 882 ($6 \times 49 \times 3$) trials.

Results and Analysis

A matrix of dissimilarities, averaged across observers and presentations,[1] was used as an input into a nonmetric MDS algorithm. The results are presented in Figure 2A.

A particular Munsell chip illuminated by a particular light will be referred as to a *test object*. A two-digit acronym is employed for each test object. The first digit denotes albedo (α_i, $i = 1, 2, \ldots, 7$), and the second
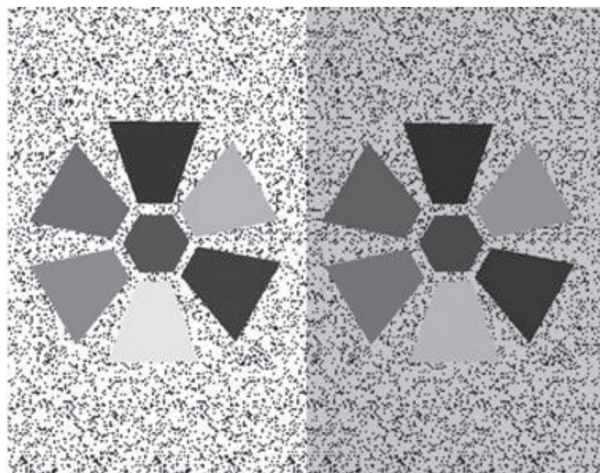

Figure 1. Example of the stimulus displays used in the first experiment. The left and right halves of the display are illuminated differently. Each half contained several surface chips with different albedos, arranged as shown. The positions were randomized from trial to trial. The same albedos are presented on the left as on the right. The observer's task was to rate the dissimilarity of two chips designated by a laser pointer, one on the left and one on the right.
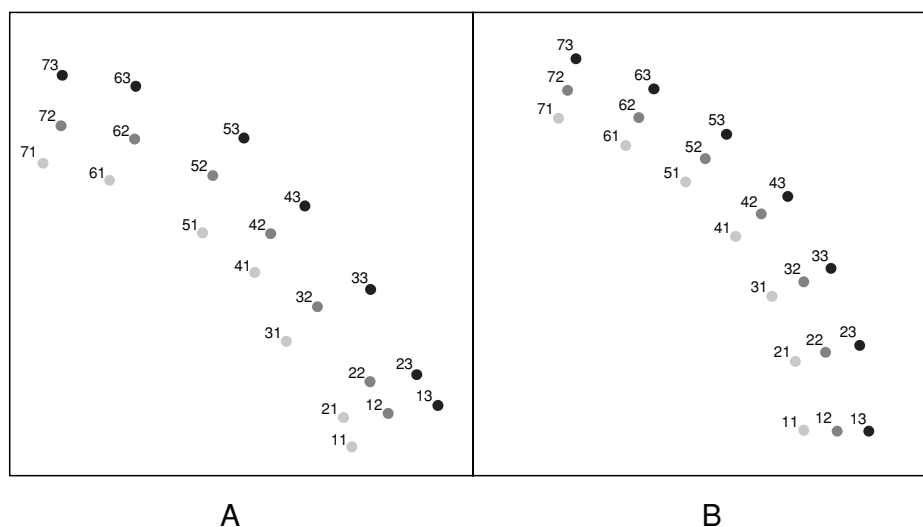
A   B

Figure 2. Results of the first experiment. (A) The dissimilarities between gray chips, such as those illustrated in Figure 1, were subjected to two-dimensional nonmetric multidimensional scaling (NMDS). Each point *ik* represents the light and surface pair $\langle a_i, L_k \rangle$. The figure shows the final configuration, the stress for which was 0.06, indicating a good fit. In the NMDS model, the ordering of dissimilarities corresponds to the ordinary Euclidean distance between pairs of points. (B) A schematization of the maximum likelihood parametric scaling model that best fit the observers' data. The points lie along the radii of concentric circles. The light–surface pair $\langle a_i, L_k \rangle$ falls on the *k*th equi-illuminant circle and the *i*th equi-albedo radius. The metric is not the usual Euclidean metric. Distance is measured by the shortest path along radii and circles from a point *ik* to any other point *jl*, as illustrated.

