Color constancy and hue scaling

Sven Schultz   Institut für Psychologie, Martin-Luther-Universität, Halle-Wittenberg, Germany

Katja Doerschner   Department of Psychology, New York University, New York, NY, USA

Laurence T. Maloney   Department of Psychology and Center for Neural Science, New York University, New York, NY, USA

In this study, we used a hue scaling technique to examine human color constancy performance in simulated three-dimensional scenes. These scenes contained objects of various shapes and materials and a matte test patch at the center of the scene. Hue scaling settings were made for test patches under five different illuminations. Results show that subjects had nearly stable hue scalings for a given test surface across different illuminants. In a control experiment, only the test surfaces that belonged to one illumination condition were presented, blocked in front of a black background. Surprisingly, the hue scalings of the subjects in the blocked control experiment were not simply determined by the color codes of the test surface. Rather, they depended on the sequence of previously presented test stimuli. In contrast, subjects' hue scalings in a second control experiment (with order of presentations randomized) were completely determined by the color codes of the test surface. Our results show that hue scaling is a useful technique to investigate color constancy in a more phenomenological sense. Furthermore, the results from the blocked control experiment underline the important role of slow chromatic adaptation for color constancy.

Keywords: surface color perception, hue scaling, color naming, stereo vision, chromatic adaptation, color constancy

Introduction

The term "color constancy" is often used to describe the phenomenon of stable color appearance of surfaces with changing illumination (Jameson & Hurvich, 1989; Kaiser & Boynton, 1996). This formulation of color constancy refers directly to human perceptual experience. However, quantitative measurements of color constancy under various experimental conditions have shown that human color constancy is far from perfect (Arend & Reeves, 1986; Kraft & Brainard, 1999; Kuriki & Uchikawa, 1996). Several explanations have been suggested that may account for this imperfect performance, including effects of instruction (Arend & Reeves, 1986), inappropriateness of the stimuli used (Kraft & Brainard, 1999), increment–decrement asymmetries (Bäuml, 2001), and the size of the illuminant shift (Craven & Foster, 1992).

Psychophysical methods such as achromatic settings or asymmetric matching promise accurate, quantitative assessment of color constancy. It is not clear how well such tasks characterize the stability of observer's color appearance across the full range of surface colors. However, although spatial or temporal asymmetric matching can be used to characterize perception of colors away from the neutral point, recent results suggest that it might be inappropriate to evaluate human color constancy for a variety of reasons (Foster, 2003; Logvinenko & Maloney, 2006; Maloney, 1999). In particular, when observers are asked to set a test surface under a given illumination so that it appears perceptually indistinguishable to a reference patch under a second illumination, they sometimes report that they cannot find a satisfying setting. This problem was first noted by David Katz, who reported that when observers make a match in a lightness or color constancy experiment, there is usually a residual difference (Katz, 1911, p. 82). The following comment from a recent asymmetric color matching study may illustrate this problem:

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

If asymmetric matches do not, in fact, match, then it is questionable whether they can be used to characterize color constancy.

Achromatic setting measures, for example, essentially measure only the location of the observer's neutral point, and their use to characterize color appearance away from the neutral point involves assumptions. Speigle and Brainard (1999) report, for example, that achromatic settings can be used to predict asymmetric matches, but, as just noted, it is unclear whether the latter can be used to measure stability of color appearance under changes in illumination.

A task that is more directly related to the purpose of color constancy in terms of object recognition is color

categorization (Boynton & Olson, 1987). The grouping of colors into a small number of discrete categories seems to be a universal feature of the visual system (Kay & Regier, 2003). Jameson and Hurvich (1989) argue that color categories of objects are preserved with changes of the illuminant, if one takes compensatory mechanisms like chromatic adaptation into account. Troost and deWeert (1991) investigated the color constancy performance of observers in a color categorization task and asymmetric matchings. They found that observers reliably used the same color category for a test patch with changes of the illumination. Although color categorization seems to be a reliable and appropriate measure of human color constancy, two disadvantages of this task are the lack of comparability of the color categorization data to quantitative measures and the potential that the conclusions drawn depend on an arbitrary choice of categories.

In this study, we introduce hue scaling as an alternative method to investigate color constancy (Abramov & Gordon, 1994; Boynton & Gordon, 1965). On one hand, this task is evidently based on judgments of the appearance of chromatic surfaces under varying illumination and is therefore appropriate to study human color constancy. On the other hand, the resulting data are potentially comparable to quantitative measures of color constancy. Moreover, with the usage of a hue scaling task, we hope to eliminate disadvantages that are inherent in asymmetric matchings and achromatic settings. We also test a generalization of the model proposed by Speigle and Brainard (1999) to link achromatic settings and asymmetric matching but in the context of hue scaling. Specifically, we test whether knowledge of the hue scalings of a single test patch under two lights can be used to predict scaling of all test patches under these two lights. We will describe their model and our tests in more detail in the Data analysis section of the first experiment.

Previous studies of color constancy have employed stimuli consisting of patterns of flat surfaces embedded in a fronto-parallel plane, often called "Mondrians" (Land & McCann, 1971). A disadvantage of this configuration is that more complex light–surface interactions are not taken into account (Boyaci, Doerschner, Snyder, & Maloney, 2006). Moreover, reported indices of color constancy are typically low. In the experiments reported here, we used simulated three-dimensional (3D) stimuli presented binocularly to more closely approximate natural viewing conditions (Boyaci et al., 2006; Boyaci, Maloney, & Hersh, 2003; Maloney, 1999). These scenes contain additional cues to the illuminant such as specular highlights or shadows. The use of these enriched stimuli was also motivated by the idea that the visual system seems to combine cues that are available in a scene to estimate the illuminant (Boyaci, Doerschner, & Maloney, 2006; Kraft & Brainard, 1999; Snyder, Doerschner, & Maloney, 2005; Yang & Maloney, 2001).

In addition to the cues presented in a scene, chromatic adaptation plays a crucial role in the adjustment of the visual system to the illuminant (Kuriki & Uchikawa, 1996). Studies of the time course of chromatic adaptation have revealed that this process consists of a fast and a slow phase of adaptation (Fairchild & Reniff, 1995; Rinner & Gegenfurtner, 2000). The slow phase of adaptation might be related to slow changes of daylights that occur in the daytime. In Experiment 2 of this study, we were able to demonstrate, by accident, the strength of the isolated slow adaptation mechanism.

Experiment 1: 3D scenes

Methods

Stimuli

The stimuli in this experiment were computer-rendered 3D scenes that consisted of simple objects (such as columns and spheres) with different colors and reflectance properties (e.g., shiny, matte). These objects served as potential cues to the chromaticity and intensity of the light sources. A matte test patch was presented at the center of the scene.

The scenes were illuminated by a mixture of a simulated diffuse and a simulated punctate source. To present the 3D scenes stereoscopically, for each scene, we rendered two images from two different viewpoints [1] corresponding to the locations of the observer's eyes in the virtual scene (Figure 1). The scenes differed in the chromaticity of the test patch and in the color of the punctate source.

Software and apparatus

The stimuli were presented to the subjects stereoscopically on two Sony Trinitron Multiscan GDM-F500 21-in. CRT screens (Figure 2). We ran the experiments under Fedora Linux Core 3 using a Dell workstation with an NVIDIA dual-head graphics card that controlled both monitors. The two images of each scene were rendered with the Radiance package (Larson & Shakespeare, 1996). The two output files of the rendering procedure contained relative RGB triplets. These values were corrected for nonlinearities of the gun responses using measured lookup tables for each monitor. The lookup tables were based on direct measurements of the luminance values on each monitor with a Photo Research PR-650 spectroradiometer. Finally, the corrected RGB triplets were translated to 24-bit graphics codes. The experiments were programmed by us in the C language using the X Window System, Version 11R6 (Scheifler & Gettys, 1996). The settings were made by the subjects using a mouse.

