The Verriest Lecture: Color vision in an uncertain world

David H. Foster

University of Manchester, School of Electrical & Electronic Engineering, Sackville Street Building, Manchester M13 9PL, UK
(d.h.foster@manchester.ac.uk)

The natural world is optically unconstrained. Surface properties may vary from one point to another, and reflected light may vary from one instant to the next. The aim of this work is to quantify some of the physical failures of color vision performance that result from uncertainty. In computational simulations with images of vegetated and nonvegetated outdoor scenes, it is shown that color provides an unreliable guide to surface identity. It is also shown that changes in illuminant may cause colors to no longer match and the relations between individual colors to vary. These failures are generally well described by a measure of the randomness of the colors in scenes, the Shannon entropy. Although uncertainty is intrinsic to the environment, its consequences for color vision can be predicted.

1. INTRODUCTION

Color provides us with information about the environment. It allows us to divide scenes into distinct regions and, within them, to detect, distinguish, and identify objects. Yet what we know about color vision in such tasks comes largely from laboratory measurements with highly simplified stimuli, typically colored planar geometrical shapes, repetitive patterns, and blocks tableaux, which rarely capture the unconstrained, nonuniform optical structure of the natural world [1]. Scenes containing trees, shrubs, herbs, flowers, grasses, soil, stones, and rock are spatially and spectrally complex. Surface reflectance properties may vary from one point to another, and the light reflected from those surfaces may vary from one instant to the next due to changes in the illumination [2,3].

How does this uncertainty of the environment limit the utility of color? The aim of this work is to quantify in computational simulations several basic physical failures in color vision performance. These failures are then related to a measure of the randomness of the colors in scenes, the Shannon entropy [4].

This paper is based in part on the author's Verriest Medal Lecture delivered to the International Colour Vision Society, Erlangen, 2017. Except for Fig. 5, none of the data presented in the figures has been published previously.

2. TRANSLATING TASKS

One way to generalize traditional observations of surface color properties is to translate tasks to a more naturalistic setting. Consider an experiment to measure color constancy, the degree to which the color of a surface is seen or inferred to be constant despite changes in the spectral composition of the illumination [5]. Instead of presenting a test surface or object in an abstract geometric ("Mondrian") array [6] or in a simple tableau [7], it can be placed physically in a natural scene, as in Fig. 1. The test object is a small sphere attached to the trunk of a tree, in the center right of each image. The sphere is covered in Munsell neutral matte paint. The images were rendered from a hyperspectral reflectance image of the scene under an illuminant representing, on the left, direct sunlight with correlated color temperature 4000 K, and on the right, a mixture of sunlight and skylight with correlated color temperature 6700 K [8]; see [9] for more comprehensive daylight spectra.

In laboratory experiments, the spectral reflectance of the sphere was changed by a random amount so that when rendered it had a different surface color. Presented with two images of the scene in succession, the task of the observer was to discriminate between an illuminant change alone, as in Fig. 1, and an illuminant change accompanied by a change in the reflecting properties of the test surface. This task converts a subjective approach to color-constancy judgments, based on
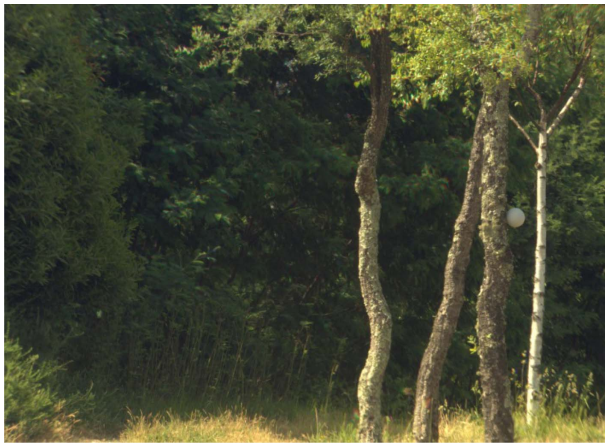
Fig. 1. Translating a color-constancy discrimination task to a natural scene. The images were rendered from a hyperspectral reflectance image of a woodland scene under an illuminant representing on the left, direct sunlight with correlated color temperature 4000 K, and on the right, a mixture of sunlight and skylight with correlated color temperature 6700 K. The test object is a small sphere attached to the trunk of a tree, center right [10]. The scene is from Sameiro, Braga, Portugal [11].

color appearance, to a more objective, operational one [12] that can be performed quickly and accurately [13,14]. As expected, observers' performance varies with the scene, but it can be broadly accounted for by variations in the spatial ratios of cone excitations across pairs of points under the two illuminants [10]. More generally, color constancy is supported by a variety of physical, physiological, and cognitive cues [2,5,7,15–17].

Although these laboratory experiments are informative, they do not really address the problem of uncertainty in natural scenes:

1. Translated tasks remain primarily deterministic. For each scene, the test object is usually fixed, and the only uncertainty is the extent to which its surface reflectance properties vary from trial to trial.

2. Much of the information in the image may be redundant. In surface-color judgments, for example, the task can be executed by comparing the test surface with just one other surface in the scene [18].

3. Performance is subject to physical constraints that are not always made explicit. Trichromatic sampling leads to phenomena such as metamerism, whereby surfaces may appear the same under one illuminant and different under another [8,19].

4. Observers may or may not use the cues available. Those they prefer may be imperfect, leading to suboptimal performance [20].

For the effects of scene uncertainty to be properly tested, all the elements of the scene should be given equal prior status, metamerism should be taken into account where relevant, and sampling should be random. Moreover, the effects of scene uncertainty should be distinguished from the effects of observer uncertainty, since they are intrinsically different [21,22].

3. COUNTING COLORS AND SURFACES

The fundamental role of scene uncertainty is evidenced in discriminating surfaces by their color. A naïve limit on realizable performance is set by the number of discriminable colors available to the trichromatic eye, or camera, with appropriate allowance for different sensor spectral sensitivities. Traditionally,

this number is estimated from a theoretical color gamut. A three-dimensional representation of the collection of all possible surface colors, the object color solid [19,23] is expressed in a uniform color space such as CIECAM02 [24,25] or the less uniform space CIELAB [8], which can be augmented by a chromatic adaptation transform CMCCAT2000 [25,26] and color-difference formula CIEDE2000 [27,28]. For further details, see [29]. The solid is divided into unit cells of side equal to the minimum discriminable, usually defined by a hard discrimination threshold, ΔE^{thr} say, and the number of cells then counted. By these methods, the number of discriminable colors is around 2.0–2.3 million [30,31].

