Chapter 5

Experiments II: Color constancy and Hue Scaling

In the previous chapter color constancy was investigated assuming a simplistic two-dimensional model of the environment which we referred to as the Mondrian World. The series of experiments presented here serves two main purposes. First, we wish to study color constancy in an enriched three-dimensional environment that resembles natural scenes. Second, hue scaling will be introduced as an alternative method to measure color constancy. This method promises to represent aspects of human color constancy performance more adequately than traditional measures (Foster, 2003).

5.1 Introduction

5.1.1 Motivation

The term color constancy is used to describe the phenomenon of stable color appearance of surfaces with changing illumination (Jameson & Hurvich, 1989; Kaiser & Boynton, 1996). This conception 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; Kuriki & Uchikawa, 1996; Kraft & Brainard, 1999). 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).

As pointed out by Foster (2003), it is questionable if frequently used quantitative measures are appropriate instruments to determine the degree of color constancy experienced by human observers. Therefore, he claims that alternative
methods are needed to describe the characteristics of human color constancy more adequately. In this sense, the present study is intended as a first step toward an understanding of color constancy that is more directly related with the demands on observers in natural environments.

5.1.2 Measuring Color Constancy

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, while 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 (Maloney, 1999; Foster, 2003; Logvinenko & Maloney, 2006). 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 used reliably 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, a disadvantage of this task is the lacking comparability of the color categorization data and quantitative measures and the potential that the conclusions drawn depend on an arbitrary choice of categories.

In the experiments reported here, hue scalings will be introduced as an alternative method to investigate color constancy (Boynton & Gordon, 1965; Abramov & Gordon, 1994). 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 is potentially comparable to quantitative measures of color constancy. Therefore, hue scalings of observers will also be compared with corresponding achromatic settings. Moreover, the usage of a hue scaling task promises to eliminate disadvantages that are inherent in asymmetric matchings and achromatic settings.

5.1.3 The Speigle-Brainard conjecture

Another purpose of the present study is to test a generalization of the model proposed by Speigle and Brainard (1999) to link achromatic settings and asymmetric matching. Speigle and Brainard 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 the context of hue scaling it will be tested 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.

5.1.4 Models of the environment

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). Although main principles assumed to underlie color constancy have been identified presupposing this model of environment (McCann, McKee & Taylor, 1976; Arend & Reeves, 1986; Brainard & Wandell, 1992),
three-dimensional aspects of vision are largely ignored. A further disadvantage of this configuration is that more complex light-surface interactions are not taken into account and reported indices of color constancy are typically low. In the experiments reported here simulated three-dimensional stimuli were presented binocularly to more closely approximate natural viewing conditions (Maloney, 1999; Boyaci, Maloney & Hersh, 2003). These scenes are viewed binocularly and contain additional cues to the illuminant such as specular highlights or shadows. Previous studies showed the relevance of these cues for color constancy performance of observers (Yang & Maloney, 2001; Boyaci, Doerschner & Maloney, 2006). The usage 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 (Maloney, 2002).

5.1.5 The Role of Daylights

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, 2004b). According with these findings and the results from Experiment 1 reported in Chapter 4 it is not expected that illuminant adjustment is better for daylights compared with illuminations off the daylight locus.

5.1.6 The Role of Chromatic Adaptation

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 the second of the experiments reported in this chapter I was able to demonstrate, by accident, the strength of the isolated slow adaptation mechanism.

5.1.7 Hypotheses

From the discussion above we can derive the following hypotheses:

1. Color constancy performance of observers measured with hue scalings and achromatic settings should be comparable.

2. Observers’ adjustment to natural daylights should not be better than to red and green illumination.
3. The appearance of a virtual achromatic surface under test illumination can be predicted by hue scalings of observers for a set of chromatic surfaces.

4. Color constancy of observers should break down when no spatial cues are given.

5.2 General Methods

5.2.1 Stimuli

Visual display

The stimuli in the experiments 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 possible cues to the chromaticity and intensity of the light sources. A matte test patch was presented at the center of the scene. In Experiments 2 and 3 the scene consisted only of the isolated test patch in front of a black homogeneous background.