Test patch

We used 16 different test patches that represented simulations of Munsell chips (Table 1). All test patches had identical Munsell value and chroma of 7 and 4, respectively. They differed only in hue and formed a color
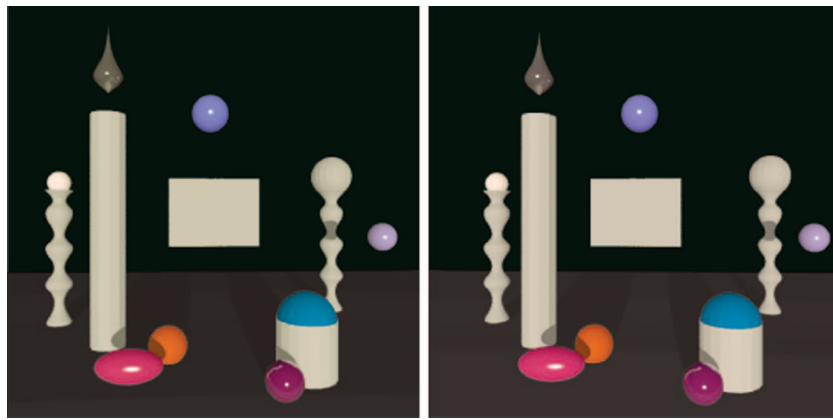
Figure 1. Typical stimulus. Left and right images of the scene under neutral illumination.

"circle" in the CIE $u'v'$ diagram (Figure 3). We used this set of test patches in each of the three experiments.

To increase the accuracy of stimuli, we rendered the test surfaces separately from the rest of the scene. Rendering packages such as Radiance simulate light–surface interactions as the product of the respective RGB codes. This "RGB heuristic" (Maloney, 1999) does not always lead to adequate simulation of light–surface interaction. Therefore, we computed the light signal that reached the eye from the test patch according to the model described in Boyaci et al. (2003). The spectral energy distribution of the light signal $L(\lambda)$ from a Lambertian surface $S(\lambda)$ that is illuminated by a diffuse source $E_D(\lambda)$ and a punctate source $E_P(\lambda)$ is given by

$$L(\lambda) = S(\lambda) \cdot (E_P(\lambda) \cdot \cos\theta + E_D(\lambda)), \qquad (1)$$

where λ indexes the wavelength of light in the visible spectrum and θ is the angle between the incident light from the punctate source and the surface normal. In our case, θ was always set to zero. The reflectance functions

Figure 2. The apparatus used in this study.

of the surfaces are based on spectral reflectance measurements made with a spectrophotometer on 1,269 color chips from the 1976 Munsell Book of Color at a 1-nm resolution from 380 to 800 nm. This data set was obtained from http://spectral.joensuu.fi/.

The size of the test patch was 2.9×2.2 cm. The distance between the test surface and the observer's viewpoint was 70 cm. The surface normal of the test patch was identical to the viewing axis and to the direction of the incident light from the punctate source. We held the orientation of the test patch constant throughout all experiments.

Light sources

As noted above, we used simulations of a diffuse and a punctate light source. The diffuse source was always a neutral daylight (D65). The chromaticity of the punctate source was manipulated in the experiment. Five different illuminations were used. Three of them were simulations of CIE daylights with correlated color temperatures of 6500K (D65), 4000K (D40), and 10000K (D10) that appeared achromatic, yellow, and blue to the observer, respectively. Furthermore, two artificial illuminants that appeared red and green to the observer were used (Figure 4). The xy and $u'v'$ chromaticities of the punctate sources are given in Table 2. For the five punctate sources, spectral energy distributions were computed from the CIE daylight basis functions (Wyszecki & Stiles, 1982). We calculated the light signal that reached the eye of the observer from the test patch using these distributions and the reflectance functions of the Munsell surfaces. The virtual punctate source was simulated to be behind the observer. The distance between the test patch and the punctate light source was 670 cm. The position of the punctate source was held constant throughout all experiments. The punctate–total luminance ratio was always $\pi = 0.67$ (see Boyaci et al., 2003).

Task

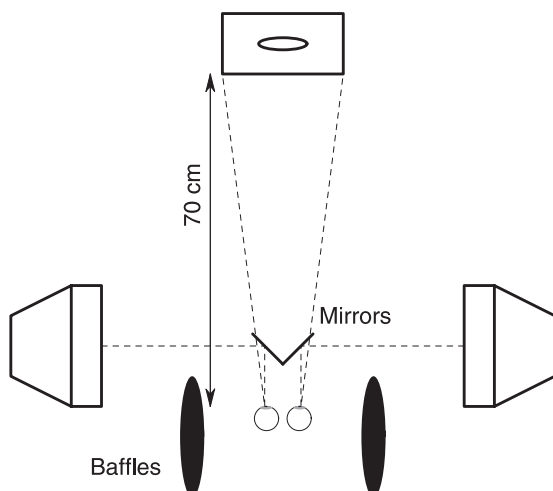Subjects were told that they were viewing a scene under a certain illumination. The task of the subject was to set hue

Munsell notation	*x*	*y*	*L*	*u′*	*v′*
5R7/4	0.357	0.333	18.62	0.228	0.477
10R7/4	0.366	0.346	19.15	0.228	0.485
5YR7/4	0.373	0.362	18.82	0.226	0.494
10YR7/4	0.375	0.383	18.83	0.219	0.504
5Y7/4	0.369	0.397	18.37	0.210	0.509
10Y7/4	0.360	0.405	18.74	0.201	0.511
5GY7/4	0.341	0.400	17.88	0.192	0.506
10GY7/4	0.313	0.387	18.31	0.178	0.496
5G7/4	0.290	0.365	17.97	0.170	0.483
10G7/4	0.281	0.353	18.36	0.168	0.476
5BG7/4	0.271	0.338	17.96	0.166	0.467
10BG7/4	0.265	0.323	18.12	0.167	0.458
5B7/4	0.262	0.307	18.24	0.170	0.449
10B7/4	0.272	0.301	18.36	0.179	0.446
5P7/4	0.304	0.296	18.56	0.205	0.448
5RP7/4	0.337	0.314	18.26	0.222	0.464

Table 1. CIE *xy* chromaticity coordinates, luminance, and *u′v′* chromaticity coordinates of the 16 test patches under illumination D65.

scalings for the test patch. Subjects were asked to judge how blue, yellow, red, and/or green a test surface appeared to them on four respective scales ranging from 0 (*none*) to 6 (*very saturated*). The active scale was presented monocularly as the number of the current value in the respective color of the scale. The subject saw this number to the right of the fused image. At any given time, only one hue scale was active. The subject could increase or decrease the value of the active scale by pressing the left or the right mouse button, respectively. The subject moved to a different scale by scrolling the mouse wheel up or down. The settings of one trial were confirmed by pressing the mouse wheel.