stands for illumination level (L_k, $k = 1, 2, 3$) as defined in Table 1. Note that the albedos and illumination intensities are ordered so that $\alpha_1 < \alpha_2 < \ldots < \alpha_7$ and $L_1 < L_2 < L_3$, for convenience. In Figure 2A, the 21 test objects (7 reflectances $\times$ 3 illuminations) are systematically arranged in a two-dimensional Euclidean plane, and the Euclidean distances between them are, generally, in the same order as averaged dissimilarities. The stress of the configuration in Figure 2A (i.e., an index showing how well the distances match the dissimilarities) is 0.06. The stress of the solution in three dimensions is negligibly smaller (0.05), whereas the optimal solution in one dimension had a very high stress (exceeding 0.50). By ordinary criteria for MDS, the dissimilarity judgments are two-dimensional. The ordering of pairwise dissimilarity judgments can be reproduced as distances between points in two-dimensional space, but not in one dimension. Note that any MDS solution is equivalent to any other that is a rotation, reflection, or scaling of the first.

It must be mentioned that the data unequivocally rule out luminance as a possible determinant of dissimilarity. Indeed, Objects 71 and 13, with nearly the same luminance (46.5 and 52.2 cd/m², respectively), make a pair with one of the largest dissimilarities. Notice also that Object 73 (of the highest luminance: 416.5 cd/m²) is much closer to Object 71 (46.5 cd/m²) than to Object 13 (52.2 cd/m²). Therefore, the log luminance difference produced by change in illumination intensity weighed considerably less than the same log luminance difference produced by change in reflectance, when contributing to dissimilarity

judgments. This is in line with Kardos's (1934) finding that observers do not pay much attention to shadows when describing real scenes. We refer to this asymmetry as *discounting of illumination*.[2]

The points corresponding to test objects in Figure 2A are arranged in a two-dimensional fanlike pattern. It can be described either as three arcs (more or less parallel to each other) or as seven converging radii. Objects of the same reflectance lie roughly along the same radius (equi-albedo loci), and objects having the same illumination lie along the same arc (equi-illuminant loci). All of the three objects lying along the same radius are ordered with respect to their illumination (the lower the object's illumination, the closer it is to the origin). All the radii are ordered with respect to reflectance, the leftmost radius representing the highest reflectance.

Inspection of Figure 2A leads to several questions. Is a Euclidean metric an accurate model of how light and surface differences are combined? Is the apparent curvature of the three arcs real or an artifact of the fitting procedure? MDS is primarily a method for exploratory data analysis, and although we can use it to formulate hypotheses about dissimilarity perception, in itself it does not allow us to confirm or reject these hypotheses. Therefore, we developed explicit parametric models of dissimilarity perception and fit them to our data. These models were arranged hierarchically, and models higher in the hierarchy had more free parameters. Within this nested hierarchy, we sought the model of minimum complexity that accounted for our data. Due to limitations of space, we will present

an abbreviated subset of the models considered and their corresponding hypotheses.

Testing Parametric Models of Dissimilarity Judgments

For two illumination–albedo pairs ik and jm, each model was intended to predict the observer's mean rated dissimilarity $\delta(ik, jm)$ for that pair. The observer's dissimilarity rating on each trial was modeled as a random variable

$$\Delta(ik, jm) = \delta(ik, jm) + \varepsilon, \qquad (1)$$

where ε was a Gaussian random variable with a mean of 0 and a variance σ^2. The models embodied a series of hypotheses about the observer's judgments of dissimilarity. We fit each model to the observers' data by the method of maximum likelihood and compared the goodness of fit of each model by nested hypothesis tests.[3] The model at the "top" of the hierarchy has six parameters (w_s, w_l, c, d, r, σ^2) and is of the form of Equation 2 below.

We will postpone a detailed discussion of $\Gamma(k, m \mid c, w_l)$ until after we define the remaining terms in Equation 2. For now, we note that $\Gamma(k, m \mid c, w_l)$ controls the increase in spacing between the radii as we move to arcs corresponding to more intense illuminants.

The basic form of the model is a Minkowski r metric combination of differences in albedo and in light intensity. The MDS solution shown in Figure 2A is based on the assumption that $r = 2$ (the Euclidean metric). We will test this hypothesis and also consider the hypothesis that $r = 1$ (the *city block* metric). If $r = 1$, dissimilarity could be expressed as the weighted sum of the log-differences in illumination intensity and surface intensity [scaled by the term $\Gamma(k, m \mid c, w_l)$ which is effectively an illumination–albedo interaction].

The parameters w_s and w_l control the weights given to the log surface albedo difference and the log light difference. We interpret the ratio $\kappa = w_l / w_s$ as an index of illumination discounting. It is the relative impact of a change in luminance induced by changing light intensity and the same change induced by changing surface albedo. When κ is 1, a proportional change in surface albedo has the same effect as a proportional change in illumination intensity of the same magnitude (no illumination discounting). When κ is 0, changes in illumination have no effect on perceived dissimilarity (full illuminant discounting). The quantity $1 - \kappa$ is a candidate index of lightness constancy.

