

Illuminant estimation as cue combination

Laurence T. Maloney

Psychology and Neural Science,
New York University, New York, NY, USA



This work briefly describes a model for illuminant estimation based on combination of candidate illuminant cues. Many of the research issues concerning cue combination in depth and shape perception translate well to the study of surface color perception. I describe and illustrate a particular experimental approach (perturbation analysis) employed in the study of depth and shape that is useful in determining whether hypothetical illuminant cues are actually used in color vision.

Keywords: Cue combination, surface color perception

Introduction

In the simple scene illustrated in Figure 1, there is a single light source, and light reaches the eye after being absorbed and reemitted by just one surface. We can express the excitations of photoreceptors at each location xy in the retina by the equation

$$\rho_k^{xy} = \int E(\lambda) S^{xy}(\lambda) R_k(\lambda) d\lambda, \quad k = 1, 2, 3, \quad (1)$$

where $S^{xy}(\lambda)$ is the surface spectral reflectance function of a surface patch imaged on retinal location xy , $E(\lambda)$ is the spectral power distribution of the light incident on the surface patch, and $R_k(\lambda)$, $k = 1, 2, 3$ are photoreceptor spectral sensitivities, all indexed by wavelength λ in the electromagnetic spectrum. A more realistic model of light flow in a scene would include the possibility of multiple light sources and inter-reflections between surfaces, and would take into account the orientation of surfaces. But in both the simple and the realistic models, the initial retinal information, the excitations of photoreceptors, depends on the spectral properties of both the illuminant and the surfaces present in a scene.

Illumination Estimation Hypothesis

Under some experimental conditions, human judgments of surface color are little affected by the spectral properties of the illuminant (see, in particular, Brainard, Brunt, & Speigle, 1997; Brainard, 1998). Although this constancy of perceived surface color has intrigued researchers for over a century, there is still no explanation of how human color visual processing effectively discounts the contribution of the illuminant in Equation 1. One hypothesis, originating with von Helmholtz (1962, p. 287), is that the human visual system estimates the chromaticity of the illuminant and then uses this estimate to discount the illuminant. The goal of this work is to investigate the theoretical and experimental issues involved in determining how the human visual system arrives at estimates of illuminant chromaticity. First, however, I will briefly describe psy-

chophysical results and computational work that supports the notion that the human visual system is engaged in illuminant estimation.

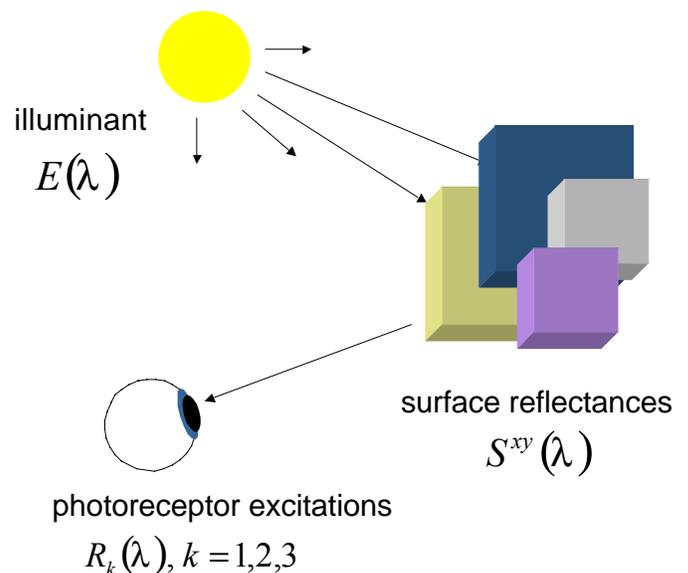


Figure 1. A simplified model of surface color perception. $S^{xy}(\lambda)$ is used to denote the surface spectral reflectance function of a surface patch imaged on retinal location xy . $E(\lambda)$ is the spectral power distribution of the light incident on the surface patch, and $R_k^{xy}(\lambda)$, $k = 1, 2, 3$ are the photoreceptor sensitivities, all indexed by wavelength λ in the electromagnetic spectrum.

Psychophysical Work

Brainard and colleagues (Brainard et al., 1997; Brainard, 1998) note that the patterns of errors in surface color estimation are those to be expected if the observer incorrectly estimates scene illumination and then discounts the illuminant using the incorrect estimate (“the equivalent illuminant” in their terms). Their observation supports the hypothesis that the observer is explicitly estimating the illuminant at each point of the scene. Mausfeld and colleagues (Mausfeld, 1997) ad-

vance the hypothesis that the visual system explicitly estimates illuminant and surface color at each point in a scene (the “dual code hypothesis”), and their empirical results support this claim.

Computational Work

In the last 20 years, researchers have sought to develop computational models of biologically plausible, color-constant visual systems (for reviews, see [Hurlbert, 1998](#); [Maloney, 1999](#)). Many of these algorithms share a common structure: first, the chromaticity of the illuminant (or equivalent information) is estimated. This illuminant estimate is then used in inverting [Equation 1](#) to obtain invariant surface color descriptors, typically by using a method developed by [Buchsbbaum \(1980\)](#). The algorithms differ from one another in how they estimate illuminant chromaticity, and it is reasonable to consider each algorithm as a potential cue to the illuminant present in a scene. There are currently algorithms that make use of surface specularity ([Lee, 1986](#); [D’Zmura & Lennie, 1986](#)), shadows ([D’Zmura, 1992](#)), mutual illumination ([Drew & Funt, 1990](#)), reference surfaces ([Brill, 1978](#); [Buchsbbaum, 1980](#)), subspace constraints ([Maloney & Wandell, 1986](#); [D’Zmura & Iverson, 1993a](#); [D’Zmura & Iverson, 1993b](#)), scene averages ([Buchsbbaum, 1980](#)), and more (see [Maloney, 1999](#)). An evident conclusion is that there are many potential cues to the illuminant in everyday, three-dimensional scenes, and it is of interest to consider the status of each of these algorithms as a possible component of a model of human color visual processing.

Cues to the Illuminant

Given that there are several possible cues to the illuminant, and that not all will provide accurate estimates of illuminant chromaticity in every scene, it is natural to consider illuminant estimation as a cue combination problem. This idea is not new. [Kaiser and Boynton \(1996, p. 521\)](#), for example, suggest that illuminant estimation is best thought of as combination of multiple illuminant cues. They leave unresolved several important theoretical and methodological problems surrounding cue combination. A theoretical issue, for example, is to develop a criterion for what counts as a possible illuminant cue. A methodological issue is how to determine experimentally that the human color vision system makes use of a particular cue.

The goal here is to describe a plausible framework for the study of illuminant cue combination in human surface color perception and to illustrate its use. The term framework is employed because the outcome is far from being a model of cue combination. The intent is to develop enough structure to allow us to translate basic questions about cue combination into experiments. As such, the assumptions made in developing the model may all be taken as provisional and open to empirical test. Their

purpose is to permit us to focus on devising experiments that tell us something useful about illuminant cue combination and help resolve the theoretical and methodological problems mentioned above.