We could have chosen a rating scale with more or fewer steps. We chose six based on previous work. De Valois,

De Valois, Switkes, and Mahon (1997, p. 887) reported that "In preliminary trials, we found that observers preferred, and differentially used, a scale finer than the three-level scale, e.g., GGG, GGY, GYY, YYY, etc., used by Boynton and Gordon (1965), but did not require a 100-point scale as used by Abramov, Gordon, and Chan (1990)." Even if subjects do not make nonzero setting on both scales of an opponent pair such as blue–yellow, they can still make 13 distinct ratings for each opponent pair and, combining ratings from the two opponent scales, they can classify each stimulus in 1 of 169 (13 × 13) different categories. Our method allows much finer comparison than color naming procedures where only 11 categories are typically allowed.
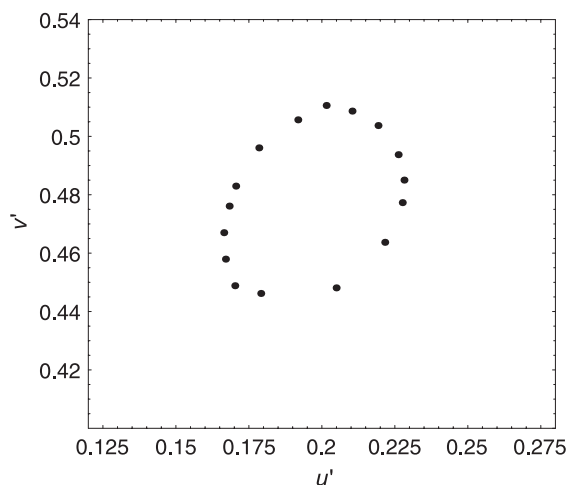

Figure 3. Chromaticities of the 16 Munsell surfaces under neutral illumination in *u′v′*. Under chromatic illumination, the set of Munsell surfaces is shifted toward the illuminant.
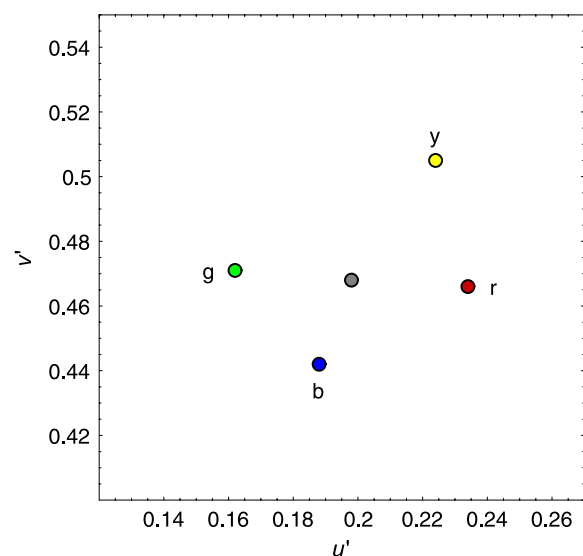

Figure 4. Chromaticities of the punctate illuminants in *u′v′*. The letters refer to the color appearance of these lights (b, blue; y, yellow; r, red; g, green).

Illumination	x	y	u′	v′
Neutral (D65)	0.313	0.329	0.198	0.468
Blue (D10)	0.279	0.292	0.188	0.442
Yellow (D40)	0.382	0.384	0.224	0.505
Red	0.354	0.313	0.234	0.466
Green	0.268	0.347	0.162	0.471

Table 2. CIE *xy* and *u′v′* chromaticity coordinates of the five punctate sources.

Procedure

Subjects practiced hue scalings in a training session before running in the experiments. Only subjects who, in their settings, had reliability of 0.8 and more took part in the study. [2] In the experiments, different illuminant conditions were blocked into different sessions. Each session started with 16 training trials. These training trials were prepended to the experimental trials to allow subjects to practice hue scalings as well as to stabilize the adaptational state of the subject. In each trial, the scene was presented to the subject and the subject carried out the task described above. Between two trials, a black screen was presented to the subject for 1 s to reduce the influence of afterimages. Observers repeated hue scalings for each of the 16 test surfaces four times. Each subject made 80 settings in one session. There were no time constraints. One session took about 20 min on the average. Each experiment consisted of five sessions that corresponded to the five different punctate sources. The order of illuminant blocks was randomized and differed across subjects. Within one illuminant block, the test surface order was randomized and different subjects saw different randomizations.

Observers

Five subjects took part in the experiment. All subjects were paid undergraduate students who were not aware of the purpose of the study. All had normal color vision, which was determined via the Ishihara color plates test (Ishihara, 1997).

Data analysis

We first give an overview of the general methods of data analysis. The same analyses were carried out for all experiments.

Absolute hue scalings

As a first step, we analyzed the absolute hue scalings of subjects under different illumination conditions. For each subject and each surface, we calculated mean hue scalings. We plotted these data as polar coordinates. In this representation, each of the 16 surfaces corresponds to a fixed angle. Hue scalings of a subject are plotted on the axis belonging to the respective surface. Each figure contains mean hue scalings of a subject on the four scales. The settings on a fixed scale are connected and form a geometric object that looks like a bubble. For a color constant subject, we expect the form and size of these bubbles not to change with changes of the illuminant. A subject shows no constancy if form and size of the bubbles are determined only by the local color signal of the test surface under each illumination.

Transformed hue scalings

In the second step of our analysis, we transformed mean hue scalings for each subject in each illumination condition to *u′v′* chromaticity coordinates. This transformation makes it easier to compare our results with those from previous studies in terms of color constancy indices. The procedure we followed to obtain these indices is explained next.

We note, first of all, that subjects made nonzero settings for a given test surface on either the blue or the yellow scale but not both and on either the red or the green scale but not both. Therefore, for each given subject and condition, we combined blue and yellow ratings into a single number a_{BY} on a BY opponent scale that ran from -6 (blue) to 6 (yellow). If, for example, the subject's rating was 2.3 on the blue scale and 0 on the yellow, the rating on the opponent scale would become $a_{BY} = -2.3$. A rating of 1.7 on the yellow scale (with a rating of 0 on the blue) would become $a_{BY} = 1.7$. We similarly combined red and green hue ratings into a rating a_{RG} on a red–green opponent scale. An observer's mean hue ratings in any given condition are then summarized by the two-dimensional column vector $\mathbf{a} = [a_{BG}, a_{RG}]′$ in a two-dimensional opponent space, $\mathbf{A}$.

Let $\mathbf{u}$ be the two-dimensional vector that denotes the *u′v′* coordinates of a surface under the neutral illuminant. We assume that the mapping between the opponent space $\mathbf{A}$ and *u′v′* space is affine but perturbed by judgment error. That is, we assume that there is a 2×2 matrix $\mathbf{M}$ and a column vector $\mathbf{c}$ such that

$$\mathbf{u} = \mathbf{Ma} + \mathbf{c} + \varepsilon, \qquad (2)$$

where $\varepsilon \sim \Phi(0, \sigma^2)$ is Gaussian judgment error with a mean of 0 and a variance of σ^2. The vector $\mathbf{c}$ captures the shift of the achromatic point.

For the neutral illumination, the chromaticity coordinates of the surfaces and mean hue scalings of the subjects

are known. We used these data to estimate the six parameters of the transformation (the four elements of **M** and the two elements of **c**) for each subject. We chose the set of parameters for which the sum of squared distances between $u'v'$ coordinates and predictions was minimized.

This set of parameters, **M** and **c**, determined by the observer's ratings under neutral illumination, was then used to compute $u'v'$ coordinates from hue scalings under other illuminations.

The Speigle–Brainard conjecture

Speigle and Brainard ([1999](#)) asked subjects to make achromatic settings and asymmetric matches under different illuminations and then sought to predict the latter from the former. The question they address can be formulated as follows: Suppose that the experiment has measured the shift in color space of the achromatic point induced by a change from one illumination to a second. Can the experimenter now predict the pattern of shift of *any* surface (not just an achromatic surface) induced by the same change in illumination? Speigle and Brainard conclude that, although the transformations on all points in color space as measured by asymmetric matching are *potentially* too complex to be predicted by knowledge of the transformation of the achromatic point, in *reality,* they can be accurately predicted.

In hue scaling terms, the conjecture of Speigle and Brainard can be stated as a claim that knowledge of the shift in hue scalings of a neutral reference surface determines all other possible hue scalings. In testing this conjecture, we will use as reference surface the centroid of all of the surface patches under a single light. We do not assume that this surface, which the subject never sees, is precisely achromatic, but, given the construction of our stimuli, it will be nearly so.