There are several commonly used indices of lightness or color constancy in the literature—notably, the Brunswik (1929) ratio and the Thouless (1931) ratio. The relationship between these different indices of constancy (including ours) is a fundamental issue, but not one that we can resolve here.

In Figure 2A, the circular arcs are visibly curved, and as a consequence, the spacing between adjacent radii increases as we move to arcs corresponding to more intense

illumination. If this change in spacing within arcs is real, it represents a form of light–surface interaction in the perception of dissimilarity of light–surface pairs. The same pair of surfaces is seen as more dissimilar under a more intense illumination. The curvature of a circular arc is the inverse of the radius of the circle. If we let ρ_k denote the radius of the arc corresponding to the kth illuminant, L_k, its curvature is $c_k = 1/\rho_k$. We will assume that the kth radius is just $\rho_k = \rho_1 + (k-1)w_l$. That is, the successive arcs are spaced at intervals of width w_l. If we know ρ_1, the radius of the innermost arc, the spacing of all of the arcs can be computed in terms of the previously defined parameter, w_l. We introduce a parameter, $c = 1/\rho_1$, the curvature of the innermost arc. It is convenient to work with the curvature parameter c rather than with the radius ρ_1, because the hypothesis of *no curvature* is then simply $c = 0$.

We define the interaction term in Equation 2 to be

$$\Gamma(k, m \mid c, w_l) = \frac{\rho_1 + (\min\{k, m\} - 1)w_l}{\rho_1}, \qquad (3)$$

which is the ratio between the radius of arc k or arc m (whichever corresponds to the dimmer illuminant) and the radius of the innermost arc. Equation 3 reduces to

$$\Gamma(k, m \mid c, w_l) = 1 + c(\min\{k, m\} - 1)w_l. \qquad (4)$$

(It will become clear in the sequel why the minimum of k and m is used.) When $c = 0$, the term $\Gamma(k, m \mid c, w_l) = 1$ and the spacing on successive arcs is the same, independently of light intensity. The light–surface pairs form a Cartesian grid. We seek to test the hypothesis $c = 0$ (no curvature) against the alternative $c > 0$ (curvature).

City block versus Euclidean metric. We first tested whether we could reject the hypothesis that $r = 2$ (Euclidean metric) versus the hypothesis that $r \neq 2$. We rejected this hypothesis [$\chi^2(1) = 111.99$, $p < .0001$] and found that the maximum likelihood estimate of r was 0.98. We could not reject the hypothesis that $r = 1$ [$\chi^2(1) = 0.0822$, $p = .774$] and adopted that value for convenience. Note that the nonmetric MDS model in Figure 2A assumes the Euclidean metric that we have just rejected.

We can interpret this result in terms of Figure 2A. Suppose that the observer is comparing ik and jm with $k < m$ (the light L_m is more intense than the light L_k). Then, the dissimilarity is the sum of a radial distance (log light intensity difference) and an arc distance (log surface albedo difference). When $c > 0$, the arc distance along the kth arc is smaller than the arc distance along the mth arc. In using $\min\{k, m\}$ in Equation 4, we removed this ambiguity by choosing to use the "shorter" arc length in computing dissimilarity.

Dissimilarity scale use. The hypothesis that $d = 1$ was similarly tested and rejected [$\chi^2(1) = 14.388$, $p < .0001$]. The maximum likelihood estimate was $d = 1.10$, suggesting that the observers used the rating scale in a slightly nonlinear manner.

$$\delta(ik, jm) = \left[w_s \Gamma(k, m \mid c, w_l) \left| \log \alpha_i - \log \alpha_j \right|^r + w_l \left| \log L_k - \log L_m \right|^r \right]^d \qquad (2)$$

Light–surface interaction. We next tested the hypothesis that $c = 0$. If this hypothesis were true, the curvature observed in Figure 2A (which represents an expansion in perceived dissimilarities of surfaces with increasing light intensities) would have proved to be just the artifact of nonmetric MDS known as Kendall's horseshoe: Solutions that are really one-dimensional tend to "curve" into two-dimensional solutions as a consequence of measurement error (Kendall, 1971). We could, however, reject the hypothesis that $c = 0$ [$\chi^2(1) = 113.56$, $p < .0001$]. We conclude that the apparent dissimilarity between surfaces did increase with increasing light intensity. There is a light–surface interaction in perceived dissimilarity. We were able to test this hypothesis in a different manner that did not depend on our choice of models. We took all dissimilarity measures between surfaces under Light 3 (most intense) and all dissimilarity measures between surfaces under Light 1 (least intense) and performed a t test (two independent samples) on the two sets of measures. We found that the dissimilarity measures under Light 3 were, on average, greater than those under Light 1 [$t(48) = 7.413$, $p < .0001$], in agreement with the outcome of the nested hypothesis test.