Clearly, natural scenes do not generally contain all possible surface colors. But counting procedures can still be applied by considering just the nonempty units cells in the representation of the scene in the chosen color space. By way of example, the image on the left in Fig. 2 shows a scene from Ruivães, Vieira do Minho, Portugal. With the side of each unit cell ΔE^{thr} in CIECAM02 set to 0.5, approximately equivalent [32] to a CIELAB threshold value of 1.0, the number of discriminable colors in the scene, i.e., the gamut number, is approximately 8.9×10^4.

Other scenes may contain less or more than this number. An analysis of 50 hyperspectral images of vegetated and nonvegetated outdoor scenes [33] suggested an average number of discriminable colors per scene of approximately 2.7×10^5. Further details of the scenes are given at the end of Section 4.

This result does not imply, however, that on average 2.7×10^5 surfaces in a scene can be discriminated by their color. The reason is that the distribution of colors within natural scenes is rarely uniform: individual instances occur with differing frequencies. In Fig. 2, the pixels making up the flower, a foxglove, in the bottom right of the image, comprise just 0.63% of the total number of pixels.

The nonuniformity in the distribution of colors has consequences. The pink-purple color of the foxglove enables it to be readily discriminated from the other constituents of the scene, but the green of a particular fern or the yellow-brown of a

Fig. 2. Example scene to illustrate counting colors and surfaces. The color image on the left was rendered from a hyperspectral radiance image of a woodland scene. The image on the right was obtained by histogram equalization of the gray-scale rendering of the hyperspectral image. The scene is from Ruivães, Vieira do Minho, Portugal [11].

particular branch does not allow it to be readily distinguished from any other fern or branch. Despite there being three clear colors in the scene, only one individual object can be distinguished by its color (the set of all ferns can of course be distinguished from the set of all branches and from the set of all foxgloves). This description is an oversimplification; even so, it captures the distinction between the two ways of discriminating.

To estimate the number of discriminable surfaces in natural scenes, a different method of counting is needed that takes into account the nonuniformity of observed color distributions.

4. ENTROPY OF COLORS AND GAMUT VOLUME

How should a random distribution of surface colors be measured? Take the scene illustrated in Fig. 2, left. Suppose that the distribution of its colors in a color space such as CIECAM02 is described by some probability density function (pdf), f say. The plots in Fig. 3 show histogram estimates of the marginal distributions of lightness J (top left), the red-green chroma component a_C (bottom left), and the yellow-blue chroma component b_C (bottom right), each calculated according to the CIECAM02 specification with its default parameters but, importantly, with full chromatic adaptation [25]. Unsurprisingly, a_C signals mainly green and b_C mainly yellow. The flattened histogram of adjusted lightness values J (top right) is discussed later.

The color triplets (J, a_C, b_C) may be treated as instances $\mathbf{u}$ of a trivariate continuous random variable $\mathbf{U}$ whose pdf is f. The uncertainty or randomness of $\mathbf{U}$ may be quantified by the Shannon differential entropy $h(\mathbf{U})$ thus:

$$h(\mathbf{U}) = -\int f(\mathbf{u}) \log f(\mathbf{u})\, d\mathbf{u}, \qquad (1)$$

where the entropy is measured in bits if the logarithm is to the base 2 [4].

Differential entropy may be interpreted as the logarithm of the volume of the smallest set containing most of the probability, otherwise known as the effective support size of the random variable [34,35]. In the present context, the differential entropy $h(\mathbf{U})$ of a trivariate continuous random variable $\mathbf{U}$, the entropy of colors for short, measures the logarithm of the volume of the most common colors in the gamut. If $h(\mathbf{U})$ is in bits, then the effective gamut volume is $2^{h(\mathbf{U})}$. In the limit where all the colors are equally likely, the differential entropy coincides with the logarithm of the gamut volume. Evidently its value depends on the units of the dimensions of the color space.

This interpretation of differential entropy can be tested by using it to estimate the gamut volume of the lightness values in the scene in Fig. 2, left. The histogram estimate of the
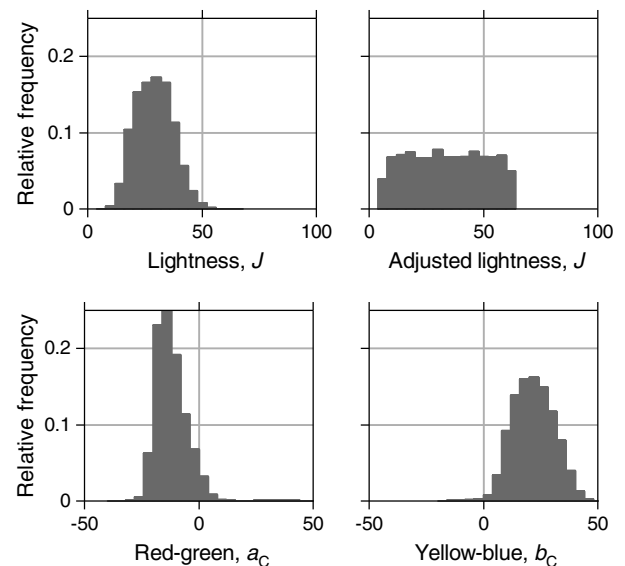
Fig. 3. Histograms of relative frequencies of lightness values J (top left) and red-green and yellow-blue chroma component values a_C (bottom left) and b_C (bottom right) of the CIECAM02 representation of the scene in Fig. 2, left. The histogram of relative frequencies of adjusted lightness values J (top right) is from the histogram-equalized gray-scale rendering of the scene in Fig. 2, right.

distribution of J values in Fig. 3, top left, was flattened by applying empirical histogram equalization, so that the relative frequency after adjustment was approximately constant over the interval from the minimum $J = 5.7$ to the maximum $J = 62.6$, as in Fig. 3, top right. The corresponding gray-scale image is shown in Fig. 2, right. Substituting the histogram for the pdf f in Eq. (1) and evaluating the integral numerically yields an estimated differential entropy $h(\mathbf{U}) = 5.80$ bits. The gamut volume $2^{h(\mathbf{U})}$, that is, the range of lightness values, is then 55.8, which is close to the correct value of $62.6 – 5.7 = 56.9$. This differential entropy calculation is approximate because empirical histogram equalization does not completely flatten the distribution.

By comparison, for the original, unflattened histogram estimate of the distribution of J values in Fig. 3, top left, the estimated differential entropy $h(\mathbf{U}) = 4.97$ bits. The effective gamut volume $2^{h(\mathbf{U})}$, the effective range of lightness values, is then 31.3, almost half the actual range.

Although these entropy estimates are approximate, the latter because of bias in the numerical integration [36,37], they illustrate how the effective gamut volume is reduced with nonuniform color distributions. Crucially, it is the effective gamut volume that describes the discriminability of surfaces in scenes. Because the distribution f is bounded, the effective gamut volume is always less than or equal to the gamut volume.

An effective support size can also be defined for discrete random variables such as counts. The Shannon entropy then corresponds to the logarithm of the effective number of elements in the discrete set [4,35]. An example is given in Section 5. The relationship between the two kinds of entropies and effective support is discussed in [4,34].