We tested the Speigle–Brainard conjecture in the following way. First, we transformed all opponent data of subject i to $u'v'$ using the specific transformation that was given by $\mathbf{M}_i$ and $\mathbf{c}_i$. Then, we fitted optimal ellipses to the $u'v'$ data of each subject for each illuminant (Halir & Flusser, [1998](#)). Finally, we compared the parameters of the ellipse for neutral illumination data with those of the ellipses for the chromatic illumination data. Analyzing the parameters of the ellipses is advantageous in that we can directly separate effects of color constancy from deviations from the Speigle–Brainard conjecture. If we analyze the ellipses for neutral and chromatic illumination, then differences in the centroids reflect incomplete color constancy. However, differences in area, eccentricity (shape), or orientation of the ellipses indicate deviations from the Speigle–Brainard conjecture. If their conjecture were valid, then changes in illuminant chromaticity would lead to simple translations of the ellipses in $u'v'$ space. They will not rotate, shrink or grow, or change shape. We next describe how we measure area, eccentricity, and orientation.

Let a be the semimajor axis and b be the semiminor axis of an ellipse. Then, the eccentricity e of the ellipse is given by

$$e = \sqrt{1 - \frac{b^2}{a^2}}. \tag{3}$$

The orientation of the ellipse is given by the radian angle θ between the major axis of the ellipse and the x-axis. The area A of an ellipse is defined as $A = \pi ab$.

Color constancy indices

We calculated color constancy indices for the transformed hue scalings of each subject. In case of perfect constancy, the $u'v'$ coordinates of the scalings under test illumination should completely overlap with those under neutral illumination. If a subject has no constancy, the $u'v'$ coordinates of the scalings under test illumination coincide with the coordinates of the test surfaces under this illumination.

The centroid of $u'v'$ chromaticities of all surfaces under neutral illumination was almost identical to the centroid of the predicted coordinates. Therefore, we used the centroids of test surfaces and settings in $u'v'$ to calculate color constancy indices. The indices have the form

$$\mathrm{CI} = 1 - \frac{|\mathbf{m}_{\mathrm{D65}} - \mathbf{m}_{\mathrm{data}}|}{|\mathbf{m}_{\mathrm{D65}} - \mathbf{m}_{\mathrm{test}}|}, \tag{4}$$

where $\mathbf{m}_{\mathrm{D65}}$ is the centroid of the surfaces under neutral illumination, $\mathbf{m}_{\mathrm{test}}$ is the centroid of the surfaces under test illumination, and $\mathbf{m}_{\mathrm{data}}$ is the centroid of the transformed hue scalings of the subject. This index is 1 in case of perfect constancy and 0 in case of no constancy. By design, it is directly comparable to Brunswik ratios typically reported in studies of color constancy (Arend, Reeves, Schirillo, & Goldstein, [1991](#); Brunswik, [1929](#); Thouless, [1931](#)).

Results

Expected results

In the 3D experiment, subjects are enabled to use various cues to the illuminant. Therefore, we expect constant hue scalings with changes of the illuminant. This constancy might not be perfect but rather lead to settings different from zero on one scale for a surface under different illuminations. It has been hypothesized that the visual system might use regularities of natural daylights, like the location of the corresponding chromaticities along the so-called daylight locus, to achieve color constancy (Judd, MacAdam, & Wyszecki, [1964](#); Shepard, [1994](#)). Recent studies have found no evidence to support this assumption (Brainard, [1998](#); Delahunt & Brainard, [2004](#)). Consistent

with these empirical findings, we do not expect color constancy performance to be better for natural daylights than for artificial red and green illuminations displaced from the daylight locus.

Experiment 1: Results

Subjects consistently used the same scale settings to judge the color appearance of a fixed surface under different illuminants. For example, a test surface that appeared blue under neutral illumination was also rated mainly blue under the chromatic illuminations. The magnitude of the scalings usually differed with changes of the illuminant. For example, the "blue" surface was rated to appear more saturated blue under blue illumination and less saturated blue under yellow illumination. In this sense, subjects showed stable hue scalings for a given test surface with changes of the illuminant (Figure 5). We find that the transformed hue scaling settings for a given test patch under a chromatic test illumination fell closer to the chromaticity of this test patch under neutral illumination (Figure 6A).

For each subject, we compared the fitted ellipses of the neutral illuminant data with those of each chromatic illuminant condition to test the Speigle–Brainard conjecture. In general, the shapes and the sizes of ellipses of the chromatic illuminant conditions did not deviate systematically from those of the neutral illuminant condition (Figures 7A and 7B). The analysis of the orientation parameter θ should be interpreted with care as the estimated ellipses of one subject (X.H.) were almost circular, which resulted in unreliable estimates of θ. In general, we found small systematic deviations for the orientation of the ellipse in the yellow and green illuminant conditions (Figure 7C). The analysis of the ellipses indicates that the Speigle–Brainard conjecture holds also for hue scaling data.

The constancy indices that we calculated from these transformed data lay between 0.58 and 0.91 with a median of 0.705 (Table 3). This indicates good to excellent color constancy (Figure 8). The results from Experiment 1 also indicate that color constancy performance is not better for daylights than for artificial red and green illuminations.

Experiment 2: Blocked control

Methods

Stimuli

In this experiment, the isolated test patch was presented in front of a black background. All stimuli of one illumination condition were shown as a block.

Software and apparatus

The same set of software and apparatus as in Experiment 1 was used in this experiment.

Test patch

The set of test patches in Experiment 2 was identical to that used in Experiment 1.

Light sources

The same light sources as in Experiment 1 were used to calculate the light signals of the test patches.

Task

In this experiment, subjects were not told that the stimuli might be interpreted as a surface under a certain illumination. Subjects set hue scalings for the test patch as in Experiment 1.

Procedure

The same procedure as in Experiment 1 was used. As before, the order of illuminant blocks was randomized and differed across subjects. Within one illuminant block, the test surface order was randomized and different subjects saw different randomizations. The randomizations used were not the same as those used in the previous experiment.

Observers

The observers for this experiment were the same five subjects who took part in Experiment 1.

Results

Expected results

In designing Experiment 2, we originally believed that we had not given the subject any cues to the illumination. Therefore, we expected the mean hue scalings of the subjects to depend only on the color codes of the surface under each illumination. We therefore expected the subjects to show no constancy in their settings.