[Maloney \(1999\)](#) contains a brief outline of these ideas. The result is analogous to a framework of depth and shape combination proposed by [Maloney and Landy \(1989\)](#) and [Landy, Maloney, Johnston, and Young \(1995\)](#).

Illuminant Cue Combination

Preliminaries

The goal is to estimate the illuminant chromaticity, $\rho^E = (\rho_1^E, \rho_2^E, \rho_3^E)$, defined as

$$\rho_k^E = \int E(\lambda) R_k(\lambda) d\lambda, \quad k = 1, 2, 3. \quad (2)$$

Illuminant chromaticity is the mean photoreceptor excitations for each class of photoreceptor when directly viewing the illuminant, and an obvious way to estimate illuminant chromaticity is to look directly at the light source(s) in a scene, a direct viewing cue ([Kaiser & Boynton, 1996](#)). An illuminant chromaticity estimate based on a direct viewing cue will be denoted by

$$\hat{\rho}^{DV} = (\hat{\rho}_1^{DV}, \hat{\rho}_2^{DV}, \hat{\rho}_3^{DV}).$$

We do not yet know whether a direct viewing cue is employed in human vision. In order for such a cue to provide accurate estimates of surface color in complex scenes, the visual system must work out which light sources illuminate which surfaces, a potentially difficult problem. The results of [Bloj, Kersten, and Hurlbert \(1999\)](#) do indicate that the visual system has some representation of how light flows from surface to surface in a three-dimensional scene.

Illuminant Cues

If a visual system cannot obtain a direct view of the light sources, then it must develop an estimate, $\hat{\rho}^E = (\hat{\rho}_1^E, \hat{\rho}_2^E, \hat{\rho}_3^E)$, of these parameters indirectly. The various algorithms above are methods for computing an estimate $\hat{\rho}^E = (\hat{\rho}_1^E, \hat{\rho}_2^E, \hat{\rho}_3^E)$ when certain assumptions about the scene are satisfied. I will restrict the use of the term ‘illuminant cue to algorithms that result in a point estimate of the chromaticity of the illuminant (I will weaken this constraint slightly in the section entitled “Promotion” below). Any illuminant cue in this sense can, in isolation, provide the information needed to discount the illuminant. There may be, of course, other sensory and nonsensory sources of information that potentially provide information about the illuminant in a scene. These sources of information do play a role in the framework developed here, but not as illuminant cues. I will return to this point below. This restriction ignores hypothetical cues that provide only ordinal or categorical information about illuminant chromaticity and may

prove to be an oversimplification if the human visual system makes use of such cues.

In this work, I will describe experimental tests of a candidate cue based on specularity, one I refer to as the *specular highlight* cue. There are other computational cues to the illuminant based on surface specularity (see Yang & Maloney, 2001, and Maloney & Yang, 2002) but consideration of this one will suffice for my purposes here. The illuminant estimates based on a specular highlight illuminant cue will be denoted $\hat{\rho}^{SH} = (\hat{\rho}_1^{SH}, \hat{\rho}_2^{SH}, \hat{\rho}_3^{SH})$, and it is the average of the chromaticities of regions of the scene corresponding to specular highlights. Evidently, the hard part of developing an explicit algorithm for estimation of this cue is the identification of the parts of the retinal images that correspond to true specular highlights in the scene.

The light reflected from a specular highlight can signal not only the chromaticity of the illuminant but also the surface material under the highlight. But, if we are certain that the light from a particular specular highlight has (almost) the same chromaticity as the light source, then we would accept the photoreceptor excitations of the highlight as a useful estimate of ρ^E , the illuminant chromaticity.

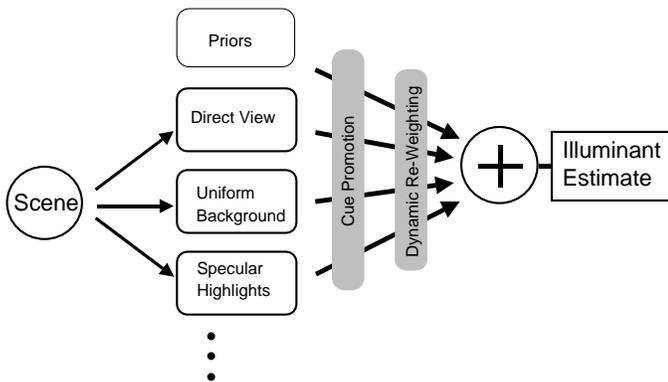


Figure 2. Illuminant cue combination. In the illuminant cue combination model of Maloney (1999), distinct illuminant cues are extracted from the scene via illuminant estimation modules, analogous to depth modules in depth perception. The different sources of information concerning the illuminant are promoted to a common format (see text) and then combined by a weighted average whose weights may vary from scene to scene as the availability and quality of illuminant cues vary. Nonsensory prior information is represented in the diagram but not further discussed here.

A third candidate illuminant cue is the chromaticity of the uniform background $\hat{\rho}^{UB} = (\hat{\rho}_1^{UB}, \hat{\rho}_2^{UB}, \hat{\rho}_3^{UB})$ when one is present in the scene. This cue would only be an accurate cue to the illuminant when the chromaticity of the light absorbed and reemitted by the background is that of the illuminant, an assumption closely related to the Grayworld Assumption (Buchsbaum, 1980). Computing this cue presents no obvious challenges beyond identi-

fying the parts of the scene that belong to the uniform background.

Several more illuminant cues, taken from the computational literature, are defined and discussed in Maloney (1999). The three just introduced are all we need to discuss the illuminant cue combination framework introduced next. In listing these candidate cues, I do not mean to imply that they are known to play any role in human color visual processing. Rather, by formalizing their role in an explicit cue combination framework, we will be in a position to test whether any of them act as a cue to the illuminant in human color vision.

Illuminant Cue Combination

Figure 2 contains a diagram illustrating the cue combination process. Explicit cues to the illuminant are derived from the visual scene and, eventually, combined by a weighted average after two intervening stages labeled promotion and dynamic reweighting and explained below. The weighted average can be written as

$$\hat{\rho}^E = \alpha_{DV} \hat{\rho}^{DV} + \alpha_{SH} \hat{\rho}^{SH} + \alpha_{UB} \hat{\rho}^{UB} + \dots \quad (3)$$

The α s are nonnegative scalar weights that sum to 1, and they can be interpreted as a measure of the importance of each of the cues in the estimation process. The cue estimates shown correspond to the hypothetical cues discussed above: direct viewing (DV), specular highlights (SH), and uniform background (UB).