In what follows, the entropies for each scene are all differential entropies, unless otherwise indicated, and they were estimated not by numerical integration but by an asymptotically bias-free nearest-neighbor method [38–40]. The scenes consisted of a set of 50 hyperspectral reflectance images [41], closely similar to those used in [33], where thumbnail color images are available. Each image had dimensions ≤ 1344 pixels $\times$ 1024 pixels and spectral range 400–720 nm sampled at 10-nm intervals. Reflectance properties are defined only for the particular viewing geometry at the time of acquisition, as represented by the spectral bidirectional reflectance distribution function, the BRDF [42]. Of the 50 scenes, 29 were mainly vegetated and 21 mainly nonvegetated. Further details are given in [41].
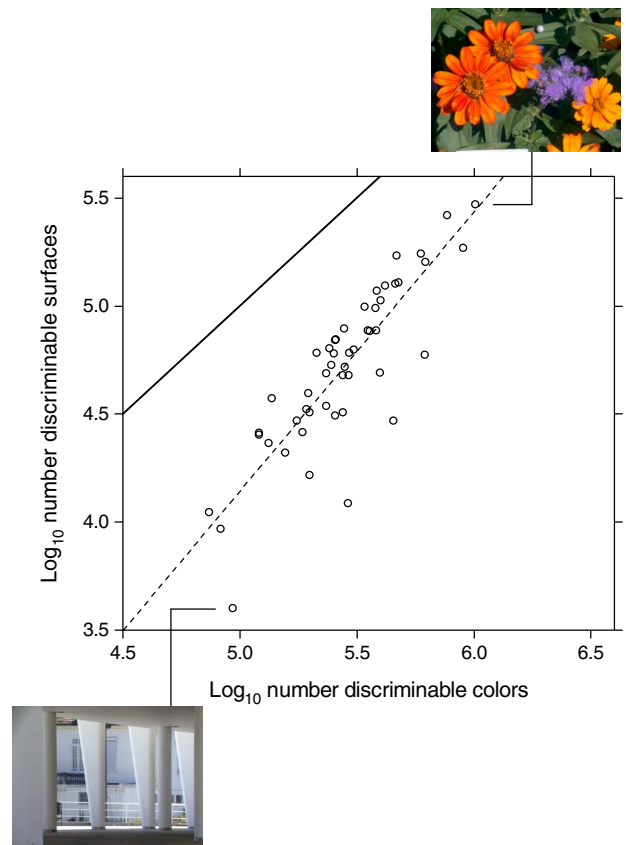
5. NUMBER OF DISCRIMINABLE SURFACES

Recall from Section 3 that the number of discriminable colors, the gamut number, is estimated by dividing the gamut volume into unit cells of side equal to the hard discrimination threshold ΔE^{thr} and then counting the number of cells, in other words, taking the quotient of the gamut volume by the cell volume. An analogous procedure can be invoked to estimate the effective gamut number from the effective gamut volume. The effective gamut number gives the number of surfaces discriminable by their color. All that is required is an appropriate formulation of cell volume.

A unit cell of side ΔE^{thr} represents the portion of the color space over which observer discrimination responses are random. These responses may be represented by a random variable, $\mathbf{W}$ say, whose uncertainty is quantified by the corresponding entropy $h(\mathbf{W})$, which, for a unit cell of side ΔE^{thr}, is equal to $3 \log \Delta E^{\mathrm{thr}}$ [4]. Alternatively, observer uncertainty may be defined by a smooth function, for example, a Gaussian distribution, whose entropy has a standard formulation [4]. Either way, the difference $h(\mathbf{U}) – h(\mathbf{W})$ between the entropy of colors $h(\mathbf{U})$ and the entropy of observer uncertainty $h(\mathbf{W})$ gives the logarithm of the number of discriminable surfaces [43]. As a difference of logarithms, it is equivalent to the logarithm of a quotient.

In such an analysis of a set of 50 outdoor scenes, the average number of discriminable surfaces per scene was found to be approximately 5.2×10^4 with a hard threshold [43], about one-fifth of the number of discriminable colors with the same threshold [33].

Notice that these entropy estimates relate solely to how color can be used to identify points or arbitrarily small surface elements in the scene, or pixels in a digital representation. The calculation is indifferent to what physically defines surfaces

Fig. 4. Number of discriminable surfaces and colors. The logarithm of the number of surfaces discriminable by their color is plotted against the logarithm of the number of discriminable colors. Each point represents data from one of 50 outdoor scenes, two of which are pictured [33,41]. The dashed line shows a linear regression, with $R^2 = 73\%$. The oblique solid line represents the upper limit on the number of discriminable surfaces for each scene. Unpublished data from [43].

and objects, for example, the foxglove and ferns in Fig. 2. Nevertheless, it supports the informal considerations of Section 3. Thus, suppose that the surface elements of the flower and of the ferns and the rest of the scene are merged into two separate uniform surfaces. The random variable $\mathbf{U}$ representing the colors in the scene is now discrete, with just two values: pink-purple and green. The probability p of a randomly chosen pixel being pink-purple is given by ratio of the areas, that is, $p = 6.3 \times 10^{-3}$, and the probability of its being green is $1 - p$. The entropy $H(\mathbf{U})$ of the colors is given by the discrete version of Eq. (1) with $(p, 1 - p)$ replacing f [4]; that is, $H(\mathbf{U}) = 0.055$ bits. The number of discriminable surfaces is then $2^{H(\mathbf{U})}$, that is, 1.04, very close to the estimate of 1 suggested in Section 3.

Despite the large differences between the number of discriminable colors and the number of discriminable surfaces, there is an interesting approximately linear relationship between the two for the majority of scenes, as illustrated in Fig. 4, taken from unpublished data in [43]. The dashed line shows a linear regression. The two scenes generating the most extreme numbers of discriminable surfaces values are pictured. Since the effective gamut volume must be less than or equal to the gamut volume (Section 4), the number of discriminable surfaces must be less than or equal to the number of discriminable colors. Equality is represented by the oblique solid line. All the data points fall below it.

6. COLOR MATCHING UNDER DIFFERENT LIGHTS

The consequences of scene uncertainty become more evident still when surfaces are sampled under different lights. Departures from ideal matching with individual surfaces may be described by various indices, including indices of metamerism [8] and the related index of inconstancy [44]. In metamerism, the illuminant under which surfaces match is usually called the reference and under which they differ the test [8,19]. For definiteness, color matching refers to the appearance of stimuli whose colorimetric specification varies continuously [8,19] and should not be confused with nearest-neighbor matching from a finite population of colors [40,45].

In principle, natural surfaces are particularly susceptible to metamerism, for their reflectance spectra generally have more than three degrees of freedom, in fact between five and eight, depending on the criterion for discrimination [46].