Experiment 2: Results

In the blocked control, all test surfaces of one illumination condition were presented to the subjects within one session in front of a black background. There were no cues to interpret the stimulus configuration as a surface under an illumination, and this possible interpretation was not mentioned explicitly. Surprisingly, the scalings of the subjects in this experiment were not simply determined by the color codes of the test surface (Figures 5 and 6B). For example, when we used a surface that appeared green under neutral illumination and a test patch that appeared red under neutral illumination, the green surface under the red illuminant and the red surface under green illumination produced identical color codes, but, even in this case, subjects rated the former green whereas the latter was judged to appear red. The constancy indices suggest that it
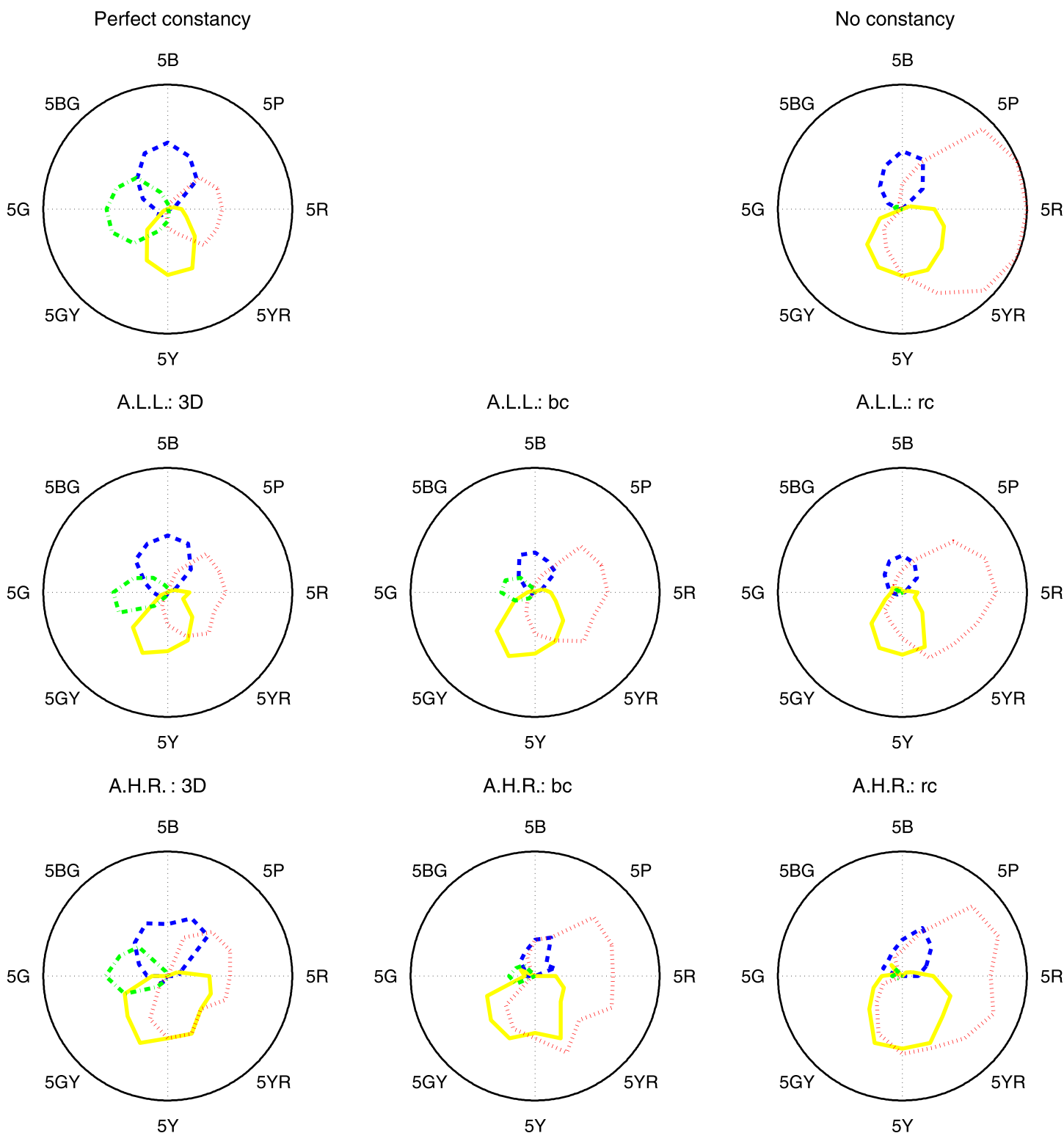
Figure 5. Mean hue scalings of all subjects and of subject A.H.R. under red punctate illumination as polar plots. Surfaces are given in Munsell notation. Different line styles and colors indicate the four scales: solid, blue; dashed, yellow; dotted, red; dashed–dotted, green. The first row gives results corresponding to perfect color constancy, that is, scalings under neutral illumination (left) and hypothetical results corresponding to complete failure of constancy (right). The plots in the three columns show data from the 3D experiment (3D), the blocked control (bc), and the randomized control (rc), respectively. The second row shows mean scalings of all subjects.
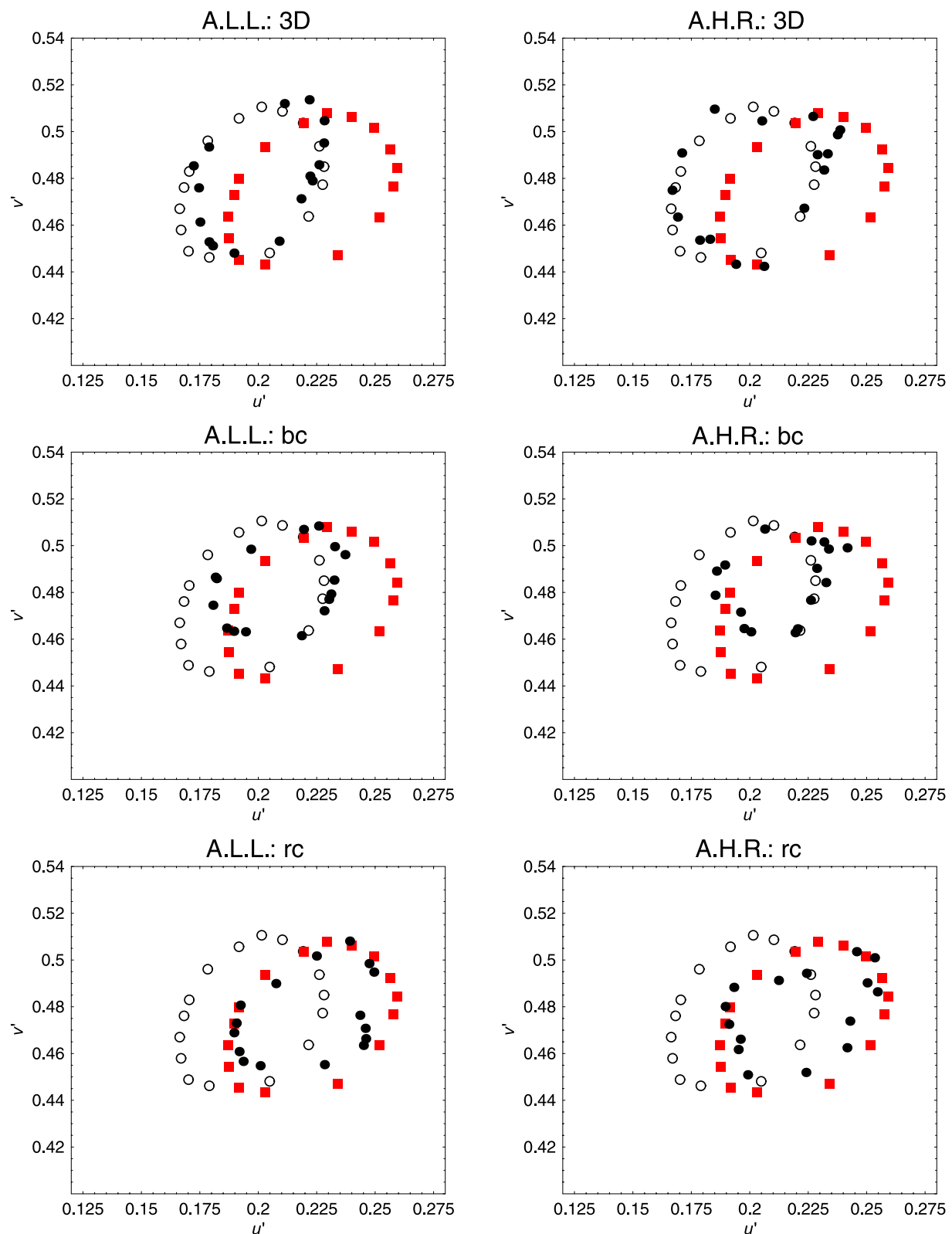
Figure 6. Transformed mean hue scalings of all subjects and of one subject (A.H.R.) under red punctate illumination. The plots in the three rows show data from (A) Experiment 1 (3D), (B) Experiment 2 (bc), and (C) Experiment 3 (rc). Open circles, surfaces under neutral illumination; red-filled squares, surfaces under red illumination; filled circles, transformed mean hue scalings. For an observer with perfect color constancy, we expect the pattern of mean hue scalings to coincide with the ellipse of color signals under neutral illumination (open circles). If the observer has no color constancy, the pattern of mean hue scalings will fall together with the ellipse of color signals under chromatic illumination (red squares).
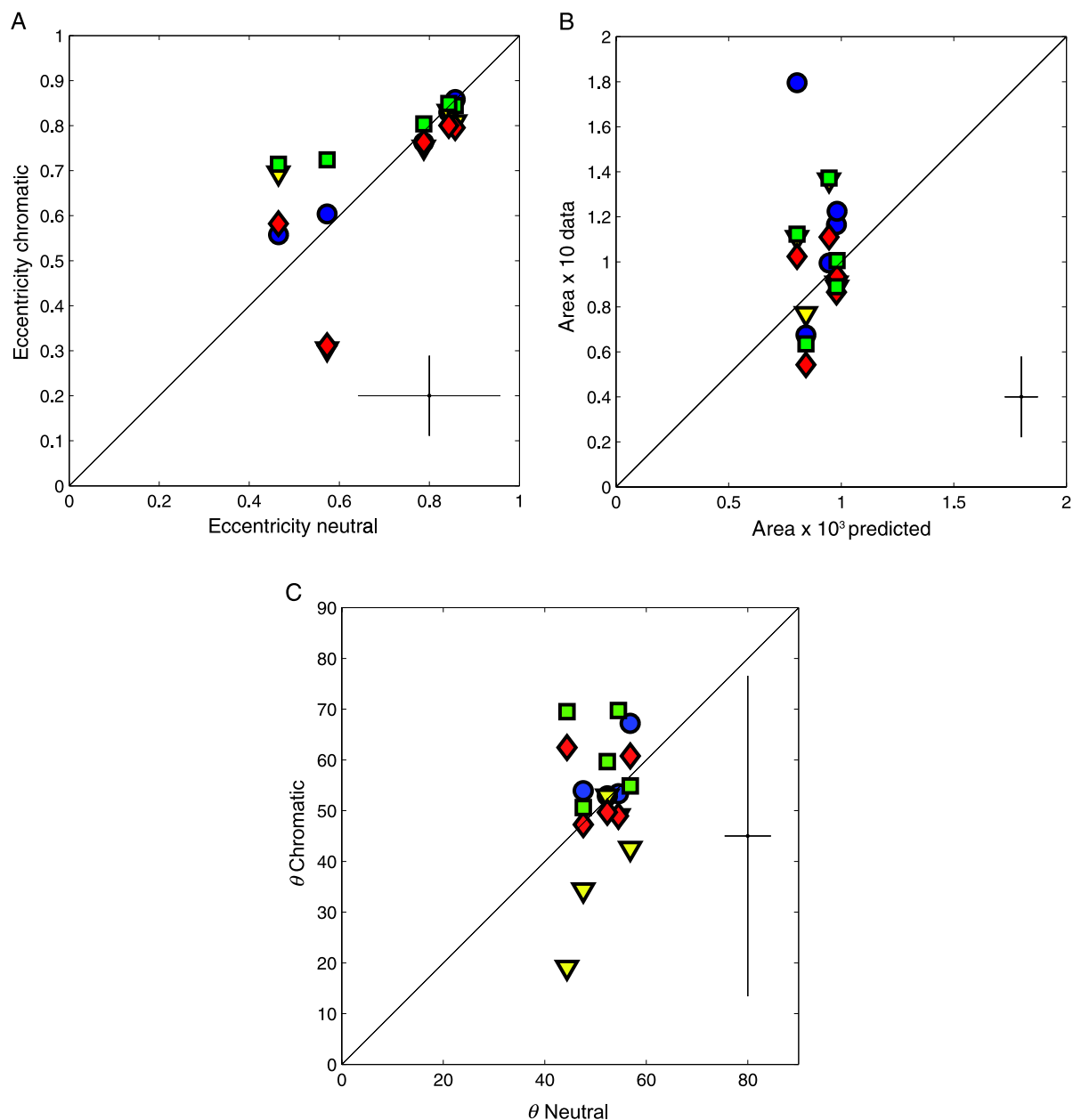
Figure 7. Parameters of the estimated ellipses for neutral and chromatic illuminant conditions. (A) Eccentricities of the estimated ellipses for neutral and chromatic illuminant conditions. Each symbol refers to a different chromatic illuminant (blue, blue circle; yellow, yellow triangle; red, red diamond; green, green square). The eccentricities of a given subject lie on a line parallel to the ordinate (from left to right: A.H.R., X.H., A.L.L., L.G.S., and A.S.T.). The standard deviations for the conditions are shown in the cross at the right part of the figure. (B) Area of the estimated ellipses for neutral and chromatic illuminant conditions. Each symbol refers to a different chromatic illuminant (blue, blue circle; yellow, yellow triangle; red, red diamond; green, green square). The eccentricities of a given subject lie on a line parallel to the ordinate (from left to right: X.H., A.S.T., A.H.R., A.L.L., and L.G.S.). The standard deviations for the conditions are shown in the cross at the right part of the figure. (C) Orientation (θ) of the estimated ellipses for neutral and chromatic illuminant conditions. Each symbol refers to a different chromatic illuminant (blue, blue circle; yellow, yellow triangle; red, red diamond; green, green square). The θ values of a given subject lie on a line parallel to the ordinate (from left to right: X.H., A.S.T., L.G.S., A.H.R., and A.L.L.). The standard deviations for the conditions are shown in the cross at the right part of the figure.