In order to apply Equation 3, the visual system needs to solve two distinct and complementary problems. The first is to determine the estimates available from each of the individual illuminant cues (cue estimation). The computational models discussed previously are models of this process. The second is to assess the relative importance of each cue in a given scene and assign appropriate weights (cue weighting). This second problem has been studied intensively only in the last 15 years (see discussions in Landy et al., 1995 and Yuille & Bülthoff, 1996), and it is in essence a statistical problem (e.g., see Geisler, 1989; Knill & Richards, 1996; Rao, Olshausen, & Lewicki, 2002). This second cue-weighting problem is of central concern here: How does the visual system assign weights in Equation 3? I will refer to algorithms that assign weights as *rules of combination*. There are many possible rules of combination, some of which are optimal by statistical criteria and some of which are not. We will soon see an example of an optimal rule that assigns weights according to the reliability of each of the cue estimates.

Rules of Combination

As a first example of a rule of combination, consider a *hierarchical rule* that assigns the three cues to positions in a hierarchy, $DV > SH > UB$. The rule of combination must

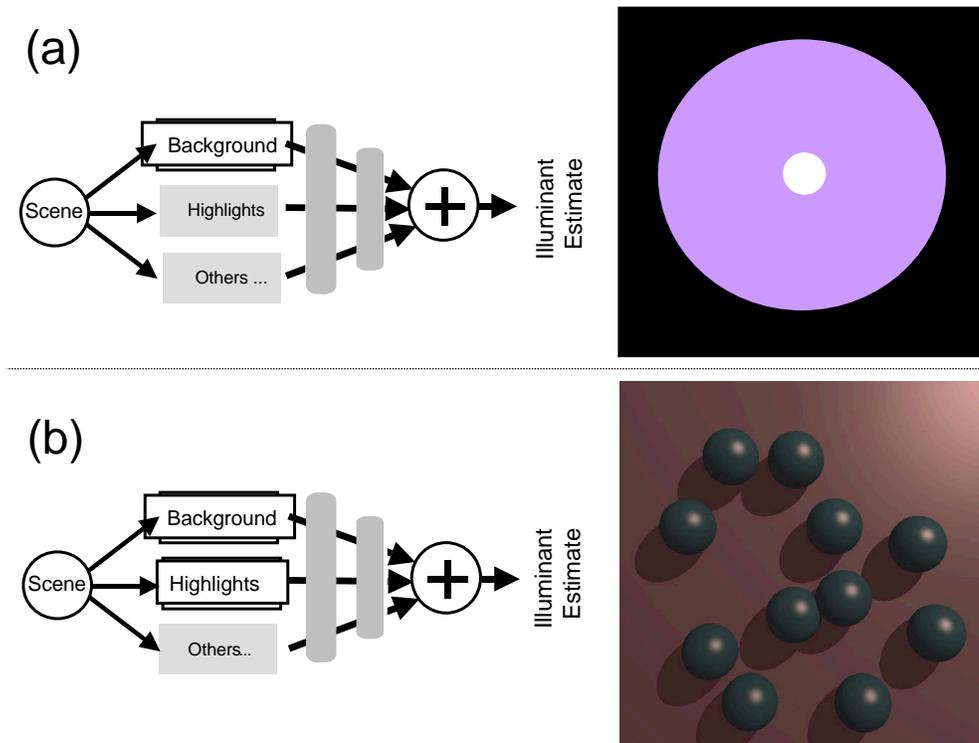


Figure 3. Dynamic reweighting. (a) A scene with only one illuminant cue (uniform background). (b) A scene with two illuminant cues, the second based on surface specularity.

first classify each cue as present or absent from a scene, and then pick the first cue in the hierarchy that is present. If the direct viewing cue is available, the visual system will use it exclusively. If the visual system judges that the light source is not visible (the DV cue is absent) and there are specular highlights available in the scene, then it will use the specular highlight cue exclusively, and so on, down the hierarchy. This rule is characterized by weights that are always 0 or 1, with exactly one weight set to 1.

A different rule of combination (*minimum variance rule*) treats the individual cues as independent trivariate Gaussian random variables $G_i \sim N(\gamma, \sigma_i^2 \Sigma)$ with a common mean γ , a common covariance matrix Σ scaled by factors σ_i^2 specific to each cue. The statistical estimator of γ that is unbiased and that has minimum total variance¹ is of the form of Equation 3 (a weighted-linear combination). The weights are functions of the covariance matrices. This is a generalization to the trivariate case of the univariate result that the choice of weights that minimizes the variance of the estimate of γ are inversely proportional to the variances of the corresponding cues (Cochran, 1937). This same univariate rule satisfies other statistical criteria of optimality: it is the maximum likelihood estimator and also the MAP estimator (Yuille & Bülthoff, 1996).

A third rule of combination takes into account the covariances of the individual illuminant cues and then assigns a weight of 1 to the cue with the lowest total variance and a weight of 0 to other rules. This *best cue rule*

selects the most reliable (as measured by total variance) cue and ignores the others, a sort of winner-take-all algorithm for cue combination.

The last two rules of combination require information about the covariance of illuminant cues. I mentioned above that there are other sources of potential source information in scenes that are not illuminant cues. For example, information that permits estimation of the covariance of illumination cues falls into this category, and Maloney and Landy (1989) refer to such sources of information as “ancillary measures.”

There are many possible rules of combination, some but not all consistent with the weighted linear rule of Equation 3. In order to discriminate among possible rules of combination, we need to be able to estimate the weights assigned to each cue experimentally. More generally, we can frame hypotheses about cue combination in terms of the values of the weights. If, for example, the direct viewing cue is never used in human vision, then $\alpha_{DV} = 0$ for all scenes. Experimental tests of the hypothesis $\alpha_{DV} = 0$ and similar hypotheses for other cues serve as a formalism that allows us to decide that a cue is used in human vision ($\alpha_{DV} > 0$) at least under some circumstances.

The linear rule in Equation 3 is provisional. The rule of combination employed by the visual system may be distinctly nonlinear. However, the weighted linear combination rule has proven to be a useful basis for investigation of cue combination in depth and shape vision (e.g.,

the elegant results of Ernst & Banks, 2002). In effect, researchers can frame hypotheses about cue combination in terms of weights in Equation 3, and then test these hypotheses experimentally by measuring the weights. Before describing how that can be done, I need to say a bit about dynamic reweighting and promotion.

Dynamic Reweighting

There may be no shadows, no specularity, or no mutual illumination between objects in any specific scene. The illuminant may be in the current visual field (directly viewable), or not. In the psychophysical laboratory, we can guarantee that any or all of the cues above are absent or present as we choose. If human color vision made use of only one cue to the illuminant, then when that cue was present in a scene, we would expect a high degree of color constancy, and when that cue was absent, a catastrophic failure of color constancy. Based on past research, it seems unlikely that there is any single cue whose presence or absence determines whether color vision is color constant. An implication for surface color perception is that the human visual system may make use of multiple cues and different cues in different scenes. The relative weight assigned to different estimates of the illuminant from different cue types may, therefore, change. Landy et al. (1995) report empirical tests of this claim, which imply that depth cue weights do change in readily interpretable ways.