The practical importance of metamerism may be gauged in two ways, one absolute, the other conditional. As before, color triplets (J, a_C, b_C) in CIECAM02 may be treated as instances of random variables $\mathbf{U}$ and $\mathbf{V}$ under reference and test illuminants, respectively. Given two instances, $\mathbf{u}$ and $\mathbf{u}'$, under the reference illuminant, their vector color difference $\Delta\mathbf{u} = \mathbf{u}' - \mathbf{u}$ is defined component-wise, that is, $\Delta\mathbf{u} = (\Delta J, \Delta a_C, \Delta b_C)$. The magnitude of the difference, usually called the total color difference or just color difference, ΔE_r say, is calculated in the usual way [8,19] by $\Delta E_r = (\Delta J^2 + \Delta a_C^2 + \Delta b_C^2)^{1/2}$. Likewise, for instances $\mathbf{v}$ and $\mathbf{v}'$ and their vector color difference $\Delta\mathbf{v}$ under the test illuminant.

For a scene of interest, suppose that a sample of N pairs of surface elements is chosen at random. Let N_0 be the number of pairs $(\mathbf{u}, \mathbf{u}')$ in this sample whose color differences ΔE_r under the reference illuminant are less than the discrimination threshold ΔE^{thr}. From this subsample, let N_1 be the number of pairs $(\mathbf{v}, \mathbf{v}')$ whose color differences ΔE_t under the test illuminant are greater than or equal to ΔE^{thr}. Such pairs are sometimes described as parameric [47]. Necessarily, $N_1 \leq N_0 \leq N$. The relative frequency of metamerism is then N_1/N, and the conditional relative frequency of metamerism is N_1/N_0. In other words, N_1/N is an estimate of the probability that two surface elements chosen at random are indiscriminable under the reference illuminant but discriminable under the test illuminant, and N_1/N_0 is an estimate of the probability that, given two surface elements that are indiscriminable under the reference illuminant, they are discriminable under the test illuminant.

In simulations with the set of 50 scenes described in Section 4, illuminants were taken from the more extreme phases of daylight with correlated color temperatures of 4000 K and 25,000 K, characteristic of direct sunlight and polar skylight, respectively [8]. The discrimination threshold ΔE^{thr} in CIECAM02 was again set to 0.5, typical for these tasks [46]. Results were found to be similar with a larger CIECAM02 threshold of 1.0. Although not explored further, thresholds may be defined with respect to acceptability criteria [48] and categorization measures [49].

Averaged over the 50 scenes, the relative frequency of metamerism is approximately 1.8×10^{-4}. Values are smaller still with more demanding threshold criteria and with illuminants that are spectrally closer. Small values have also been reported elsewhere [49]. By contrast, the conditional relative frequency of metamerism is much larger, approximately 5.7×10^{-1}, i.e., 57%, with the same scenes and illuminants, the particular value depending on the criteria and conditions [41]. By this conditional measure, metamerism is common.
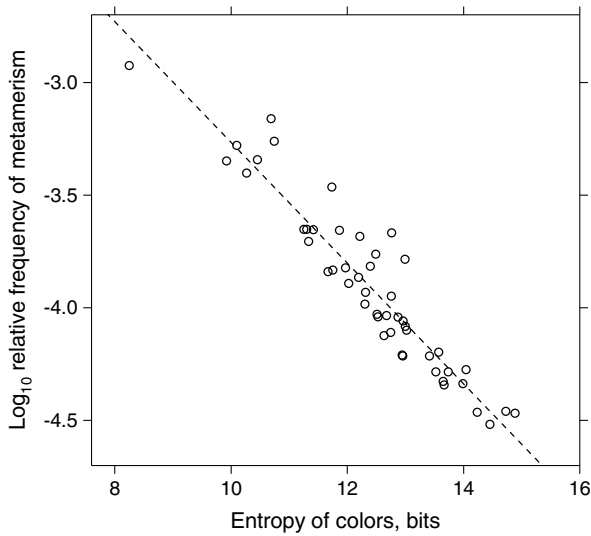
7. FREQUENCY OF METAMERISM

As with the number of discriminable surfaces, the value of the relative frequency of metamerism can be related to the appropriate entropies.

With the notation of Section 6, the quotient N_0/N is the relative frequency with which two surfaces chosen at random are indiscriminable under the reference illuminant, and the quotient N_1/N_0 is the conditional relative frequency with which those surfaces are discriminable under the test illuminant. The relative frequency N_1/N can be decomposed into their product thus:

$$\frac{N_1}{N} = \frac{N_1}{N_0} \cdot \frac{N_0}{N}.$$

Empirically, the variation in N_0/N dominates the variation in N_1/N_0 on a logarithmic scale [50], and N_0/N depends on the effective gamut: the smaller the effective gamut, the more likely that two surfaces chosen at random are indiscriminable. The logarithm of the relative frequency N_1/N should therefore be approximately inversely proportional to the entropy of colors under the reference illuminant.

Fig. 5. Relative frequency of metamerism and entropy of colors. The logarithm of the relative frequency N_1/N under 4000 K and 25,000 K daylight illuminants is plotted against the entropy of colors $h(\mathbf{U})$ under the 4000 K illuminant. Each point represents data from one of 50 outdoor scenes [33,41]. The dashed line shows a linear regression, with $R_2 = 90\%$. Earlier versions of this plot appear in [50,51].

To test this relationship, fresh computational simulations were performed with the 50 scenes described in Section 4. Because N_1/N is very small, as noted in Section 6, the accuracy of the simulations was improved by increasing the size of the sample N from each scene by several orders of magnitude to 3.4×10^8. Here and subsequently, the threshold ΔE^{thr} was set to 0.5.

Figure 5 shows data from the 50 scenes under the 4000 K and 25,000 K daylight illuminants. The logarithm of the relative frequency N_1/N is plotted against the entropy $h(\mathbf{U})$ of the colors of each scene under the 4000 K illuminant. The logarithm to the base 10 is used for ease of interpretation. Relative frequency ranges from 3.0×10^{-5} to 1.2×10^{-3}. The mean is 1.8×10^{-4}, the same as the value reported in Section 6 with fewer pairs in each scene sample.

There is a strong linear inverse relationship between the logarithm of the relative frequency of metamerism and entropy. A linear regression accounts for most of the variance, with $R^2 = 90\%$, a value closely similar to that reported in [50,51] with the same threshold criterion.

As for the relevance of the linear regression, the entropy estimates are not without error, and the relationship could instead be described by some form of orthogonal (e.g., standardized major axis) regression [52], though there is little change in slope.