Dependence on light intensity and surface albedo. We next tested the hypothesis that $w_1 = 0$. If this hypothesis were true, perceived dissimilarity would not depend on illumination (*dissimilarity constancy* with changes in illumination). We found that we could reject this hypothesis [$\chi^2(1) = 188.65$, $p < .0001$]. Not surprisingly, we could also reject the hypothesis that $w_s = 0$ [$\chi^2(1) = 604.84$, $p < .0001$]: Lightness perception is influenced by surface albedo. These results imply that the index of illuminant discounting is greater than 0; that is, changes in illumination affect perceived dissimilarity.

Order independence. The order in which the nested hypotheses were tested could conceivably affect the overall conclusions drawn. Accordingly, we verified that the same tests in different orders led to the same outcomes. Normally, we would use a Bonferroni correction for multiple tests in testing a series of hypotheses, but the statistical values reported above are so extreme that such a correction would change nothing. We will omit it from the discussion.

The resulting model with unneeded parameters omitted can be seen in Equation 5 below, with $w_s = 2.81$, $w_1 = 0.69$, and $d = 1.10$, the κ being equal to 0.25. We can interpret this value of κ as follows: Doubling the surface albedo ratio α_i/α_j had roughly the same effect on perceived dissimilarity as increasing the light intensity ratio L_k/L_m by a factor of $2^4 = 16$. The estimate of $\sigma = 1.49$ is an estimate of the reliability of the observer's judgments on the 30-point rating scale.

Discussion

To summarize, rated dissimilarity is affected by both illumination difference and surface difference, and the

arc curvature and expansion apparent in Figure 2A are statistically significant. It follows that achromatic colors constitute an essentially two-dimensional manifold with a readily interpretable parametric structure. Among the class of models considered, the model that best fit the observers' data (Equation 5) is schematized in Figure 2B. It captures the major features of Figure 2A.

In Figure 3, we plot the mean dissimilarity rating Δ for each pair of test objects versus the corresponding prediction $\hat{\Delta}$ of the model specified in Equation 5. We estimated the coefficients of a linear regression of Δ against $\hat{\Delta}$. The best-fitting regression line was $\Delta = 0.997\hat{\Delta} + 0.039$. The slope is not significantly different from 1, and the intercept is not significantly different from 0 (both tests at the .01 level). The regression line is shown in the plot. The linear regression accounts for .950 of the variances.

The residuals $\hat{\Delta} - \Delta$ appear to be unpatterned. They also appear to be homoscedastic, consistent with our modeling assumption that σ (the observer's uncertainty in rating dissimilarity) does not vary with the magnitude of dissimilarity rated. The maximum likelihood estimate of σ was 1.49, and the residuals fall roughly within $\pm 2\sigma$ of the line, as was expected. A histogram of the residuals is shown in the inset (the range of the horizontal axis is ± 4). It is roughly Gaussian in appearance, in agreement with our modeling assumption that the observer's uncertainty is Gaussian.

Returning to Figure 2A, we note that for any fixed illuminant, the locus of achromatic colors for equi-illuminated surfaces forms the familiar one-dimensional continuum that we might think of as *lightness*. However, differences in length and curvature between the arcs in Figure 2A imply that each of the three levels of illumination employed in the experiment brings about its own lightness continuum. When illumination decreases, the lightness continuum shrinks. In its limit—that is, in the darkness—the lightness continuum contracts into a point (in the dark, all cats are gray). It follows that there are, strictly speaking, no chips with the same lightness in two identical series of Munsell chips under two different illuminations. Therefore, an exact match in asymmetrical lightness matching is impossible in principle. The best that observers can do is to point out samples with minimal subjective difference, and we conjecture that that is all that they ever do.