In particular, consider the sort of experiment where almost all cues to the illuminant are missing. The observer views a large, uniform surround (Figure 3A) with a single test region superimposed. The observer will set the apparent color of the test region under instruction from the experimenter, and it is plausible that the only cue to the illuminant available is the uniform chromaticity of the surround. In very simple scenes, observers behave as if the chromaticity of the surround were the chromaticity of the illuminant (for discussion, see Maloney, 1999). An intelligent choice of weights for the scene of Figure 3A is $\alpha_{DV} = \alpha_{SH} = 0$ and $\alpha_{UB} = 1$.

Consider, in contrast, the more complicated scene in Figure 3B. There is still a large, uniform background, but there are other potential cues to the illuminant as well, notably the specular highlights on the small spheres. Will the observer continue to use only the chromaticity of the uniform background, or will he also make use of the chromaticity of the specular highlights? Will the influence of the uniform background on color appearance decrease when a second cue is available? Will α_{SH} be greater than 0 and α_{UB} less than 1?

Cue Promotion

A second and surprising analogy between depth cue combination and illuminant estimation is that not all cues to the illuminant provide full information about the

illuminant parameters $\rho^E = (\rho_1^E, \rho_2^E, \rho_3^E)$. Some of the methods lead to estimates of ρ^E up to an unknown multiplicative scale factor (e.g., D'Zmura & Lennie, 1986; for review, see Maloney, 1999). The same is, of course, true of depth cue combination where certain depth cues (such as relative size) provide depth information up to an unknown multiplicative scale factor. By analogy with Maloney and Landy (1989), I refer to cues such as illuminant cues with a missing parameter. A cue that provides an estimate of ρ^E up to an unknown scale factor is an illuminant cue missing one parameter, the scale factor. If the missing parameter or parameters can be estimated from other sources, the illuminant cue with parameters can be promoted to an estimate of the illuminant parameters, ρ^E . The problem of combining depth cues, some of which have missing parameters, is termed "cue promotion" by Maloney and Landy and is treated further by Landy et al. (1995). Here we will not be further concerned with cue promotion and will assume that all cues have been promoted.

As an aside, consider that color constancy can be very good in some scenes (Brainard et al., 1997; Brainard, 1998; Kraft & Brainard, 1999) and almost nonexistent in others (Helson & Judd, 1936). A recent special issue of *Perception* was devoted to investigating why the constancy of surface color perception varies from scene to scene (Maloney & Schirillo, 2002). The answer I propose, in the spirit of the cue combination model presented here, is that some scenes are rich in accurate illuminant cues, and the visual system makes use of them, leading to accurate estimates of illuminant chromaticity and a high degree of color constancy. Other scenes, including the sort of scene represented in Figure 3A, contain few cues to the illuminant, and we would not expect that the visual system could arrive at accurate estimates of illuminant chromaticity or surface color.

Many of the algorithms described by Maloney (1999) can be identified with potential cues to the illuminant as noted above. What the cues to the illuminant employed in human vision are and how they are combined remain open questions. In the following sections of this work, I describe experimental methods taken from Yang and Maloney (2001) that allow one to measure which illuminant cues are influencing human surface color perception.

Weight Estimation

We measured the influence of each of the two candidate cues to the illuminant using a cue perturbation approach analogous to that described by Maloney and Landy (1989) and Landy et al. (1995). The perturbation approach has the advantage that we can test whether a cue is in use in a given scene without large alternations to the scene that might trigger other unanticipated changes in visual processing.

The key idea underlying the approach is easily explained. We would like to alter the illuminant informa-

tion signaled by specularly while holding everything else in the scene constant. If this perturbation affects perceived surface color, we have evidence that the cue is being used by the visual system, and the magnitude of the effect, compared to the magnitude of the perturbation, allows us to quantify the influence of the cue in a particular scene. We next describe in more detail how to perturb illuminant cues and measure their influences when the dependent measure is an achromatic setting.

First, we create scenes where multiple candidate cues to the illuminant are available. We measure the observer's achromatic setting for two different illuminants (illuminants I_1 and I_2) applied to the scene. These achromatic settings are plotted in a standard color space as shown in Figure 4A, marked I_1 and I_2 . The direction and magnitude of any observer change in achromatic setting in response to changes in the illuminant are useful measures of the observer's degree of color constancy, but that is not of immediate concern to us. We are content to discover that the chromaticity of the surface the observer considers to be achromatic changes when we change the illuminant, presumably because of information about the illuminant signaled by illuminant cues available to the observer. However, so far, we can conclude nothing about the relative importance of any of the illuminant cues present, because all signal precisely the same illuminant in both rendered scenes.

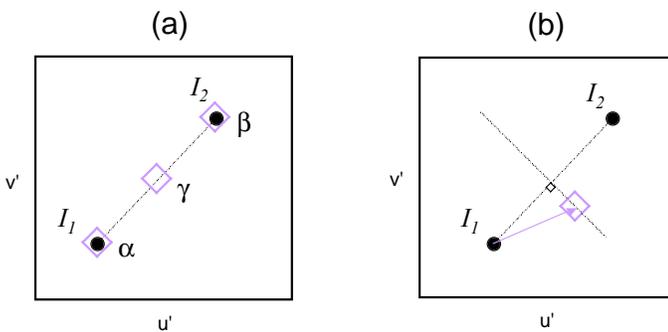


Figure 4. Hypothetical data from a perturbation experiment. (a) The point marked I_1 is the achromatic setting of a hypothetical observer when the test patch is embedded in a scene illuminated by reference illuminant 1. The point marked I_2 , is, similarly, the achromatic setting when the same scene is illuminated by reference illuminant 2. The remaining points correspond to hypothetical achromatic settings when one illuminant cue signals I_2 , and the remainder signal I_1 . The setting α is consistent with the assertion that the perturbed cue has no effect. The setting β is consistent with the assertion that the perturbed cue is the only cue that has some influence. The setting γ is consistent with an influence of 0.5 as it falls at the midpoint of the line joining A and D65. (b) Hypothetical results that include the possibility that the observer's settings are perturbed by noise. The three estimates will not, in general, be collinear.

We next ask the observer to make a third achromatic setting in a scene where the illuminant information for one cue is set to signal illuminant I_2 , while all other cues are set to signal illuminant I_1 (this sort of cue manipulation is not difficult with simulated scenes, but would be difficult to do in a real scene). The model that Joong Nam Yang and I used in rendering all of the objects used as stimuli is that of Shafer (1985).²

The experimental data we now have are composed of three achromatic settings: under illuminant I_1 , under illuminant I_2 , and under illuminant I_1 with one cue perturbed to signal illuminant I_2 . We wish to determine whether the visual system is paying attention to the perturbed cue, that is, whether the perturbed cue has a measurable influence on color perception as measured by achromatic matching.

What might happen? One possibility is that the observer's setting in the scene with one cue perturbed to signal illuminant I_2 is at the point labeled α in Figure 4A, identical to the setting that he or she chose when all cues signaled illuminant I_1 . We would conclude that the perturbed cue had no effect whatsoever on surface color perception: It is not a cue to the illuminant, at least in the scene we are considering.