The scenes were illuminated by a mixture of a simulated diffuse and a simulated punctate source. In order to present the 3D-scenes stereoscopically, for each scene two images were rendered from two different viewpoints corresponding to the locations of the observer’s eyes in the virtual scene (Color Plate D.3, Appendix D).\(^1\) The scenes differed in the chromaticity of the test patch and in the color of the punctate source.

Test patch

In Experiments 1–3 a set of 16 different test patches was used that represented simulations of Munsell chips (Table B.4, Appendix B). All test patches had identical Munsell value and chroma of seven and four respectively. They differed only in hue and formed a color circle in the CIE-\(u'v'\) diagram (Figure 5.1).

To increase the accuracy of stimuli, the test surfaces were rendered separately from the rest of the scene. Rendering packages such as Radiance simulate light-surface interactions as product of the respective RGB-codes. This RGB heuristic (Maloney, 1999) does not always lead to adequate simulation of light-surface interaction. Therefore, the light signal that reached the eye from the test patch was computed according to the Lambertian model as described in Section 3.3. Recall that the spectral power distribution of the light signal from a Lambertian

\(^1\)An interocular distance of 6.3 cm was used as a standard 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). It was checked whether subjects could achieve fusion and, had any subject reported difficulty with stereo fusion, they would have been excluded from further participation in the experiment.
surface that is illuminated by a diffuse source and a punctate source is given by:

\[ L(\lambda) = S(\lambda) \cdot (E^P(\lambda) \cdot \cos \theta + E^D(\lambda)), \]

where \( \lambda \) indexes the wavelength of light in the visible spectrum and \( \theta \) is the angle between the incident light from the punctate source and the surface normal. In our case \( \theta \) was always set to zero. The reflectance functions 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 \( \times \) 2.2cm which is equivalent to a visual angle of 2°22’ \( \times \) 1°48’\'. 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. The orientation of the test patch remained constant throughout all experiments.

**Light sources**

As noted above, simulations of a diffuse and a punctate light source were used. The diffuse source was always a neutral daylight (\( D_{65} \)). 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 6500\( K(D_{65}) \), 4000\( K(D_{40}) \) and 10000\( K(D_{10}) \) that appeared achromatic, yellow and blue to the observer respectively.

Additionally, two illuminants off the daylight locus were used that appeared red and green to the observer (Figure 5.2). The \( xy \)- and \( u'v' \)-chromaticities of
Figure 5.2: Chromaticities of the punctate illuminants in $u'v'$. Different symbols refer to different illuminations: filled square - blue (b), open square - yellow (y), filled circle - red (r), open circle - green (g), gray filled circle - D65.

the punctate sources are given in Table B.3 (Appendix B). For the five punctate sources spectral power distributions were computed from the CIE daylight basis functions (Wyszecki & Stiles, 1982). The calculation of the light signal that reached the eye of the observer from the test patch based on these distributions and the reflectance functions of the Munsell surfaces. The punctate source was situated behind the observer. The distance between the test patch and the punctate light source was 670cm. The position of the punctate source was held constant throughout all experiments. The punctate-total luminance ratio was always $\pi = 0.67$ (see Boyaci, Maloney & Hersh, 2003).

Software and apparatus

The stimuli were presented to the subjects stereoscopically on two Sony Trinitron Multiscan GDM-F500 21” CRT-screens (Figure 5.3). I ran the experiments under Red Hat Linux 6.1 using a Dell workstation with a 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 look-up tables for each monitor. The look-up tables were based on direct measurements of the luminance values on each monitor with a Pritchard PR-650 spectrometer. Finally, the corrected RGB triplets were translated to 24-bit graphics codes. The experiments were programmed by myself in the C language using the X Window
Figure 5.3: Apparatus. The observer is presented with a stereoscopic image which consists of two slightly different images displayed on two CRT-creens.

System, Version 11R6 (Scheifler & Gettys, 1996). The settings were made by the subjects using a mouse.

5.2.2 Task

The task of the subject in Experiments 1–3 was to set hue 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.

The decision to chose a rating scale with exactly six steps 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 non-zero setting on both scales of an opponent pair such as blue-yellow, then 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 one of $169 = 13 \times 13$ different categories. The method presented here allows much finer comparison than color naming procedures where only 11 categories are typically allowed.

In Experiment 4, subjects were asked to make achromatic settings for the test patch.