An examination of the output configuration in Figure 2A suggests that any surface under one illuminant is *least dissimilar* to itself under a second illuminant. In other words, it follows from our data that although an illumination change results in a change in the object's lightness, the object remains close to itself (in terms of dissimilarity), a form of *relative lightness constancy*. This perhaps explains why, despite being generally good at distinguishing between material (reflectance) and illumination changes (Craven & Foster, 1992), observers typically have prob-

$$\delta(ik, jm) = \left[w_s \Gamma(k, m \mid c, w_1) \left| \log \alpha_i - \log \alpha_j \right| + w_1 \left| \log L_k - \log L_m \right| \right]^d \tag{5}$$
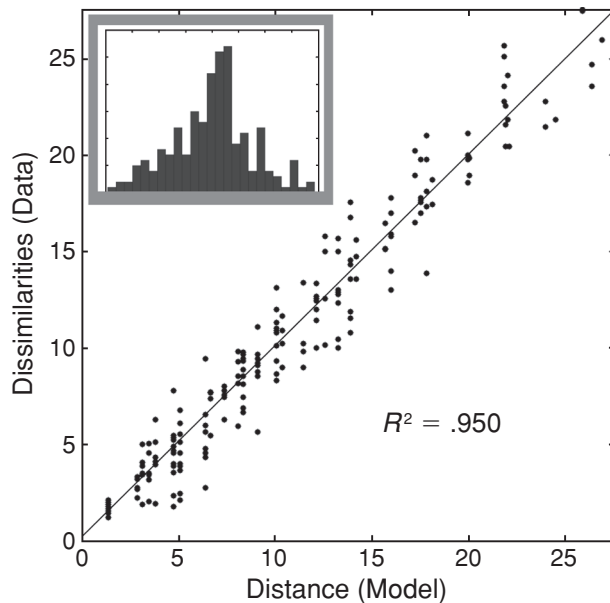
Figure 3. Dissimilarities versus predicted dissimilarities. The mean dissimilarity rating Δ for each pair of text–objects (vertical axis) is plotted versus the corresponding prediction $\hat{\Delta}$ of the model specified in Equation 5 (horizontal axis). The best-fitting regression line is shown. A linear regression accounts for .950 of the variances. The residuals are unpatterned and homoscedastic. A histogram of the residuals $\hat{\Delta} - \Delta$ is shown in the inset. It is roughly Gaussian in appearance.

lems with matching chips of the same lightness when they are differently illuminated. In the latter case, observers are facing a problem that does not have a solution.

EXPERIMENT 2
Relative Lightness Constancy

In comparing across lighting conditions, observers systematically chose the chip with roughly the same albedo as the one that was *least dissimilar* to any designated chip. However, only seven distinct albedos were employed in the experiment, and observers' performance might have been a consequence of the small number of albedos employed. With more reflectances, we might have found deviations from relative lightness constancy. Also, it remains open whether such relative lightness constancy will occur if one instructs observers directly to choose a Munsell chip under one illumination that is least dissimilar to the same chip under the other illumination. These issues were examined in the following experiment.

Method

The same observers were presented with two series of 20 gray chips randomly arranged as in Figure 4. The same two (of three) illuminants (175 and 1,694 lux) were used to illuminate the chips on the left and right halves of the display. The experiment was divided into two parts. In the first part, we used a stimulus display in which the chips were separated by a random-dot strip, an illumination edge passing through this strip (Figure 4A). In the second part, the il-

lumination edge came through the gray chips (Figure 4B), which was similar to the experimental conditions used by Gilchrist and Zdravkovic (2000; Zdravkovic & Gilchrist, 2000) in their recent experiments on lightness constancy.

Using a laser pointer, the experimenter pointed out, in random order, a chip on the left (highly illuminated) side and asked the observers to find a chip on the right (dimly illuminated) side that looked most similar to it. Each measurement was repeated three times for each observer. Note that we used the term *dissimilarity* in Experiment 1, in which we asked the subjects to judge differences. However, we avoided the phrase *least dissimilar* in Experiment 2, because we judged it to be potentially confusing. We used, instead, the term *most similar*.

Results and Analysis

The medians (of nine measurements) are presented in Figure 5. The data for the first part of the experiment (no illumination edge present) are plotted as diamonds; the data for the second part (illumination edge present) are plotted as circles.

As one can see, the data obtained in both experimental conditions (i.e., in the first and second parts) lie very close to the bisection, meaning that all the observers almost always chose the same reflectance as the one that was most similar to that pointed out by the experimenter. Although there is systematical deviation from the bisection, this deviation is very small; at least, it is much smaller than the deviation from lightness constancy found recently by Gilchrist and Zdravkovic (2000). Note that we find roughly the same degree of relative lightness constancy in Experiment 2 as we found in Experiment 1.