Suppose, on the other hand, the observer's achromatic setting in the scene with one cue perturbed to signal illuminant I_2 (and all others are set to signal illuminant I_1) is at the point marked β in Figure 4A, the same as it was when all cues signaled illuminant I_2 . This would suggest that the observer is using only the manipulated cue, and ignoring the others. A third possibility is that the observer chooses a setting somewhere between his or her settings for the two illuminants (point γ in Figure 4A), along the line joining them. Let δ be the change in setting when only the perturbed cue signals illuminant 2 (the distance from I_1 to γ) and let Δ be the change in setting when all cues signal illuminant 2 (the distance from I_1 to I_2). We define the influence of the perturbed cue to be

$$I = \frac{\delta}{\Delta} = \frac{\| \gamma - I_1 \|}{\| I_2 - I_1 \|} \quad (4)$$

The value I should fall between 0 and 1. A value of 0 implies that the perturbed cue is not used (point α); a value of 1 implies that only the perturbed cue is used (point β). Point γ corresponds to an influence of 0.5 as it falls at the midpoint of the line joining A to B. It is easy to show (Maloney & Landy, 1989; Landy et al., 1995) that the influence of a cue is precisely the weight assigned to it in Equation 3, or, allowing for measurement error, the measured influence of a cue is an estimate of the weight assigned to the cue. The empirical procedure just described allows us to estimate the weights in Equation 1.

In the perturbed scenes, the observer is free to make achromatic settings that do not fall on the line joining the settings in the two unperturbed scenes. We expect such an outcome, if only as a consequence of measurement

error. The computation of influence we actually employ is described in more detail in [Yang and Maloney \(2001\)](#) and [Brainard \(1998\)](#) and is illustrated in Figure 4B. In essence, we use the nearest point on the line segment in computing influence in [Equation 4](#) above. Note that if we can demonstrate that the deviations of observers' settings are not the result of measurement error, then we would reject the hypothesis that the weighted linear combination rule of [Equation 3](#) correctly describes human illuminant cue combination.

A critical factor in illuminant estimation studies such as those described here is that the images that are displayed on a computer monitor must be rendered correctly. Human color constancy with simulated images is markedly less than that obtained with real scenes ([Arend et al., 1991](#); [Brainard, 1998](#); [Kurichi & Uchikawa, 1998](#)). With real scenes, the index reaches an average of 0.84 ([Brainard, 1998](#)), while the values achieved with scenes presented on computer monitors are typically less than 0.5. In [Yang and Maloney \(2001\)](#), we took several steps to ensure that the scenes we present are as accurate as possible and achieved an index of 0.65, intermediate between previous research with computer monitors and with real scenes. In describing the apparatus, we will touch on some of them.

An Illustrative Experiment

Apparatus

[Yang and Maloney \(2001\)](#) built a large, high-resolution stereoscopic display (Figure 5). The observer sat at the open side of a large box, positioned in a chin rest, gazing into the box. Its interior was lined with black feltlike paper. Small mirrors directly in front of the observer's eyes permitted him or her to fuse the left and right images of a stereo pair displayed on computer monitors positioned to either side.

An example of a stimulus (image pair) is shown in Figure 6. Once an image was displayed, the observer pressed keys that altered the color of a small test patch until it appeared achromatic. The observer could adjust the color of the patch in two dimensions of color space but could not change its luminance.

We used the physics-based rendering package RADIANCE ([Larson & Shakespeare, 1997](#)) to render each of the images in a stereo pair, simulating the appearance of a spheres tangent to a plane perpendicular to the observer's line of sight, as shown in Figure 6. The objects within the scene were rendered as if they were roughly the same distance in front of the observer as the optical distance from each of the observer's eyes to the corresponding display screen (70 cm).

The matte component of each rendered surface (background, spheres) was rendered so as to match it to a particular Munsell color reference chip from the Nickerson-Munsell collection ([Kelley, Gibson, & Nickerson, 1943](#)). Computer graphics rendering does not correctly

model the spectral effects of light-surface interaction ([Maloney, 1999](#)). We modified the rendering package to correct this problem as described in [Yang and Maloney \(2001\)](#). The entire scene was illuminated by a combination of a punctate and a diffuse light. The spectral power distribution of the diffuse light was always that of either standard illuminant D65 ([Wyszecki & Stiles, 1982, p. 8](#)). The punctate illuminant was always positioned behind, to the right of and above the observer in the rendered scene. The square test patch (0.5 deg of visual angle on a side) was tangent to the front surface of one of the spheres.

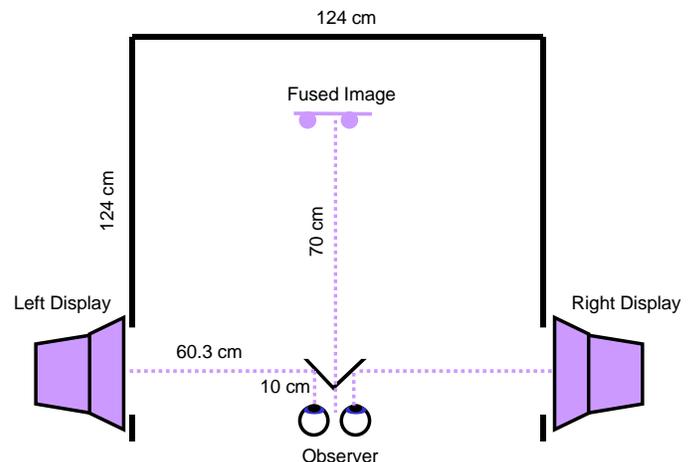


Figure 5. The experimental apparatus. Stimuli were presented in a computer-controlled Wheatstone stereoscope. Two monitors were used to present the images of a stereo image pair to the observer's left and right eyes. Two computers controlled the monitors and a third computer coordinated the presentation of stimuli and recorded the observer settings in an achromatic matching task.

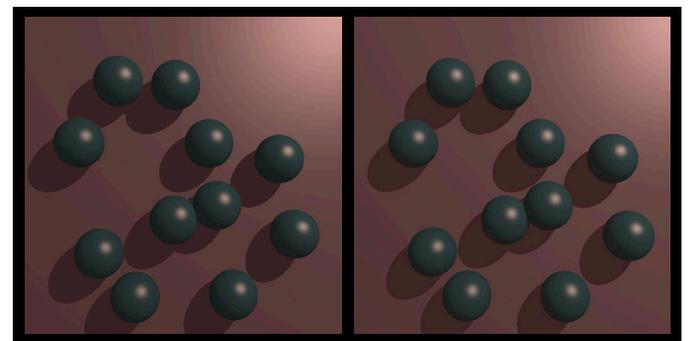


Figure 6. An example of a stimulus (binocular image pair). The figure shows a stereo image pair (for crossed fusion) similar to those employed in the experiments.