5.2.3 Procedure

Subjects practiced hue scaling in a training session before running in the experiments. Only subjects who had reliability in their settings of 0.8 and more took part in the study. Reliability was determined as follows. In the training session subjects repeated hue scaling settings for 16 surfaces under neutral illumination five times. For each subject, all pair-wise correlations between repeated measurements were transformed using Fisher’s r-to-z transformation. The subject’s reliability was defined as the correlation corresponding to the mean of the transformed values. I 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.

In the experiments, different illuminant conditions were generally 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 scaling and also to stabilize the adaptational state of the subject. In each trial the scene was presented to the subjects and they carried out the task described above.

In Experiment 4, subjects made achromatic settings for the test patch instead of hue scalings. Between two trials a black screen was presented to the subject for 1s 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 minutes 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.

5.2.4 Data analysis

First an overview of the general methods of data analysis will be given. The same analyses were carried out for Experiments 1–3.
Absolute hue scalings

As a first step, the absolute hue scalings of subjects under different illumination conditions were analyzed. For each subject and each surface mean hue scalings were calculated. These data were plotted 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 it is expected that 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 a second step of the analysis, mean hue scalings for each subject in each illumination condition were transformed to \( u'v' \)-chromaticity coordinates. This transformation makes it easier to compare the present results with those from previous studies in terms of color constancy indices. The procedure of obtaining these indices is explained next.

Subjects made non-zero settings for a given test surface only on one of the blue or the yellow scale but not both, and on only one of the red or the green scale but not both. Therefore, for each given subject and condition, blue and yellow ratings were combined 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 rating \(0\) on the blue) would become \(a_{BY} = 1.7\). Similarly, red and green hue ratings were combined 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 \(a = (a_{BY}, a_{RG})'\) in a two-dimensional opponent space, \(A\).

Let \(u\) 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 \(A\) and \(u'v'\)-space is affine but perturbed by judgment error. That is, we assume that there is a \(2 \times 2\) matrix and a column vector \(c\) such that

\[
\mathbf{u} = \mathbf{M} \mathbf{a} + \mathbf{c} + \mathbf{\varepsilon},
\]

where \(\mathbf{\varepsilon} : \mathcal{N}(0, \sigma^2)\) is Gaussian judgment error with mean \(0\) and variance \(\sigma^2\). 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. This data was used to estimate the six parameters of the transformation (the four elements of \(\mathbf{M}\) and the two
elements of \( c \) for each subject. The set of parameters was chosen so that 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) proposed a model to predict achromatic settings from asymmetric matching. 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, the centroid of all of the surface patches under a single light will be used as reference surface. It is not assumed that this surface, which the subject never sees, is precisely achromatic but, given the construction of the stimuli, it will be nearly so.

I tested the Speigle-Brainard conjecture in the following way. First, all opponent data of subject \( i \) was translated to \( u'v' \)-coordinates using the specific transformation that was given by \( M_i \) and \( c_i \). Then optimal ellipses to the \( u'v' \)-data of each subject for each illuminant were fitted (Halir & Flusser, 1998). Finally the parameters of the ellipse for neutral illumination data were compared with those of the ellipses for the chromatic illumination data.

The analysis of the parameters of the ellipses has the advantage 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.

I next describe how area, eccentricity and orientation were measured. 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:

\[
e = \sqrt{1 - \frac{b^2}{a^2}}.
\]

The orientation of the ellipse is given by the radian angle \( \theta \) 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**

Color constancy indices were calculated in order to compare performance of observers across different tasks. The rationale behind the determination of indices differed for the hue scaling and the achromatic setting task.
First, the derivation of indices from transformed hue scalings of subjects will be described. 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, the centroids of test surfaces and settings in $u'v'$ were used to calculate color constancy indices. The indices have the form:

$$CI = 1 - \frac{|m_{D65} - m_{data}|}{|m_{D65} - m_{test}|},$$

(5.4)

where $m_{D65}$ is the centroid of the surfaces under neutral illumination, $m_{test}$ is the centroid of the surfaces under test illumination and $m_{data}$ 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).

In Section 4.2.6 it was described how to derive an index of color constancy performance from subject’s achromatic settings (see also Arend, Reeves, Schirillo & Goldstein, 1991). In case of perfect color constancy, we expect the achromatic setting of the subject to coincide with the chromaticity of the test illuminant. A subject has no color constancy if the achromatic setting under chromatic test illumination is identical with the chromaticity of neutral daylight $D65$. Roughly speaking, the color constancy index sets the shift of the subject’s achromatic setting in relation to the illuminant shift. The index is 1 in case of perfect constancy and 0 in case of no constancy.