Discussion

We believe that the difference between our results and Gilchrist and Zdravkovic's (2000) are due to differences in the observers' task. Gilchrist and Zdravkovic used a classical lightness matching task, whereas we measured dissimilarities. Instructing observers to match the lightness of differently illuminated objects implies the ability, first, to abstract just one dimension of achromatic colors (i.e., lightness) and, second, to compare this abstracted value of lightness with the other lightness value abstracted from the other object lit by different light. There are at least two reasons why we doubt that observers are capable of either of these two operations. First, it is apparently difficult for observers to isolate one perceptual dimension (presumably corresponding to *lightness*) of achromatic color in the two-dimensional manifold we found. Second, even if they can accurately and consistently compare lightnesses under a single illuminant, it is not obvious that observers can compare lightness values from lightness scales induced by different illuminants. We believe that such comparisons are impossible in principle, and we will return to this point in the conclusion.

CONCLUSION

We began with a puzzle. Although asymmetric matching is widely used in the study of lightness and color matching, observers, given two surfaces, cannot always
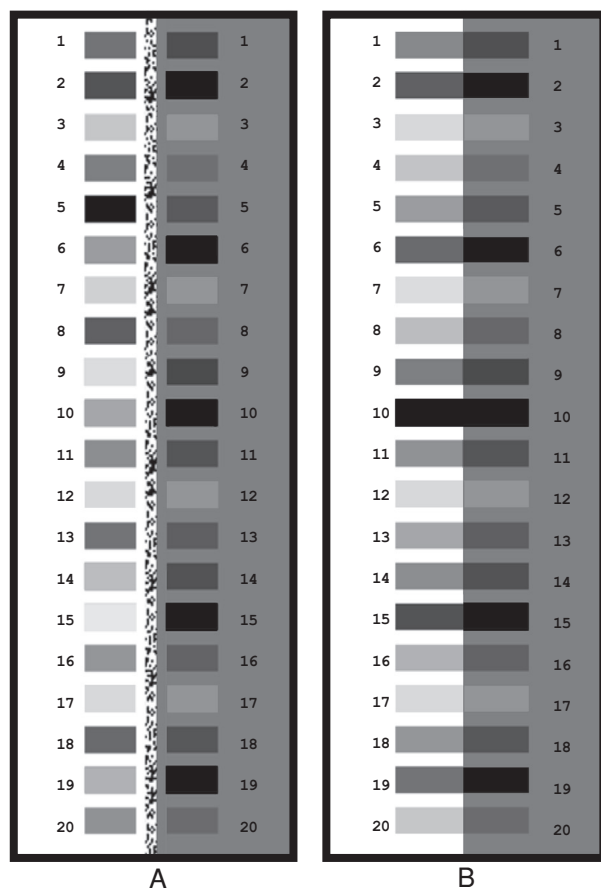
Figure 4. Example of the kinds of stimulus displays used in the second experiment. (A) The stimulus display used in the first part of the experiment. Test chips are separated by a vertical random-dot strip. The ordering of albedos in the left and right columns was different and randomized. The illumination differs on the right and left sides, with the illumination edge passing through the random-dot strip separating the chips. Using a laser pointer, the experimenter pointed out, in random order, a chip on the left (highly illuminated) side and asked the observers to find a chip on the right (dimly illuminated) side that looked most similar to it. Each measurement was repeated three times for each observer. (B) In the second part, the illumination edge ran through the gray chips. This condition was similar to the experimental conditions used by Gilchrist and Zdravkovic (2000).

adjust the albedo or surface color of the first to match the second. When observers cannot achieve a satisfactory match, it may be wondered what they might be doing instead. An evident conjecture is that they select a setting point that minimizes but does not eliminate the perceived dissimilarity between two surfaces under different lights.

Our first goal, then, was to study perceived dissimilarity between pairs of illuminated surfaces and to model it as a metric defined in terms of differences in light intensities and surface albedos. Our second goal was to ascertain which surfaces are perceived as *least dissimilar* under different illuminations.

In Experiment 1, we asked the observers to rate the dissimilarities of pairs of achromatic chips viewed under pos-

sibly different illuminations. Achromatic surface colors were found to constitute a two-dimensional manifold. We applied a novel method MLPS to test a series of hypotheses concerning the structure of this manifold and found that the manifold had a readily interpretable structure. On the basis of dissimilarity judgments only, we developed an index of illumination discounting that we estimated to be 0.25: Changing log reflectance was roughly four times as effective in producing a perceptual difference between two surfaces as changing log illumination intensity.