Illumination		A.H.R.	A.S.T.	L.G.S.	X.H.	A.L.L.
Blue (D10)	Experiment 1	0.74	0.71	0.67	0.91	0.69
	Experiment 2	0.72	0.50	0.59	0.73	0.75
	Experiment 3	0.17	−0.18	0.34	0.68	0.08
Yellow (D40)	Experiment 1	0.70	0.63	0.62	0.76	0.64
	Experiment 2	0.29	0.56	0.23	0.18	0.38
	Experiment 3	−0.04	0.13	0.06	−0.23	−0.05
Red	Experiment 1	0.65	0.58	0.58	0.85	0.75
	Experiment 2	0.16	0.46	0.37	0.63	0.46
	Experiment 3	0.06	0.21	0.22	0.05	0.03
Green	Experiment 1	0.75	0.65	0.73	0.73	0.68
	Experiment 2	0.60	0.42	0.72	0.44	0.57
	Experiment 3	0.07	−0.04	0.49	−0.01	−0.11

Table 3. Color constancy indices of mean hue scalings of subjects for all illumination conditions in Experiments 1, 2, and 3. The last column shows the indices of the mean data of all subjects.

appears as if subjects had developed a kind of partial constancy (Table 3; Figure 8). This result held for all subjects and under all illumination conditions.

Smithson and Zaidi (2004) found evidence of a similar dependence of color appearance on the chromaticities of test surfaces presented sequentially. They presented test stimuli one at a time against a fixed background. The task of the observer was to judge the color appearance of the test patch using only the terms "blue," "yellow," "red," and "green," and Smithson and Zaidi estimated the neutral points of the two opponent channels under a given illumination from observers' data. In one experiment, they rendered test patches and background under different illuminants. They found that, in this case, the judgments of observers were significantly affected by the chromaticity of the illumination of the test patches, implying that the visual system somehow estimated this chromaticity

across successive presentations in time. Their results are consistent with those we derived from Experiment 2. We note, however, that there was no visible background in Experiment 2. Consequently, in our experiment, there is no possibility that the effect depended in some way on the presence of a background.

The most plausible explanation for our results and those of Smithson and Zaidi is that the visual system is averaging chromaticities from multiple surfaces across time and is using this average as an estimate of the chromaticity of the illuminant in estimating surface colors. We will discuss this point further in the Summary and discussion section.

Experiment 3: Randomized control

Methods

Stimuli

The stimuli used were the same as in Experiment 2, except that the isolated test patch was presented in front of a black background. In this experiment, however, test surfaces from all illumination conditions were shown in random order. Subjects made hue scalings for 80 different test patches within one session.

Software and apparatus

The same set of software and apparatus as in the previous experiments was used.