### 5.3 Experiment 1: 3D Scenes – Hue Scalings

#### 5.3.1 Methods

**Stimuli and apparatus**

In Experiment 1, the observers were presented with simulations of 3D scenes rendered under five different illuminations. The apparatus and stimuli that were used in Experiment 1, namely the visual display, the set of test patches and the simulated light sources, were already described in the General Methods Section.

**Task and procedure**

In Experiment 1, the observers were asked to set hue scalings for the test patch in the center of the scene. The details of task and procedure were already described
in the General Methods Section. Subjects were told that they are viewing a scene under a certain illumination.

Observers

Five subjects took part in Experiment 1. All subjects were paid undergraduate students who were not aware of the purpose of the study. All had normal color vision as tested with Ishihara color plates (Ishihara, 1997).

5.3.2 Results

Expected results

In the 3D-experiment subjects are enabled to use various cues to the illuminant. Therefore it is expected that hue scalings are constant 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. Consistent with previous empirical findings (Brainard, 1998; Delahunt & Brainard, 2004b) it is not expected that color constancy performance will be better for natural daylight 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 5.4). 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 5.5A).

For each subject, the fitted ellipses of the neutral illuminant data were compared 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 5.6–5.7). The analysis of the orientation parameter \( \theta \) should be interpreted with care as the estimated ellipses of one subject (XH) were almost circular which resulted in unreliable estimates of \( \theta \).

In general, small systematic deviations occurred for the orientation of the ellipse in the yellow and green illuminant conditions (Figure 5.8). The analysis of the ellipses indicates that the Speigle-Brainard conjecture holds also for hue
CHAPTER 5. EXPERIMENTS II

Figure 5.4: Mean hue scalings of all subjects and of subject AHR under red punctate illumination as polar plots. Surfaces are given in Munsell notation. Different line styles and colors indicate the four scales: solid black – blue, solid gray – yellow, bold dashed – red, dotted – green. The first row gives results corresponding to perfect color constancy, i.e. 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 5.5: Transformed mean hue scalings of all subjects and of one subject (AHR) 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, open 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 (open squares).
Figure 5.6: Eccentricities of the estimated ellipses for neutral and chromatic illuminant conditions. Each symbol refers to a different chromatic illuminant (blue – filled square, yellow – open square, red – filled circle, green – open circle). The eccentricities of a given subject lie on a line parallel to the ordinate (from left to right: AHR, XH, ALL, LGS, AST). The standard deviations for the conditions are shown in the cross at the right part of the figure.

Figure 5.7: Area of the estimated ellipses for neutral and chromatic illuminant conditions. Each symbol refers to a different chromatic illuminant (blue – filled square, yellow – open square, red – filled circle, green – open circle). The area values of a given subject lie on a line parallel to the ordinate (from left to right: AHR, XH, ALL, LGS, AST). The standard deviations for the conditions are shown in the cross at the right part of the figure.
Figure 5.8: Orientation ($\theta$) of the estimated ellipses for neutral and chromatic illuminant conditions. Each symbol refers to a different chromatic illuminant (blue – filled square, yellow – open square, red – filled circle, green – open circle). The $\theta$ values of a given subject lie on a line parallel to the ordinate (from left to right: XH, AST, LGS, AHR, ALL). The standard deviations for the conditions are shown in the cross at the right part of the figure.

scaling data. The constancy indices that were calculated from these transformed data lay between 0.58 and 0.91 with median 0.705 (Table C.5, Appendix C). This indicates good to excellent color constancy (Figure 5.9). The results from Experiment 1 also indicate that color constancy performance is not better for daylights than for artificial red and green illuminations.