The advantages of MLPS over MDS are evident: We can explicitly test conjectures concerning the parameters of an explicit model of dissimilarity judgment and estimate these parameters with MLPS.

In Experiment 2, we found that observers chose surfaces with nearly the same albedos as the ones that were *least dissimilar* when they were lit by different lights.

Our results imply that there are two distinct perceptual dimensions of achromatic surface color. Consider an ideal dissimilarity observer whose judgments are perfectly predicted by the model summarized in Figure 2B. For this observer, the first dimension corresponds to lightness or perceived albedo and is encoded as position along the circular arcs in Figure 2B. Two distinct light–surface pairs that fall on the same radius in Figure 2B have the same albedo but are illuminated by lights differing in intensity. For the ideal observer, the dissimilarity between these two light–surface pairs cannot be decreased further by adjusting the albedo of either surface. The chips still differ
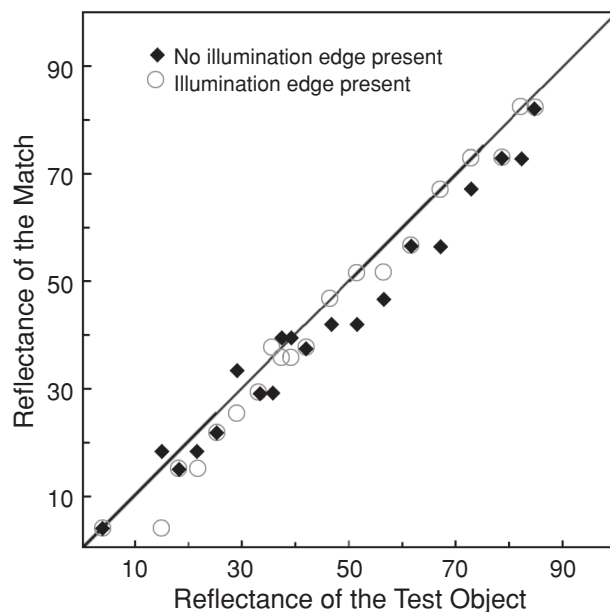

Figure 5. The results of the second experiment. The medians (of nine measurements) for the first part of the experiment (no illumination edge present, corresponding to the stimulus illustrated in Figure 4A) are plotted as solid diamonds. The medians (of nine measurements) for the second part of the experiment (illumination edge present, corresponding to the stimulus illustrated in Figure 4B) are plotted as open circles.

in appearance "[y]et turning any of the knobs or combinations of knobs [increases] the perceptual difference" (Brainard et al., 1997, p. 2098). This perceptual difference, noted by Brainard and colleagues, corresponds to a second dimension of achromatic surface color perception that is associated with changes in illumination, rather than with changes in surface albedo. The two dimensions are incommensurable: A change in one dimension cannot be compensated for by a change in the other. We refer to this second dimension as *surface brightness*[4] and conjecture that it is the reason why asymmetric matching is not always possible, for real, as well as for ideal, observers. Achromatic color appearance encodes more than just surface albedo.

We found that perceived dissimilarity judgment is, effectively, the weighted combination of light and surface *cues to dissimilarity*. Given that cue weights in depth cue combinations typically change in response to changes in the reliability of the individual cues (Landy, Maloney, Johnston, & Young, 1995; Young, Landy, & Maloney, 1993), it is interesting to examine whether κ changes in an interpretable way as viewing conditions change—for example, as the number of distinct surfaces in the scene N (the scene articulation) is increased. If it does, then the relationship between κ and N might give some insight into the mechanisms governing surface color perception.

REFERENCES

BRAINARD, D. H. (2003). Color appearance and color difference specification. In S. K. Shevell (Ed.), *The science of color* (2nd ed., pp. 191-216). Washington, DC: Optical Society of America.

BRAINARD, D. H., BRUNT, W. A., & SPEIGLE, J. M. (1997). Color constancy in the nearly natural image: I. Asymmetric matches. *Journal of the Optical Society of America A*, **14**, 2091-2110.

BRUNSWIK, E. (1929). Zur Entwicklung der Albedowahrnehmung. *Zeitschrift für Psychologie*, **109**, 40-115.

COX, T. F., & COX, M. A. A. (2001). *Multidimensional scaling* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

CRAVEN, B. J., & FOSTER, D. H. (1992). An operational approach to colour constancy. *Vision Research,* **32**, 1359-1366.

FOSTER, D. H. (2003). Does colour constancy exist? *Trends in Cognitive Sciences*, **7**, 439-443.