Test patch

The set of test patches in Experiment 3 was identical to that in Experiment 1.
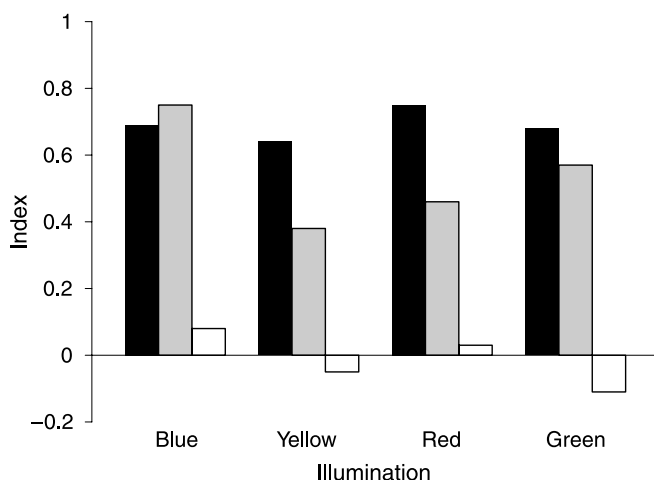

Figure 8. Color constancy indices for the mean hue scalings of all subjects. Black bars, 3D experiment; gray bars, blocked control; white bars, randomized control.

Light sources

The same light sources as in Experiment 1 were used to calculate the light signals of the test patches.

Task

In this experiment, subjects were not told that the stimuli might be interpreted as a surface under a certain illumination. Subjects set hue scalings for the test patch as in Experiment 1.

Procedure

The procedure of Experiment 3 differed from those of Experiments 1 and 2 in the following aspects. Each session included all 16 different surfaces rendered under all five illuminants. The 80 different stimuli of each session were presented in random order, and different subjects saw different randomizations. Each subject completed five sessions.

Observers

Four of the five subjects from the previous experiments took part in Experiment 3.

Results

Expected results

In Experiment 3, no cues to the illumination were given. The sequential presentation of the test surface × illumination combinations was fully randomized. Therefore, we expect the mean hue scalings of the subjects to depend only on the color codes of the surface under test illumination. This means that subjects should show no constancy in their settings.

Experiment 3: Results

The hue scalings in this control were completely determined by the color codes of the test surface (Figure 5). For example, subjects made identical scalings for the green surface under red illumination and the red test patch under green illumination. The transformed hue scalings of the subjects under a chromatic test illumination fell together with the color codes of the surfaces under this illuminant (Figure 6C). According to this pattern, most of the constancy indices in this experiment are close to zero and indices are always smaller than those from Experiments 1 and 2 (Table 3; Figure 8).

Summary and discussion

In this study, we measured color constancy performance using a hue scaling task. In the first experiment, subjects showed constant hue scalings for a given test patch with changes of the illuminant. The degree of color constancy that we observed in the first experiment was comparable to that found in previous studies (Brainard, 1998; Foster, Amano, & Nascimento, 2001). In Experiment 2, subjects were presented only with a sequence of isolated surfaces that were related to a given illumination condition. Nevertheless, hue scalings of the subjects still indicated a moderate degree of constancy. It seemed as if subjects had adapted to the whole sequence of previously presented stimuli. In contrast, the constancy disappeared when the surfaces were presented completely randomized to the subjects (Experiment 3).

Three results stand out.

First, the results from Experiment 1 indicate that hue scaling is an appropriate technique to investigate color constancy in a more phenomenological sense. Furthermore, we found that a measure of color constancy based on the hue scaling data is comparable to previous quantitative measures of color constancy performance.

In a recent essay on color constancy, David Foster pointed out that achromatic settings and asymmetric matchings are problematic methods to measure color constancy and that more direct approaches that serve this purpose are needed (Foster, 2003). This study is a first step toward developing and validating measures of color constancy that are more closely related to the phenomenon of color constancy and that can be used to quantify the degree of color constancy.

Second, over the range of stimuli considered, we found that the Speigle–Brainard conjecture (generalized to hue scaling) held. Speigle and Brainard (1999) were able to predict asymmetric matchings from achromatic settings. In our analysis of the data ellipses, we could show that the centroid of the ellipse under a given chromatic illumination (i.e., the achromatic point) is needed to reconstruct an approximation to the locations of the respective chromatic test surfaces. This result, based on hue scaling, supports the view of Speigle and Brainard.

Third, the surprising results from Experiment 2 in comparison to those from Experiment 3 showed the strong dependence of settings from previously presented surfaces. A comparison across all panels of Figure 6 illustrates the differences between the three experimental conditions. This finding underlines the strong contribution of slow chromatic adaptation to the illuminant adjustment of the visual system.

In the literature, the distinction between successive and simultaneous color constancy has been made (Bäuml, 1999; Brainard, 1998; Brainard et al., 1997). Both terms can be related to two different situations in the natural environment. Successive color constancy refers to gradual changes of one uniform illumination as it happens when daylight changes from dawn to noon. Simultaneous color constancy can be related to a situation where one part of a scene is exposed to direct sunlight whereas another part is situated in the shade. Based on these terminologies, our study refers to successive color constancy. Our results are

consistent with the claim that successive color constancy is simply due to chromatic adaptation (Kuriki & Uchikawa, 1996). However, further investigation of the factors that contribute to changes in adaptational state across time is needed before drawing conclusions.

Recent studies on the time course of chromatic adaptation revealed that chromatic adaptation consists of at least two processes, a fast and a slow phase of adaptation, and that chromatic adaptation is complete after about 2 min (Fairchild & Reniff, 1995; Rinner & Gegenfurtner, 2000; Shevell, 2001; Werner, Sharpe, & Zrenner, 2000). In these studies, test stimuli were presented to the subjects on larger adaptation backgrounds. As in this stimulus configuration, simultaneous and successive contrasts are confounded, and the two suggested processes of chromatic adaptation might be related to these two phenomena (Rinner & Gegenfurtner, 2000). In Experiment 2, we presented only isolated test patches to the subjects. The results from Experiment 2 can be interpreted as the isolated effect of slow (sensory) chromatic adaptation. The observer, given samples of surfaces from a scene, one at a time, is able to develop an estimate of the illuminant and correct surface color appearance for this illuminant estimate, at least in part. More research is needed to clarify the relationships between the phenomena of successive and simultaneous contrasts, the two processes of chromatic adaptation, and stability of color appearance in ongoing scenes.

Another interesting outcome of Experiment 2 is the fact that the visual system is able to integrate information from isolated stimuli temporally. As we noted above, Smithson and Zaidi (2004) found a similar effect in scenes where single test surfaces were presented in sequence against a fixed background. It is interesting to speculate what role this temporal integration of surface chromaticities might play in surface color perception. D'Zmura and Lennie (1986) conjectured that the visual system estimates illuminant chromaticity by integration of chromaticities across successive eye movements, in effect developing an estimate of mean chromaticity of the scene by sampling across time. Our results and those of Smithson and Zaidi suggest that the similar averaging process occurs even when the change from one surface to the next is not due to an eye movement but is under experimenter control. There is considerable evidence for slow adaptation processes to exchange very intense backgrounds differing in chromaticity (Augenstein & Pugh, 1977; Reeves, 1982). It remains to be seen how and when such slow adaptational processes affect color appearance.

Acknowledgments

Commercial relationships: none.
Corresponding author: Sven Schultz.
Email: s.schultz@psych.uni-halle.de.
Address: Institut für Psychologie, Brandbergweg 23c, D-06099 Halle (Saale), Germany.

Footnotes

[1]We used, as a standard, an interocular distance of 6.3 cm in rendering. This separation has proven to be sufficient for subjects in these and previous experiments using the same apparatus. The normal human range of interocular distances is 6.0 to 7.0 cm (French, 1921). We checked whether subjects could achieve fusion, and, had any subject reported difficulty with stereo fusion, we would have excluded them from further participation in the experiment.