5.4 Experiment 2: Blocked Control

5.4.1 Methods

Stimuli and apparatus

In Experiment 2, the isolated test patch was presented in front of a black background. All stimuli of one illumination condition were shown as a block. The apparatus, the set of test patches and the simulated light sources in Experiment 2 were as described in the General Methods Section. The same rationale as in Experiment 1 was used to calculate the light signals of the test patches.
Figure 5.9: Color constancy indices for subjects XH (top), AHR (center) and the mean data of all subjects (bottom) under four chromatic illuminant conditions and each of the Experiments 1–4. Different shadings refer to different experiments: Dark gray bars – hue scalings 3D (Exp. 1), light gray bars – blocked control (Exp. 2), white bars – randomized control (Exp. 3), black bars - achromatic settings (Exp. 4). Note that the rationale the determination of indices is based upon differs for the hue scaling and the achromatic setting task. Therefore, interpretation and comparison of indices across different tasks should be done carefully.
CHAPTER 5. EXPERIMENTS II

Task and procedure

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

5.4.2 Results

Expected results

In designing Experiment 2, it was originally assumed that no cue to the illumination had been given to the subject. Therefore, mean hue scalings of the subjects were expected to depend only on the color codes of the surface under each illumination. In this sense, subjects should 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 neither cues to interpret the stimulus configuration as a surface under a illumination nor was this possible interpretation mentioned explicitly.

Surprisingly, the scalings of the subjects in this experiment were not simply determined by the color codes of the test surface (Figure 5.4, Figure 5.5). For example, when a surface was used 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 appears as if subjects had developed a kind of partial constancy (Table C.5, Appendix C; Figure 5.9). 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’. 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 can draw from Experiment 2. It should be noted, however, that in Experiment 2, there was no visible background. Consequently, in the present experiment there is no possibility that the effect depended in some way on the presence of a background. The most plausible explanation for my results and those of Smithson and Zaidi is that the visual system is averaging chromaticities from multiple surfaces across time and using this average as an estimate of the chromaticity of the illuminant in estimating surface colors. We return to this point in the discussion.

5.5 Experiment 3: Random Control

5.5.1 Methods

Stimuli and apparatus

As in Experiment 2, only 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. The apparatus, the set of test patches and the simulated light sources in the present experiment were as described in the General Methods Section. The same rationale as in Experiment 1 was used to calculate the light signals of the test patches.

Task and procedure

Subjects set hue scalings for the test patch as in Experiment 1. In this experiment, subjects were not told that the stimuli might be interpreted as a surface under a certain illumination. The procedure of Experiment 3 differed in the following respect from those of Experiments 1 and 2. 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.
5.5.2 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, mean hue scalings of the subjects are expected 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 5.4). 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 5.5). 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 C.5, Appendix C; Figure 5.9).

5.6 Experiment 4: 3D Scenes – Achromatic Settings

5.6.1 Methods

Stimuli and apparatus

In Experiment 4, the observers were presented with simulations of 3D scenes rendered under either one of the five different illuminations. The visual display was identical to Experiment 1 except for the color of the test patches. Achromatic settings for the test patches were obtained at four fixed luminance levels (8, 12, 18, 27 cd/m²). The initial chromaticity of a given test patch was chosen at random from a line D65 – test illuminant. The apparatus was identical to the one used in the previous experiments which was already described in the General Methods Section.

Task and procedure

In Experiment 4, observers made achromatic settings for the test patch. They were asked to adjust the chromaticity of the test patch until it appeared neither bluish nor yellowish and/or neither reddish nor greenish to them. Results from previous studies have shown that achromatic settings of subjects are essentially aligned on a line that is determined by the direction of the adaptational stimulus.
in chromaticity space (Bäuml, 1994; Rinner & Gegenfurtner, 2002). Therefore, the subjects made their settings on only one line which was defined as the line D65 – test illuminant in $xy$-chromaticity diagram. This modification considerably reduced the complexity of the task. The subjects reported being able to always find satisfying settings. Under D65 punctate illumination observers used two axes to make their adjustments that corresponded to the two opponent channels $RG$ and $BY$. Subjects practiced achromatic settings in a preliminary training session before running in the experiment. Like in Experiment 1, they were told that they are viewing a scene under a certain illumination.

Different illuminant conditions were blocked into different sessions. Subjects repeated settings for each of the four different luminance levels of the test surface five times resulting in 20 settings per block. In order to control the adaptational state of the subject, the first test patch appeared only after 2 min of presentation of the simulated 3D-scene. There were no time constraints. One session took about 15 minutes on the average. 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

Four of the five subjects from the previous experiments took part in Experiment 4.