GILCHRIST, A. L., KOSSYFIDIS, C., BONATO, F., AGOSTINI, T., CATALIOTTI, J., LI, X., ET AL. (1999). A new theory of lightness perception. *Psychological Review*, **106**, 795-834.

GILCHRIST, A. L., & ZDRAVKOVIC, S. (2000). Mining lightness errors. *Investigative Ophthalmology & Visual Science*, **41**(Suppl.), 226.

HURLBERT, A. C. (1998). Computational models of color constancy. In V. Walsh & J. J. Kulikowski (Eds.), *Perceptual constancy: Why things look as they do* (pp. 283-322). Cambridge: Cambridge University Press.

KARDOS, L. (1934). Ding und Schatten: Eine experimentelle Untersuchung über die Grundlagen des Farbensehens. *Zeitschrift für Psychologie*, **23**(Suppl.), 1-184.

KATZ, D. (1935). *The world of colour* (R. B. MacLeod & C. W. Fox, Trans.). London: Kegan Paul, Trench, Trubner.

KENDALL, D. G. (1971). Seriation from abundance matrices. In F. R. Hodson, D. G. Kendall, & P. Tautu (Eds.), *Mathematics in the archae-ological and historical sciences* (pp. 215-252). Edinburgh: Edinburgh University Press.

LANDY, M. S., MALONEY, L. T., JOHNSTON, E. B., & YOUNG, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, **35**, 389-412.

MALONEY, L. T. (1999). Physics-based approaches to modeling surface color perception. In K. R. Gegenfurtner & L. T. Sharpe (Eds.), *Color vision: From genes to perception* (pp. 387-422). Cambridge: Cambridge University Press.

MALONEY, L. T., WUERGER, S. M., & KRAUSKOPF, J. (1995). A method for testing Euclidean representations of proximity judgments in linear psychological spaces. In R. D. Luce, M. D'Zmura, D. Hoffman, G. J. Iverson, & A. K. Romney (Eds.), *Geometric representations of perceptual phenomena: Papers in honor of Tarow Indow on his 70th birthday* (pp. 137-152). Mahwah, NJ: Erlbaum.

MALONEY, L. T., & YANG, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision*, **3**, 573-585.

MAUSFELD, R. (1998). Color perception: From Grassmann codes to a dual code for object and illuminant colors. In W. G. K. Backhaus, R. Kliegl, & J. S. Werner (Eds.), *Color vision: Perspectives from different disciplines* (pp. 219-250). Berlin: De Gruyter.

MOOD, A. M., GRAYBILL, F. A., & BOES, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.

THOULESS, R. H. (1931). Phenomenal regression to the real object: I. *British Journal of Psychology*, **21**, 339-359.

WUERGER, S. M., MALONEY, L. T., & KRAUSKOPF, J. (1995). Proximity judgments in color space: Tests of a Euclidean color geometry. *Vision Research*, **35**, 827-835.

YOUNG, M. J., LANDY, M. S., & MALONEY, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*, **33**, 2685-2696.

ZDRAVKOVIC, S., & GILCHRIST, A. L. (2000). Lightness determination for an object under two illuminations. *Perception*, **29**(Suppl.), 73.

NOTES

1. The data were averaged across observers after regressing the mean ratings for each observer against each other observer to see whether the observers had any appreciable differences in how they used the rating scales. They did not, a result we attribute to the care we took in instructing the subjects on scale use.

2. This notion is usually applied to discounting of temporal changes of an ambient illumination. We use it here to refer to discounting of spatial variations of local illumination as well.

3. Two parametric models are nested when one (the *nested*) is a special case of the other, achieved by setting some of the parameters of the other to prespecified values. We could, for example, test the model $y = ax^b$ (*power law*) against the nested model $y = ax$ (*linear law*) achieved by setting $b = 1$. A nested hypothesis test compares the difference of the log likelihoods of maximum likelihood fits to both models (the decision variable) and determines whether the nested model should be rejected in favor of the nonnested. In the example above, we are testing the null hypothesis $H_0: b = 1$ against the alternative $H_1: b \neq 1$. Twice the decision variable for the hypothesis test is distributed as χ^2_n when the null hypothesis holds. The degrees of freedom is the number of parameters set to specified values in passing from the nonnested to the nested hypothesis (here, $n = 1$; see Mood, Graybill, & Boes, 1974, pp. 440ff).

4. We emphasize that the observers in our experiments rated dissimilarities of surface appearance. Although *surface brightness* covaries with illumination intensity, it describes a perceptual dimension of achromatic surface appearance.