8. CONDITIONAL FREQUENCY OF METAMERISM

Because the relative frequency of metamerism is governed by the probability of finding surface colors that are indiscriminable under the reference illuminant, the corresponding entropy of colors is uninformative on whether the surfaces

are discriminable under another illuminant, that is, the conditional relative frequency of metamerism. What is required is a conditional version of the entropy of colors.

For two random variables $\mathbf{U}$ and $\mathbf{V}$, the conditional differential entropy $h(\mathbf{V}|\mathbf{U})$ is defined by

$$h(\mathbf{V}|\mathbf{U}) = - \int f(\mathbf{v}, \mathbf{u}) \log f(\mathbf{v}|\mathbf{u}) \, d\mathbf{v} d\mathbf{u},$$

or, equivalently,

$$h(\mathbf{V}|\mathbf{U}) = h(\mathbf{V}, \mathbf{U}) - h(\mathbf{U}), \qquad (2)$$

where $h(\mathbf{V}, \mathbf{U})$ is the joint entropy of $\mathbf{U}$ and $\mathbf{V}$ [4]. As before, the entropy is measured in bits if the logarithm is to the base 2. In the present context, it is not the triplets (J, a_C, b_C) but the vector color differences $(\Delta J, \Delta a_C, \Delta b_C)$ that need to be considered. The random variable under the reference illuminant is $\Delta \mathbf{U}$ with values $\Delta \mathbf{u}$ subject to the constraint that their magnitudes $\Delta E_r < \Delta E^{thr}$. Likewise, the random variable under the test illuminant is $\Delta \mathbf{V}$ with values $\Delta \mathbf{v}$, where to count as metamers, their magnitudes $\Delta E_t \geq \Delta E^{thr}$. The conditional entropy $h(\Delta \mathbf{V}|\Delta \mathbf{U})$ then measures the logarithm of the effective gamut of the differences $\Delta \mathbf{v}$ under the test illuminant, given knowledge of the differences $\Delta \mathbf{u}$ under the reference illuminant. The larger this gamut, the more likely that two surfaces chosen at random are discriminable under the test illuminant. The logarithm of the conditional relative frequency of metamerism should therefore be approximately proportional to the corresponding conditional entropy of the vector color differences.

Figure 6 shows the logit of the conditional relative frequency N_1/N_0 plotted against the conditional entropy $h(\Delta \mathbf{V}|\Delta \mathbf{U})$ of the difference $\Delta \mathbf{V}$ under a 25,000 K daylight test illuminant,
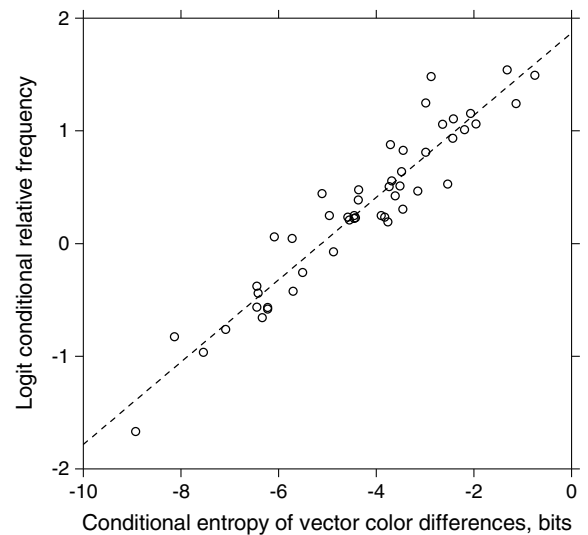
Fig. 6. Conditional relative frequency of metamerism and conditional entropy of vector color differences. The logit conditional relative frequency N_1/N_0 is plotted against the conditional entropy $h(\Delta \mathbf{V}|\Delta \mathbf{U})$ of the vector color difference $\Delta \mathbf{V}$ under a 25,000 K daylight illuminant given the subthreshold vector color difference $\Delta \mathbf{U}$ under a 4000 K daylight illuminant. Each point represents data from one of 50 outdoor scenes [33,41]. The dashed line shows a linear regression, with $R^2 = 90\%$.

given the subthreshold difference $\Delta \mathbf{U}$ under a 4000 K daylight reference illuminant. Each point represents data from one of 50 outdoor scenes. Other details of the sampling regime are the same as in Section 7. Exceptionally, the logit transform $\ln[q/(1-q)]$ of the conditional relative frequency q is used rather than a logarithmic transformation $\log q$ in order to treat small and large values symmetrically [53], a problem that does not occur with relative frequency, which is always small. Values of the conditional relative frequency range from 16% to 83% over the 50 scenes.

As anticipated in the preceding analysis, the logit of the conditional relative frequency increases as the conditional entropy increases, and there is a strong linear relationship between the two, with $R^2 = 90\%$. This result should be distinguished from that reported in a previous analysis [54], where the chosen explanatory variable was the conditional entropy $h(\mathbf{V}|\mathbf{U})$ rather than $h(\Delta \mathbf{V}|\Delta \mathbf{U})$. It provided a poorer description, with $R^2 = 64\%$.

The opposite directions of dependence in Figs. 5 and 6 are not inconsistent. In Fig. 5, the relative frequency of metamerism decreases as the entropy of colors increases because the chances of surfaces that are chosen at random being indiscriminable under a reference illuminant (and therefore potentially discriminable under the test) decrease as the effective gamut of colors gets larger. Conversely, in Fig. 6, the conditional relative frequency of metamerism increases with the conditional entropy of vector color differences because the chances of surfaces that are indiscriminable under a reference illuminant being discriminable under a test illuminant increase as the effective gamut of the differences under the test illuminant gets larger.

9. MAGNITUDE OF METAMERISM

For single pairs of surfaces, the usual measure of the magnitude of metamerism is the metamerism index, that is, the total color difference ΔE_t induced by substituting the test illuminant for the reference illuminant [8,19]. For multiple pairs of surfaces, the natural measure is the mean or the more robust median value of ΔE_t taken over all the pairs in the sample. Because the pairs $(\mathbf{u}, \mathbf{u}')$ under the reference illuminant are not exactly metameric, but instead have vector color differences $\Delta \mathbf{u}$ with magnitudes $\Delta E_r < \Delta E^{thr}$, an analog of the standard CIE correction was applied ([8], Section 9.2.2.3). In this correction, the residual vector color differences $\Delta \mathbf{u}$ under the reference illuminant are subtracted from the corresponding vector color differences $\Delta \mathbf{v}$ under the test illuminant, and the total color difference quantified by ΔE_t. This notion is developed further in Section 10. Since the mean or median value of ΔE_t is proportional to the effective gamut, its logarithm should be approximately proportional to the corresponding conditional entropy of the vector color differences.