The methods used to effect perturbations of the illuminant chromaticity of the specular highlight cue are complicated and are described in detail in [Yang and Maloney \(2001\)](#). When the specularity cue in a scene rendered under the nearly neutral illuminant D65 was altered to signal Illuminant A, the pixels forming the specu-

lar highlights in images reddened with little or no apparent change anywhere else in the images. The number of pixels altered in perturbing a cue was small and the effect on average scene chromaticity was negligible.

Results

Yang and Maloney (2001) studied surface color perception in scenes made up of spheres placed against a uniform background surface. The spheres were highly specular, the background slightly specular, and the matte components of all of the spheres were homogeneous and identical. The Munsell coordinates for the matte components of each sphere were BG 5/4, and for the matte component of the background, N 3/ (Kelley et al., 1943). One of our stimuli is shown in Figure 6. In this section, I summarize the results of the first experiment in Yang and Maloney. The goal of this experiment was to determine whether the visual system makes use of the specular highlights on the spheres as a cue to the illuminant, using the perturbation method just described.

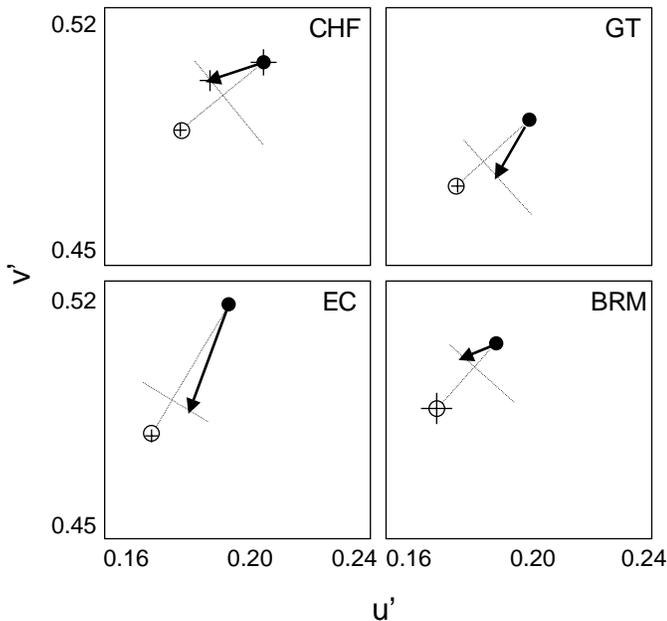


Figure 7. Specular Illuminant cues: results of Experiment 1. The achromatic settings for four observers are shown, plotted in $u'v'$ coordinates in CIE chromaticity space. In each small plot, an open circle marks the mean of multiple settings by one observer for the illuminant D65 consistent-cue condition, a filled circle marks the mean for multiple settings by the same observer for the illuminant A consistent-cue condition, and the center of the head of the vector marks the mean of multiple settings for the perturbed-cue condition. The base of the vector is connected to the consistent cue setting corresponding to the illuminant signaled by the nonperturbed cues. Horizontal and vertical bars indicate one SE for each setting. The projection of the perturbed setting onto the line joining the unperturbed settings is marked. The perturbed cue signaled D65, all others signaled A. Taken from Yang and Maloney (2001).

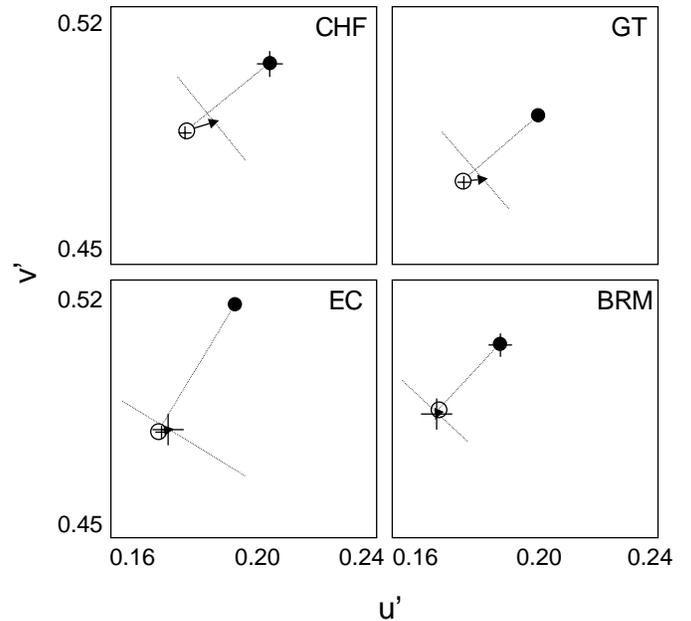


Figure 8. Specular illuminant cues: results of Experiment 1. The format is identical to that of the previous plot. The perturbed cue signaled A, all others signaled D65. Taken from Yang and Maloney (2001).

There were two perturbation conditions in the experiment. In the first, all cues except for the specular highlight cue signaled illuminant A while the specular highlight cue signaled illuminant D65. In the second, all cues except the specular highlight cue signaled D65 and the specular highlight cue signaled A. Figure 7 contains the results for four observers in the first perturbation condition, Figure 7 contains the results for the same four observers in the second perturbation condition. In each small plot in Figure 7 and in Figure 8, the horizontal and vertical axes are the u' and v' coordinates of the CIE chromaticity diagram in the same format as the hypothetical data of Figure 4B. The open circle in each small plot corresponds to the observer's mean achromatic setting when the scene was rendered under illuminant A; the filled circle corresponds to the mean achromatic setting under illuminant D65. In the four plots in Figure 7, the tip of the arrow corresponds to the observer's mean achromatic setting when the specular highlight cue signaled illuminant D65 while all other cues signaled illuminant A. Figure 8 shows the effect of perturbing the specular highlight cue toward A, when all of the other illuminant cues signal D65.

Each observer's setting in the unperturbed condition for illuminant D65 (open circle) is evidently different from his setting in the unperturbed condition for illuminant A (filled circle). The observers are responding to changes in the illuminant, and the direction and magnitude of response are similar to those found in previous studies (e.g., Arend, Reeves, Shirillo, & Goldstein, 1991; Brainard, 1998).

Note that the influence is asymmetric, in that the cue perturbation from illuminant A in the direction of illuminant D65 has a much greater influence than that from illuminant D65 in the direction of illuminant A. For the former settings, specular information had significant influence on achromatic settings: The measured influence ranged from 0.3 to 0.83.

We repeated this experiment with a different choice of Munsell surface for the objects and the background. (10GY 5/6 for the objects and 10P 4/6 for the background). When the colors of the objects and background were altered, the achromatic settings changed little, consistent with results reported in previous studies (Brainard, 1998; Kurichi & Uchikawa, 1998). The effect of perturbation changed very little as well, and there was still a marked asymmetry in the effect of perturbation between the illuminant conditions. The outcome of this experiment indicates that the illuminant information conveyed by specular information can affect the apparent colors of surfaces in a scene.