[2]In the training session, subjects repeated hue scaling settings for 16 surfaces under neutral illumination five times. For each subject, we transformed all pairwise correlations between repeated measurements using Fisher's r-to-z transformation. The subject's reliability was defined as the correlation corresponding to the mean of the transformed values. We did not measure consistency of hue scalings over time (like over a few days). However, Boynton and Gordon (1965) report that their subjects made very reliable settings with mean correlation coefficients of 0.96.

References

Abramov, I., & Gordon, J. (1994). Color appearance: On seeing red—Or yellow, or green, or blue. *Annual Review of Psychology, 45,* 451–485. [PubMed]

Abramov, I., Gordon, J., & Chan, H. (1990). Using hue scaling to specify color appearances and to derive color differences. Perceiving, measuring, and using color. *Proceedings of SPIE, 1250,* 40–51.

Arend, L., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A, Optics and Image Science, 3,* 1743–1751. [PubMed]

Arend, L., Reeves, A., Schirillo, J., & Goldstein, R. (1991). Simultaneous color constancy: Patterns with diverse Munsell values. *Journal of the Optical Society of America A, Optics and Image Science, 8,* 661–672. [PubMed]

Augenstein, E. J., & Pugh, E. N., Jr. (1977). The dynamics of the Π_1 colour mechanism: Further evidence for two sites of adaptation. *The Journal of Physiology, 272,* 247–281. [PubMed] [Article]

Bäuml, K. H. (1999). Simultaneous color constancy: How surface color perception varies with the illuminant. *Vision Research, 39,* 1531–1550. [PubMed]

Bäuml, K. H. (2001). Increments and decrements in color constancy. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 18,* 2419–2429. [PubMed]

Boyaci, H., Doerschner, K., & Maloney, L. T. (2006). Cues to an equivalent lighting model, *Journal of Vision, 6*(2), 106–118, http://journalofvision.org/6/2/2/, doi:10.1167/6.2.2. [PubMed] [Article]

Boyaci, H., Doerschner, K., Snyder, J. L., & Maloney, L. T. (2006). Surface color perception in three-dimensional scenes. *Visual Neuroscience, 23,* 311–321. [PubMed]

Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision, 3*(8), 541–553, http://www.journalofvision.org/3/8/2/, doi:10.1167/3.8.2. [PubMed] [Article]

Boynton, R. M., & Gordon, J. (1965). Bezold–Brücke hue shift measured by color-naming technique. *Journal of the Optical Society of America, 55,* 78–86.

Boynton, R. M., & Olson, C. X. (1987). Locating basic colors in the OSA space. *Color Research and Application, 12,* 94–105.

Brainard, D. H. (1998). Color constancy in the nearly natural image: 2. Achromatic loci. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 15,* 307–325. [PubMed]

Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image: I. Asymmetric matches. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 14,* 2091–2110. [PubMed]

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

Craven, B. J., & Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research, 32,* 1359–1366. [PubMed]

Delahunt, P. B., & Brainard, D. H. (2004). Does human color constancy incorporate the statistical regularity of natural daylight? *Journal of Vision, 4*(2), 57–81, http://www.journalofvision.org/4/2/1/, doi:10.1167/4.2.1. [PubMed] [Article]

De Valois, R. L., De Valois, K. K., Switkes, E., & Mahon, L. (1997). Hue scaling of isoluminant and cone-specific lights. *Vision Research, 37,* 885–897. [PubMed]

D'Zmura, M., & Lennie, P. (1986). Mechanisms of color constancy. *Journal of the Optical Society of America A, Optics and Image Science, 3,* 1662–1672. [PubMed]

Fairchild, M. D., & Reniff, L. (1995). Time course of chromatic adaptation for color-appearance judgements. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 12,* 824–833. [PubMed]

Foster, D. H. (2003). Does colour constancy exist? *Trends in Cognitive Sciences, 7,* 439–443. [PubMed]

Foster, D. H., Amano, K., & Nascimento, S. M. C. (2001). Colour constancy from temporal cues: Better matches with less variability under fast illuminant changes. *Vision Research, 41,* 285–293. [PubMed]

French, J. W. (1921). The interocular distance. *Transactions of the Optical Society, 23,* 44–55.

Halir, R., & Flusser, J. (1998). Numerically stable direct least squares fitting of ellipses. In V. Skala (Ed.), *Proceedings of the International Conference in Central Europe on Computer Graphics and Visualisation* (pp. 125–132). WSCG98: Plzen.

Ishihara, S. (1997). *Ishihara's test for colour deficiency.* Tokyo: Kanehara & Co.

Jameson, D., & Hurvich, L. M. (1989). Essay concerning color constancy. *Annual Review of Psychology, 40,* 1–22. [PubMed]

Judd, D. B., MacAdam, D. L., & Wyszecki, G. W. (1964). Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America, 54,* 1031–1040.

Kaiser, P. K., & Boynton, R. M. (1996). *Human color vision* (2nd ed.). Washington, DC: Optical Society of America.

Katz, D. (1911). *Die Erscheinungsweisen der Farben und ihre Beeinflussung durch die individuelle Erfahrung.* Leipzig: Verlag von J. A. Barth.

Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences of the United States of America, 100,* 9085–9089. [PubMed] [Article]

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 307–312. [PubMed] [Article]

Kuriki, I., & Uchikawa, K. (1996). Limitations of surface-color and apparent-color constancy. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 13,* 1622–1636. [PubMed]

Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America, 61,* 1–11. [PubMed]

Larson, G. W., & Shakespeare, R. (1996). *Rendering with radiance: The art and science of lighting and visualization.* San Francisco: Morgan Kaufmann Publishers, Inc.

Schultz, Doerschner & Maloney

Logvinenko, A. D., & Maloney, L. T. (2006). The proximity structure of achromatic surface colors and the impossibility of asymmetric lightness matching. *Perception & Psychophysics, 68,* 76–83. [PubMed]

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

Reeves, A. (1982). Exchange thresholds for green tests. *Vision Research, 22,* 961–966. [PubMed]

Rinner, O., & Gegenfurtner, K. R. (2000). Time course of chromatic adaptation for color appearance and discrimination. *Vision Research, 40,* 1813–1826. [PubMed]

Scheifler, R. W., & Gettys, J. (1996). *X Window system: Core library and standards.* Boston: Digital Press.

Shepard, R. N. (1994). Perceptual–cognitive universals as reflections of the world. *Psychonomic Bulletin & Review, 1,* 2–28.

Shevell, S. K. (2001). The time course of chromatic adaptation. *Color Research & Application, 26,* S170–S173.

Smithson, H., & Zaidi, Q. (2004). Colour constancy in context: Roles for local adaptation and levels of reference. *Journal of Vision, 4*(9), 693–710, http://www.journalofvision.org/4/9/3/, doi:10.1167/4.9.3. [PubMed] [Article]

Snyder, J. L., Doerschner, K., & Maloney, L. T. (2005). Illumination estimation in three-dimensional scenes with and without specular cues. *Journal of Vision, 5*(10), 863–877, http://www.journalofvision.org/5/10/8/, doi:10.1167/5.10.8. [PubMed] [Article]

Spectral Database, University of Joensuu Color Group, http://spectral.joensuu.fi/

Speigle, J. M., & Brainard, D. H. (1999). Predicting color from gray: The relationship between achromatic adjustment and asymmetric matching. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 16,* 2370–2376. [PubMed]

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

Troost, J. M., & deWeert, C. M. M. (1991). Naming versus matching in color constancy. *Perception & Psychophysics, 50,* 591–602. [PubMed]

Werner, A., Sharpe, T., & Zrenner, E. (2000). Asymmetries in the time-course of chromatic adaptation and the significance of contrast. *Vision Research, 40,* 1101–1113. [PubMed]

Wyszecki, G., & Stiles, W. S. (1982). *Color science: Concepts and methods, quantitative data and formulae.* 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]