Figure 7 shows the logarithm of the median value of ΔE_t plotted against the conditional entropy $h(\Delta \mathbf{V}|\Delta \mathbf{U})$ of the vector color difference $\Delta \mathbf{V}$ under a 25,000 K daylight illuminant given the subthreshold vector color difference $\Delta \mathbf{U}$ under a 4000 K daylight illuminant. Each point represents data from one of 50 outdoor scenes. Unsurprisingly, the variance about
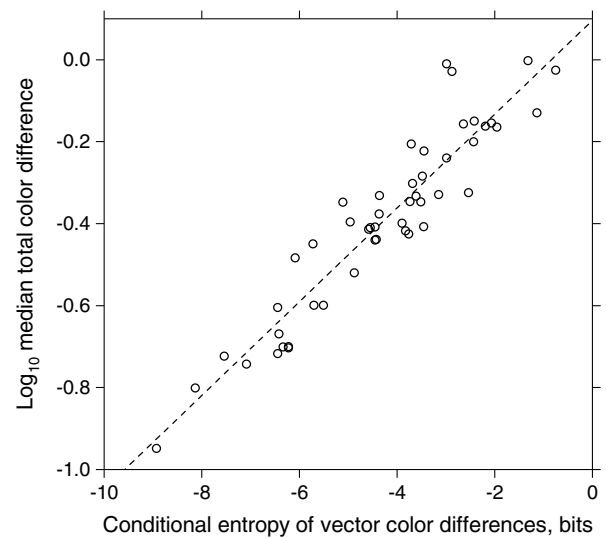
Fig. 7. Magnitude of metamerism and conditional entropy of vector color differences. The logarithm of the median total color difference ΔE_t under a 25,000 K daylight illuminant is plotted against the conditional entropy of the vector color difference $\Delta \mathbf{V}$ under a 25,000 K daylight illuminant given the subthreshold vector color difference $\Delta \mathbf{U}$ under a 4000 K daylight illuminant. Each point represents data from one of 50 outdoor scenes [33,41]. The dashed line shows a linear regression, with $R^2 = 88\%$.

the regression line is much the same as with the conditional relative frequency of metamerism in Fig. 6, with $R^2 = 88\%$.

10. MAGNITUDE OF GENERALIZED METAMERISM

The measures of metamerism described in Sections 6–9 contain an arbitrary component, the threshold ΔE^{thr}, below which colors are classified as indiscriminable and above which they are classified as discriminable. This arbitrariness is absent in a more comprehensive measure of the changes in surface color under a change in lighting, namely, generalized metamerism.

Generalized metamerism refers to pairs of surfaces failing to maintain their color relations with a change in illuminant ([41], Section 3.G, 4.C). Color relations or relative color cues are known to be important in surface color judgments [20,55–57], and their use has been identified in other species [58]. In the absence of a uniform color space, the color relations within a scene may be represented by the spatial ratios of cone excitations across pairs of points [59] or spatial ratios of linear combinations of these quantities [60]. As noted in Section 2, spatial ratios may be used to explain some color-constancy judgments. Within a uniform color space such as CIECAM02, however, color relations are more naturally represented by their vector color differences $(\Delta J, \Delta a_C, \Delta b_C)$. The two kinds of representations are plainly interdependent. Empirically, the median canonical correlation over scenes between vector color differences and the logarithm of spatial ratios is 97%.

Metamerism is a special case of generalized metamerism, in which color relations under the reference light are relations of approximate equality. Given the results of Section 9,

can changes in arbitrary color relations under a test light also be described by the conditional entropy of vector color differences?

As in Section 9, let the vector color differences be $\Delta\mathbf{u}$ under the reference illuminant and $\Delta\mathbf{v}$ under the test illuminant, but without the constraint that the $\Delta\mathbf{u}$ have magnitudes $\Delta E_r < \Delta E^{\text{thr}}$. Generalized metamerism is expressed by non-zero values of the vector difference of vector color differences $\Delta\mathbf{v} - \Delta\mathbf{u}$. Its magnitude may be described by the color difference $\Delta E_{r,t}$ in an obvious way [61]; that is, if $\Delta\mathbf{u} = (\Delta J_r, \Delta a_{C,r}, \Delta b_{C,r})$ and $\Delta\mathbf{v} = (\Delta J_t, \Delta a_{C,v}, \Delta b_{C,t})$, then

$$\Delta E_{r,t} = [(\Delta J_t - \Delta J_r)^2 + (\Delta a_{C,t} - \Delta a_{C,r})^2 + (\Delta b_{C,t} - \Delta b_{C,r})^2]^{1/2}.$$

Although the computation is analogous to that in Section 9, the symbol $\Delta E_{r,t}$ is used instead of ΔE_t to reflect the fact that the $\Delta\mathbf{u}$ are unconstrained.

Figure 8 shows the logarithm of the median value of $\Delta E_{r,t}$ plotted against the conditional entropy $h(\Delta\mathbf{V}|\Delta\mathbf{U})$ of the vector color difference $\Delta\mathbf{V}$ under a 25,000 K daylight illuminant given the vector color difference $\Delta\mathbf{U}$ under a 4000 K daylight illuminant. The variance about the regression line is somewhat greater than with strictly metameric pairs, as in Fig. 7, though it still accounts for most of the variance, with $R^2 = 75\%$. The median value of $\Delta E_{r,t}$ ranges from 0.74 to 6.37 in CIECAM02. The conditional entropies $h(\Delta\mathbf{V}|\Delta\mathbf{U})$ are larger than in Figs. 6 and 7 because, as just noted, the $\Delta\mathbf{u}$ are unconstrained.

It is worth emphasizing an essential distinction between the observed and explanatory variables in Fig. 8. The observed variable is the logarithm of the vector color difference
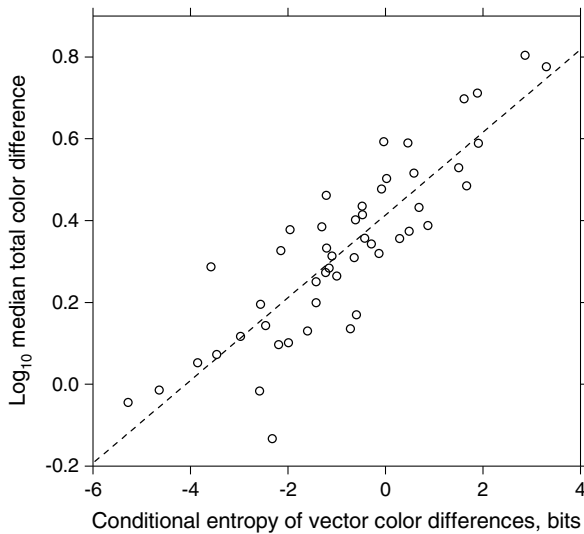
Fig. 8. Magnitude of generalized metamerism and conditional entropy of vector color differences. The logarithm of the median total color difference $\Delta E_{r,t}$ under a 25,000 K daylight illuminant is plotted against the conditional entropy $h(\Delta\mathbf{V}|\Delta\mathbf{U})$ of the vector color difference $\Delta\mathbf{V}$ under a 25,000 K daylight illuminant given the vector color difference $\Delta\mathbf{U}$ under a 4000 K daylight illuminant. Each point represents data from one of 50 outdoor scenes [33,41]. The dashed line shows a linear regression, with $R^2 = 75\%$. Notice the difference between the vertical scale and that in Fig. 7.