Dynamic Reweighting Revisited

The stimuli shown in Figure 6 contain 11 spheres, each with a single specular highlight. Yang and Maloney (2001) investigated the effect of changing the number of specular highlights in the scene. We repeated only the perturbation condition where perturbations in the illuminant signaled by the specular highlight cue did influence achromatic settings (Figure 7). We found that with 1, 2, or 6 spheres, there was no statistically significant effect of perturbation but that with 9 or 11 spheres, there was an effect of perturbation. The measured influence with 9 spheres was approximately 0.25, with 11 spheres, 0.5. These results suggest that the visual system is assigning different weights to the specular highlight cue, depending on the number (or possibly the density) of specular highlights available in the scene. These results are consistent with those of Hurlbert (1989), who found that the specular highlight cue had little effect on surface color appearance in scenes containing only one sphere and its specular highlight.

Discussion

The results reported here, together with previous research, indicate that there are at least two cues to the illuminant active in human vision. The first, the uniform background cue or perhaps average background cue, is known to affect surface color perception in very simple scenes, as described above. The results just described suggest that there is a second cue, present when specular highlights are present. Of course, given the empirical results, it is natural to propose alternative illuminant cues (algorithms) that could also account for the results of Yang and Maloney (2001), and then to devise experiments that discriminate among them. Yang and Maloney,

for example, tested a second algorithm (cue) based on specular information due to D'Zmura and Lennie (1986) and Lee (1986). They found that this cue did influence achromatic settings.

Our results suggest that the influence of this cue varies with the number of specular objects present in the scene (or alternatively, with the density of specular objects). This result is consistent with the claim that the weights given to different illuminant cues can change (dynamic reweighting). A plausible role for dynamic reweighting in Equation 3 is to reduce or eliminate the contribution of illuminant cues that do not provide reliable estimates of illuminant chromaticity in particular scenes. Of course, the visual system can adjust the weight assigned to a cue to reflect its reliability only if it has some method of assessing cue reliability. The number, density, location, and size of specular highlights are all possibly employed in assessing the reliability of the specular highlight cue. Determining the rule that the visual system uses to assign weights to the specular highlight cue and other illuminant cues is evidently important.

Dynamic reweighting has implications for experimental method. In a series of experiments, Brainard and colleagues investigated the effect of particular illuminant cues in a series of experiments where they added or removed cues from real (not virtual) scenes (Kraft & Brainard, 1999; Kraft, Maloney, & Brainard, 2002; Brainard, Kraft, & Longère, in press). For example, they added a highly specular cylinder to a scene or removed it. In cue-rich scenes, they found that adding or subtracting cues had little effect. This outcome is what would be expected with appropriate dynamic reweighting. If each illuminant cue is an unbiased estimate of the illuminant chromaticity, then any weighted linear mixture of cues with weights that sum to 1 is also an unbiased estimate of illuminant chromaticity. If the visual system sets the weight that corresponds to a deleted cue to 0 and renormalizes the remaining weights to sum to 1, then the expected value of the estimate would be unchanged. Adding or deleting cues would not be expected to affect the expected value of the illuminant chromaticity estimate and color appearance should be little affected.³

In other scenes, containing few illuminant cues, they found that removing cues typically reduces an index of color constancy that they used to summarize each observer's performance. There is no ready explanation for this result in terms of Equation 3. The key challenge arising from their results is to understand why, in some cases, the measured index of color constancy changed and what this tells us about illuminant cue combination. These results hint that the visual system has a default or prior assumption concerning illuminant chromaticity that manifests itself when the illuminant cues available in the scene are judged to be unreliable. It would be very natural to model such prior information within a Bayesian framework (see Yuille & Bülthoff, 1996; Mamassian, Landy, & Maloney, 2002).

The asymmetry observed in the first experiment of [Maloney and Yang \(2002\)](#) is intriguing. Possibly the visual system gives very little weight to illuminant cues that are far from neutral. The visual system may be organized so as to discard specularities that are intensely colored simply to avoid errors introduced by nonneutral specular surfaces (e.g., gold). That is, a specularly signaling a neutral D65 illuminant is given much higher weight than a specularly signaling a reddish illuminant A, leading to the observer asymmetry. If so, then a replication of Experiment 1 with smaller perturbations away from illuminant D65 may disclose some effect of the specular highlight cue. This outcome would reject the simple weighted linear model of [Equation 3](#).

Alternatively, it is possible that specularly cues that signal changes toward a neutral point are assigned greater weight than those that signal changes in other directions in the space. This would also account for the observed asymmetry in Experiment 1 of [Yang and Maloney \(2001\)](#). We could test this possibility by repeating the experiment of Yang and Maloney but using pairs of lights placed symmetrically around a neutral point in illuminant chromaticity space or that fall at different points along a radius leading from a neutral point to illuminant A.

The framework is, as I noted earlier, provisional. It serves two purposes. The first is to provide a natural way to frame hypotheses about cue combination in terms of the weights assigned to cues in [Equation 3](#). The second is to permit estimation of these same weights experimentally. Once we do so, we may discover that the pattern of results leads us to reject the model in [Equation 3](#). We may discover that weights are negative or that the mean perturbed setting in the diagram of [Figure 4B](#) falls so far from the line segment that we can reject the weighted linear model. [Maloney and Landy \(1989\)](#) and [Landy et al. \(1995\)](#) interpreted the linear rule as valid for only small perturbations in depth and shape vision, and assumed that large discrepancies between cues might lead to suppression of some cues at the expense of others (they refer to this issue as “robustness”). This may prove to be the case in illuminant cue combination as well, but that is an empirical question. Disproving this model or failing to disprove will, in either case, tell us something about illuminant cue combination.

It is also interesting to consider how these experiments highlight certain unspoken assumptions in the study of depth, shape, and color. In [Figure 6](#), each sphere and even the background exhibit a wide range of discriminable colors in both of the stereo images, even though each is made of a single surface material. The stimulus can be described parsimoniously in terms of surfaces and illuminants and their relative locations, in essentially something like the graphical language we employed in specifying the scenes to the rendering package we used. The resulting pair of retinal images is (superficially) much more complex. Shading, shadows,

inter-reflections, specularity, and the like have conspired to produce very complex stimuli, if we insist on describing them retinally. If, however, we wish to study surface color perception, the estimation of objective surface properties through human color vision, then it would make sense to describe the stimuli and their manipulation in terms of the environment, and not in terms of an arbitrary, intermediate, retinal stage in color processing.

Acknowledgments

Preparation of this article was supported in part by grant EY08266 from the National Institute of Health, National Eye Institute, and by Human Frontiers Science Program Grant RG0109/1999-B. The work is based on a talk prepared for the Color & Vision Satellite Meeting of the 2001 Optical Society Meeting at the University of California, Irvine. It summarizes material taken from [Maloney \(1999\)](#), [Yang and Maloney \(2001\)](#), and [Maloney and Yang \(2002\)](#). The author is particularly indebted to Joong Nam Yang. Commercial relationships: none.