5.6.2 Results

Expected results

Like in Experiment 1, subjects are provided with various cues to the illumination from the simulated 3D-scene. Therefore, achromatic settings of subjects are expected to be located closely at the chromaticity of the punctate light source. The degree of observers’ color constancy should be comparable to results from previous studies using a similar paradigm (Kraft & Brainard, 1999; Rinner & Gegenfurtner, 2002). The degree of color constancy reported in these studies typically lies in the range 60–90%. According to the outcome of the experiments reported earlier in this work, color constancy performance is not expected to be better for natural daylights than for red and green illuminations displaced from the daylight locus.

Experiment 4: results

The interpretation of results in the achromatic setting task in terms of color constancy follows a rationale which differs from the hue scaling task. In the present case, observer has perfect color constancy if chromaticities of his achromatic settings coincide with the chromaticity of the test illuminant. In contrast, it is
expected that chromaticities of the observers achromatic settings do not shift with the illuminant in case of missing color constancy.

Figure 5.10 shows achromatic settings of two observers and the pooled data of all observers. The results indicate strong differences across subjects which are probably due to the experience with the achromatic setting task. This interpretation emerges from comparison with corresponding hue scaling data of these observers and the different degree of experience with the achromatic setting task. Only one subject, XH, who had participated in comparable psychophysical experiments before showed a pattern of results that one would expect for an observer with good color constancy. Achromatic settings of subjects AHR, AST and LGS who were not experienced with the task varied only little across different illuminants.

As a consequence, observers’ adjustment to the illuminant measured with achromatic settings was generally much smaller compared with the hue scaling data obtained in Experiment 1. The degree of color constancy was expressed in terms of a constancy index. In Figure 5.9 constancy indices of observers XH, AHR and indices determined from the pooled data of all observers are compared across the different experiments. Constancy indices of individual observers are given in Table C.5 (Appendix C). Only the indices of subject XH are comparable across the two different tasks. There is no evidence in the data for a systematic difference between the adjustment to daylights and to lights off the daylight locus.

5.7 Discussion

In the present study, color constancy performance was measured using two different paradigms: a hue scaling task and achromatic settings. In the first experiment, subjects showed constant hue scalings for a given test patch with changes of the illuminant. The degree of color constancy observed in the first experiment was comparable to that found in previous studies (Brainard, 1998; Foster, Amano & Nascimento, 2001a). 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 (Exp. 3). In Experiment 4, color constancy performance of observers measured with achromatic settings was generally much smaller in comparison with the hue scaling data from Experiment 1. This finding was probably due to the lack of experience of most subjects with the achromatic setting task.

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
Figure 5.10: Achromatic settings of subjects under chromatic punctate light sources in $xy$-chromaticity diagram. The top row shows data of an experienced subject (XH, top left) and an inexperienced subject (AHR, top right). Achromatic settings of all subjects are shown below (ALL, bottom). Different symbols refer to different illuminant conditions: filled square - blue, open square - yellow, filled circle - red, open circle - green. The asterisks denote chromaticities of the illuminations which correspond with settings of an perfectly color constant observer.
phenomenological sense. Furthermore, the data shows that a measure of color constancy based on hue scaling is comparable to 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 are needed that serve this purpose (Foster, 2003). The present study is a first step toward developing and validating measures of color constancy that on one hand are more closely related to the phenomenon of color constancy and on the other hand can be used to quantify the degree of color constancy.

Second, over the range of stimuli considered, the results indicate that the Speigle-Brainard conjecture (generalized to hue scaling) held. Speigle and Brainard (1999) were able to predict asymmetric matchings from achromatic settings. In the analysis of the data ellipses it was shown that the centroid of the ellipse under a given chromatic illumination (that is 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 (1999).

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 5.4 illustrates the differences between the three experimental hue scaling 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, Brunt & Speigle, 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 until 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. In this terminology the present study refers to successive color constancy. My 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 two minutes (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 contrast are confounded, the two suggested processes
of chromatic adaptation might be related to these two classes of phenomena (Rinner & Gegenfurtner, 2000). In Experiment 2, only isolated test patches were presented 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 relations between the phenomena of successive and simultaneous contrast, 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 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 serve 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. The present results and those of Smithson and Zaidi suggest that a similar averaging process occurs even when the change from one surface to the next is not due to an eye movement but is under experimenters control. There is considerable evidence for slow adaptation processes to exchange of 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.