$\Delta\mathbf{V} - \Delta\mathbf{U}$ and, by the definition in Eq. (2), the explanatory variable is the difference of entropies $h(\Delta\mathbf{V}, \Delta\mathbf{U}) - h(\Delta\mathbf{U})$. In general, there is no simple relationship between entropies of differences and differences of entropies [35].

11. COMMENT AND CONCLUSION

The tasks considered in this analysis all illustrate the physical limitations on the utility of color in the environment, whether scenes are viewed by the trichromatic eye or camera. First, and most fundamentally, color provides an unreliable guide to surface identity. The average number of surfaces per scene that may be discriminated by their color is about one-fifth the number of discriminable colors. Second, colors may differ when the illumination changes. The probability that a pair of surfaces matching under a 4000 K daylight and not matching under a 25,000 K daylight is about 60%. Third, the relations between individual colors may vary with the illuminant. The median color difference associated with generalized metamerism ranges from about 0.7 to 6.4 times the minimum discriminable under a change from a 4000 K daylight to a 25,000 K daylight. These failures in realizable performance are generally well described by the estimates of the entropies of the colors involved.

There are two caveats to this analysis: one theoretical, the other practical. First, entropy estimates with other combinations of color signals may provide still better fits, especially for generalized metamerism. The present entropy estimates were chosen for simplicity and relevance, and need not be unique.

Second, there is a difference between the properties of a particular scene and the changes in those properties with illumination. The marked difference between the number of discriminable surfaces and the number of discriminable colors is not peculiar to a particular illumination. But the failures in color matching and variations in color relations with illumination do depend on the nature of those changes. The assumption that illumination changes are spatially uniform is acceptable in a task where individual pairs of surfaces are tested first under one illuminant and then under another [41] or where surfaces are both in direct illumination or both in the shade [3], but not necessarily under all illumination changes in the natural world.

The problem with natural illumination changes is that variations in the spectrum of the illumination are almost always accompanied by variations in geometry, as a result of movement of cloud, differences in atmospheric scatter, and local fluctuations in mutual illumination and shadows, attached and cast. These geometric variations [1,62] make the prediction of changes in reflected spectra more difficult. Knowing the colors of surfaces under a reference illumination is less useful in reducing the uncertainty of those colors under a test illumination that varies from one point to the next. Conditional entropies based solely on colors are therefore likely to be much larger.

For the illuminant changes considered here, however, entropy descriptions appear efficient. Uncertainty may be intrinsic to the environment, but its consequences for color vision can be predicted.