Footnotes

¹The total variance is the trace of the 3x3 covariance matrix.

²The Shafer model is inaccurate as a description of certain naturally occurring surfaces ([Lee, Breneman, & Shulte, 1990](#)) but it not known how well it approximates surfaces in the everyday environment. It is, however, an accurate approximation of a large class of surfaces known as dielectrics, that includes plastics.

³The loss of the cue may be reflected in the observer's setting variability but not his mean setting.

References

- Arend, L. E., Reeves, A., Shirillo, J., & Goldstein, R. (1991). Simultaneous color constancy: Papers with diverse Munsell values. *Journal of the Optical Society of America A*, 8, 661-672. [[PubMed](#)]
- Bloj, M., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences color perception through mutual illumination. *Nature*, 402, 877-879. [[PubMed](#)]
- Brainard, D. H. (1998). Color constancy in the nearly natural image. 2. Achromatic loci. *Journal of the Optical Society of America A*, 15, 307-325. [[PubMed](#)]
- Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image. 1. Asymmetric matches. *Journal of the Optical Society of America A*, 14, 2091-2110. [[PubMed](#)]

- Brainard, D. H., Kraft, J. M., & Longère, P. (in press). Color constancy: Developing empirical tests of computational models. In R. Mausfeld & D. Heyer (Eds.), *Colour perception: From light to object*. Oxford, UK: Oxford University Press.
- Brill, M. H. (1978). A device performing illuminant-invariant assessment of chromatic relations. *Journal of Theoretical Biology*, *71*, 473-476. [PubMed]
- Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute*, *310*, 1-26.
- Cochran, W. G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society*, *4*(Suppl. 1), 102-118.
- Drew, M. S., & Funt, B. V. (1990). Calculating surface reflectance using a single-bounce model of mutual reflection. *Proceedings of the Third International Conference on Computer Vision, Osaka, Japan, December 4-7, 1990* (pp. 394-399). Washington, DC: IEEE Computer Society.
- D'Zmura, M. (1992). Color constancy: Surface color from changing illumination. *Journal of the Optical Society of America A*, *9*, 490-493.
- D'Zmura, M., & Iverson, G. (1993a). Color constancy. I. Basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces. *Journal of the Optical Society of America A*, *10*, 2148-2165. [PubMed]
- D'Zmura, M., & Iverson, G. (1993b). Color constancy. II. Results for two-stage linear recovery of spectral descriptions for lights and surfaces. *Journal of the Optical Society of America A*, *10*, 2166-2180. [PubMed]
- D'Zmura, M., & Lennie, P. (1986). Mechanisms of color constancy. *Journal of the Optical Society of America A*, *3*, 1662-1672. [PubMed]
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429-433. [PubMed]
- Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discrimination. *Psychological Review*, *96*, 1-71.
- Helson, H., & Judd, D. B. (1936). An experimental and theoretical study of changes in surface colors under changing illuminations. *Psychological Bulletin*, *33*, 740-741.
- Hurlbert, A. (1998). Computational models of color constancy. In V. Walsh & J. Kulikowski (Eds.), *Perceptual constancies: Why things look as they do* (pp. 283-322). Cambridge, UK: Cambridge University Press.
- Hurlbert, A. C. (1989). *The computation of color*. Unpublished doctoral dissertation, Cambridge, MA: Harvard Medical School/Massachusetts Institute of Technology.
- Kaiser, P. K., & Boynton, R. M. (1996). *Human color vision* (2nd ed.). Washington, DC: Optical Society of America.
- Kelley, K. L., Gibson, K. S., & Nickerson, D. (1943). Tristimulus specification of the Munsell Book of Color from spectrophotometric measurements. *Journal of the Optical Society of America A*, *33*, 355-376.
- Knill, D. C., Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press.
- Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences U S A*, *96*, 307-312. [PubMed]
- Kraft, J. M., Maloney, S. I., & Brainard, D. H. (2002). Surface-illuminant ambiguity and color constancy: Effects of scene complexity and depth cues. *Perception*, *31*(2), 247-63. [PubMed]
- Kurichi, I., & Uchikawa, K. (1998). Adaptive shift of visual sensitivity balance under ambient illuminant change. *Journal of the Optical Society of America A*, *15*, 2263-2274. [PubMed]
- 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. [PubMed]
- Larson, G. W., & Shakespeare, R. (1997). *Rendering with radiance*. San Francisco, CA: Morgan Kaufmann.
- Lee, H. -C. (1986). Method for computing the scene-illuminant chromaticity from specular highlights. *Journal of the Optical Society of America A*, *3*, 1694-1699. [PubMed]
- Lee, H. -C., Breneman, E. J., & Schulte, C. P. (1990). Modeling light reflection for computer color vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *12*, 402-409.
- Maloney, L. T. (1999). Physics-based models of surface color perception. In K. R. Gegenfurtner & L. T. Sharpe (Eds.), *Color vision: From genes to perception* (pp. 387-418). Cambridge, UK: Cambridge University Press.
- Maloney, L. T., & Landy, M. S. (1989). A statistical framework for robust fusion of depth information. In W. A. Pearlman (Ed.), *Visual communications and image processing IV. Proceedings of the SPIE*, *1199*, 1154-1163.
- Maloney, L. T., & Schirillo, J. A. (2002). Color constancy, lightness constancy, and the articulation hypothesis. *Perception*, *31*, 135-139. [PubMed]

- Maloney, L. T., and Wandell, B. A. (1986). Color constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America A*, 3, 29-33. [\[PubMed\]](#)
- Maloney, L. T., & Yang, J. N. (in press). The illumination estimation hypothesis and surface color perception. In R. Mausfeld & D. Heyer (Eds.), *Colour vision: From light to object*. Oxford, UK: Oxford University Press.
- Mamassian, P., Landy, M. S., & Maloney, L. T. (2002). Bayesian modelling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki, (Eds.), *Probabilistic models of perception and brain function* (pp. 13-36). Cambridge, MA: MIT Press.
- Mausfeld, R. (1997). Colour perception: From Grassman codes to a dual code for object and illuminant colours. In W. Backhaus, R. Kliegl, & J. Werner (Eds.), *Color vision*. Berlin, Germany: De Gruyter.
- Rao, R. P. M., Olshausen, B. A., Lewicki, M. S. (Eds.). (2002). *Probabilistic models of perception and brain function*. Cambridge, MA: MIT Press.
- Shafer, S. A. (1985). Using color to separate reflectance components. *Color Research and Applications*, 4, 210-218.
- von Helmholtz, H. (1962). *Helmholtz's treatise on physiological optics*. New York: Dover.
- Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulas* (2nd ed.). New York: Wiley.
- Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vision Research*, 41, 2581-2600. [\[PubMed\]](#)
- Yuille, A. L., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123-161). Cambridge, UK: Cambridge University Press.