REFERENCES

1. L. Arend, "Environmental challenges to color constancy," in *Human Vision and Electronic Imaging VI*, B. E. Rogowitz and T. N. Pappas, eds. (SPIE, 2001), pp. 392–399.
2. A. Werner, "Spatial and temporal aspects of chromatic adaptation and their functional significance for colour constancy," Vision Res. **104**, 80–89 (2014).
3. D. H. Foster, K. Amano, and S. M. C. Nascimento, "Time-lapse ratios of cone excitations in natural scenes," Vision Res. **120**, 45–60 (2016).
4. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. (Wiley, 2006).
5. D. H. Foster, "Color constancy," Vision Res. **51**, 674–700 (2011).
6. L. Arend and A. Reeves, "Simultaneous color constancy," J. Opt. Soc. Am. A **3**, 1743–1751 (1986).
7. J. M. Kraft and D. H. Brainard, "Mechanisms of color constancy under nearly natural viewing," Proc. Natl. Acad. Sci. USA **96**, 307–312 (1999).
8. CIE, "Colorimetry, 3rd ed.," CIE Publication 15:2004 (CIE Central Bureau, 2004).
9. J. Hernández-Andrés, J. Romero, and J. L. Nieves, "Color and spectral analysis of daylight in southern Europe," J. Opt. Soc. Am. A **18**, 1325–1335 (2001).
10. D. H. Foster, K. Amano, and S. M. C. Nascimento, "Color constancy in natural scenes explained by global image statistics," Vis. Neurosci. **23**, 341–349 (2006).
11. S. M. C. Nascimento, K. Amano, and D. H. Foster, "Spatial distributions of local illumination color in natural scenes," Vision Res. **120**, 39–44 (2016).
12. B. J. Craven and D. H. Foster, "An operational approach to colour constancy," Vision Res. **32**, 1359–1366 (1992).
13. D. H. Foster, B. J. Craven, and E. R. H. Sale, "Immediate colour constancy," Ophthalmic Physiol. Opt. **12**, 157–160 (1992).
14. A. J. Reeves, K. Amano, and D. H. Foster, "Color constancy: phenomenal or projective?" Percept. Psychophys. **70**, 219–228 (2008).
15. R. J. Lee and H. E. Smithson, "Motion of glossy objects does not promote separation of lighting and surface colour," R. Soc. Open Sci. **4**, 171290 (2017).
16. D. Weiss, C. Witzel, and K. Gegenfurtner, "Determinants of colour constancy and the blue bias," I-Perception **8**, 1–29 (2017).
17. R. Lafer-Sousa and B. R. Conway, "#TheDress: categorical perception of an ambiguous color image," J. Vis. **17**(12), 25 (2017).
18. K. Amano, D. H. Foster, and S. M. C. Nascimento, "Minimalist surface-colour matching," Perception **34**, 1009–1013 (2005).
19. G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed. (Wiley, 1982).
20. S. M. C. Nascimento and D. H. Foster, "Detecting natural changes of cone-excitation ratios in simple and complex coloured images," Proc. R. Soc. London Ser. B. Biol. Sci. **264**, 1395–1402 (1997).
21. K. Chowdhary and P. Dupuis, "Distinguishing and integrating aleatoric and epistemic variation in uncertainty quantification," ESAIM Math. Model. Numer. Anal. **47**, 635–662 (2013).
22. C. Soize, *Uncertainty Quantification: An Accelerated Course with Advanced Applications in Computational Engineering*, Vol. **47** of Interdisciplinary Applied Mathematics (Springer Nature, 2017).
23. J. Morovič, *Color Gamut Mapping* (Wiley, 2008).
24. M. R. Luo, G. Cui, and C. Li, "Uniform colour spaces based on CIECAM02 colour appearance model," Color Res. Appl. **31**, 320–330 (2006).
25. CIE, "A Colour Appearance Model for Colour Management Systems: CIECAM02," CIE Publication 159:2004 (CIE Central Bureau, 2004).
26. C. Li, M. R. Luo, B. Rigg, and R. W. G. Hunt, "CMC 2000 chromatic adaptation transform: CMCCAT2000," Color Res. Appl. **27**, 49–58 (2002).
27. M. R. Luo, G. Cui, and B. Rigg, "The development of the CIE 2000 colour-difference formula: CIEDE2000," Color Res. Appl. **26**, 340–350 (2001).
28. M. Melgosa, R. Huertas, and R. S. Berns, "Performance of recent advanced color-difference formulas using the standardized residual sum of squares index," J. Opt. Soc. Am. A **25**, 1828–1834 (2008).
29. S. Westland, C. Ripamonti, and V. Cheung, *Computational Colour Science Using MATLAB*, 2nd ed. (Wiley, 2012).
30. M. R. Pointer and G. G. Attridge, "The number of discernible colours," Color Res. Appl. **23**, 52–54 (1998).
31. F. Martínez-Verdú, E. Perales, E. Chorro, D. de Fez, V. Viqueira, and E. Gilabert, "Computation and visualization of the MacAdam limits for any lightness, hue angle, and light source," J. Opt. Soc. Am. A **24**, 1501–1515 (2007).
32. P.-L. Sun and J. Morovic, "Inter-relating colour difference metrics," in *Tenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications* (Society for Imaging Science and Technology, 2002), pp. 55–60.
33. J. M. M. Linhares, P. D. Pinto, and S. M. C. Nascimento, "The number of discernible colors in natural scenes," J. Opt. Soc. Am. A **25**, 2918–2924 (2008).
34. M. Grendar, "Entropy and effective support size," Entropy **8**, 169–174 (2006).
35. I. Kontoyiannis and M. Madiman, "Sumset inequalities for differential entropy and mutual information," in *IEEE International Symposium on Information Theory (ISIT)* (IEEE, 2012).
36. A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," Phys. Rev. E **69**, 066138 (2004).
37. T. Schürmann, "Bias analysis in entropy estimation," J. Phys. A Math. Gen. **37**, L295–L301 (2004).
38. L. F. Kozachenko and N. N. Leonenko, "Sample estimate of the entropy of a random vector," Probl. Inf. Transm. (Tr. *Problemy Peredachi Informatsii*) **23**, 9–16, 95–101 (1987).
39. M. N. Goria, N. N. Leonenko, V. V. Mergel, and P. L. Novi Inverardi, "A new class of random vector entropy estimators and its applications in testing statistical hypotheses," J. Nonparametr. Stat. **17**, 277–297 (2005).
40. I. Marín-Franch and D. H. Foster, "Estimating information from image colors: an application to digital cameras and natural scenes," IEEE Trans. Pattern Anal. Mach. Intell. **35**, 78–91 (2013).
41. D. H. Foster, K. Amano, S. M. C. Nascimento, and M. J. Foster, "Frequency of metamerism in natural scenes," J. Opt. Soc. Am. A **23**, 2359–2372 (2006).
42. F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis, *Geometrical Considerations and Nomenclature for Reflectance* (Institute for Basic Standards, National Bureau of Standards, 1997).
43. I. Marín-Franch and D. H. Foster, "Number of perceptually distinct surface colors in natural scenes," J. Vis. **10**(9), 9 (2010).
44. M. R. Luo, C. J. Li, R. W. G. Hunt, B. Rigg, and K. J. Smith, "CMC 2002 colour inconstancy index: CMCCON02," Color. Technol. **119**, 280–285 (2003).
45. D. H. Foster, S. M. C. Nascimento, and K. Amano, "Information limits on identification of natural surfaces by apparent colour," Perception **34**, 1003–1008 (2005).
46. S. M. C. Nascimento, D. H. Foster, and K. Amano, "Psychophysical estimates of the number of spectral-reflectance basis functions needed to reproduce natural scenes," J. Opt. Soc. Am. A **22**, 1017–1022 (2005).
47. R. G. Kuehni, "Metamerism, exact and approximate," Color Res. Appl. **8**, 192 (1983).
48. H. Wang, M. R. Luo, G. Cui, and H. Xu, "A comparison between perceptibility and acceptability methods," in *11th Congress of the International Colour Association (AIC)*, D. Smith, P. Green-Armytage, M. A. Pope, and N. Harkness, eds. (AIC International Colour Association, 2009), pp. 1–7.
49. A. Akbarinia and K. Gegenfurtner, "Metameric mismatching in natural and artificial reflectances," J. Vis. **17**(10), 390 (2017).
50. G. Feng and D. H. Foster, "Predicting frequency of metamerism in natural scenes by entropy of colors," J. Opt. Soc. Am. A **29**, A200–A208 (2012).
51. D. H. Foster, "Estimating limits on colour vision performance in natural scenes," in *AIC Colour 2013, 12th Congress of the*

International Colour Association, L. MacDonald, S. Westland, and S. Wuerger, eds. (AIC International Colour Association, 2013), pp. 633–636.

52. D. I. Warton, I. J. Wright, D. S. Falster, and M. Westoby, "Bivariate line-fitting methods for allometry," Biol. Rev. **81**, 259–291 (2006).

53. D. R. Cox and E. J. Snell, *The Analysis of Binary Data*, 2nd ed. (Chapman and Hall/CRC, 1989).

54. D. H. Foster and G. Feng, "Visual and material identity in natural scenes: predicting how often indistinguishable surfaces become distinguishable," in *Predicting Perceptions: 3rd International Conference on Appearance* (Lulu Press, 2012), pp. 79–81.

55. S. Westland and C. Ripamonti, "Invariant cone-excitation ratios may predict transparency," J. Opt. Soc. Am. A **17**, 255–264 (2000).

56. J. Gert, "Color constancy, complexity, and counterfactual," Noûs **44**, 669–690 (2010).

57. F. Faul and V. Ekroll, "Transparent layer constancy," J. Vis. **12**(12), 7 (2012).

58. P. Olsson and A. Kelber, "Relative colour cues improve colour constancy in birds," J. Exp. Biol. **220**, 1797–1802 (2017).

59. D. H. Foster and S. M. C. Nascimento, "Relational colour constancy from invariant cone-excitation ratios," Proc. R. Soc. London Ser. B. Biol. Sci. **257**, 115–121 (1994).

60. S. M. C. Nascimento and D. H. Foster, "Relational color constancy in achromatic and isoluminant images," J. Opt. Soc. Am. A **17**, 225–231 (2000).

61. J. Liang, M. Georgoula, N. Zou, G. Cui, and M. R. Luo, "Colour difference evaluation using display colours," Lighting Res. Technol., 1–13 (2017).

62. J. A. Endler, "The color of light in forests and its implications," Ecol. Monogr. **63**, 1–27 (1993).