12-2023

Novel Psychophysics of Cognitive Color Appearance Phenomena

Luke Hellwig
lfh9787@rit.edu

Novel Psychophysics of Cognitive Color Appearance Phenomena

by

Luke Hellwig

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Munsell Color Science Laboratory
College of Science
Rochester Institute of Technology

December 2023

Signature of the Author _____

Accepted by _____
Coordinator, Ph.D. Degree Program Date

MUNSELL COLOR SCIENCE LABORATORY
COLLEGE OF SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Luke Hellwig
has been examined and approved by the
dissertation committee as satisfactory for the
dissertation required for the
Ph.D. degree in Color Science

_____
Dr. Mark D. Fairchild, Dissertation Advisor

_____
Dr. Jeff Pelz, Chair

_____
Dr. Michael Murdoch

_____
Dr. Charles Poynton

_____
Date

Novel Psychophysics of Cognitive Color Appearance Phenomena

by

Luke Hellwig

Submitted to the
Munsell Color Science Laboratory
in partial fulfillment of the requirements
for the Doctor of Philosophy Degree
at the Rochester Institute of Technology

Abstract

Color appearance models can be used as an experimental design and analysis tool to better study high-level color perception involving brightness and chromatic adaptation. We investigate the experimental basis for the Helmholtz-Kohlrausch effect, the contribution of chromatic intensity to our perception of brightness. A new experimental method for measuring the brightness of chromatic colors leads to a model of the Helmholtz-Kohlrausch effect, which we use to extend CIECAM16, the color appearance model recommended by the Commission Internationale de l'Eclairage. The model is tested on high-dynamic-range images. The process of building this model also leads to several improvements in CIECAM16 itself.

We then investigate how color appearance models can similarly be used to design experiments and model cognitive mechanisms of discounting the color of illumination. Two different experimental modalities are used to separately measure sensory and cognitive mechanisms of chromatic adaptation to heterochromatic lighting conditions. The results provide insight into this cognitive phenomenon while also setting new benchmark processes for studies that use asymmetric color matching.

Acknowledgements

I would like to thank Mark Fairchild for his genuine interest in my ideas and for trusting me to follow them. Thanks to Dale Stolitzka for his strong advocacy for this research. Thanks to my committee members for their reliable support, helpful feedback, and patience. Thank you to Susan Farnand, Michael Murdoch, Dave Wyble, Jeff Pelz, and Roy Berns for teaching me color science. Thanks to Val Hemink and Steph Livingston-Heywood for being on top of everything. Completing this degree would have been truly impossible without the wonderful community of color science students at RIT; thank you to all those with whom I shared time, ideas, and struggles. To my friends and family: thank you for the joy you bring to my life.

Sections 3.2, 3.3, 3.4, 3.5 were funded by Samsung Semiconductor, Inc.

This thesis is dedicated to the teachers who helped me find color: John Mueller, Marty Baylor, and Silya Kiese.

Contents

List of Figures

Chapter 1

Introduction

Color appearance models seek to quantify and predict the human perception of visual stimuli. Whereas color spaces (i.e., CIELAB) only use physical information about the stimulus and white point (i.e. tristimulus values) to describe color, color appearance models account for the effect of the viewing environment in their modeling of visual perception. Their added complexity gives color appearance models greater accuracy in mathematically describing the perceptual experience generated by our visual system. This complexity also has its downsides; it discourages adoption in industrial applications compared to easier-to-use models. Additionally, achieving the required level of rigor in these models compared to their competitor color spaces is made more difficult by the complexity of the operations involved. This thesis highlights the potential use of color appearance models (CAMs) in designing and analyzing psychophysical experiments and also shows how this application of a CAM can create a useful feedback loop by which the results of CAM-based experiments can reveal areas of improvement in models, such as CIECAM16, which is the current model recommended by the Commission Internationale de l'Eclairage.

These models' potential as tools of experiment design is especially relevant in cases where the object of study is cognitive visual phenomena that sit downstream of the early stages of visual processing modeled by CAMs. My thesis explores two such high-level phenomena: perceptions of brightness (Chapter 3) and cognitive chromatic adaptation (Chapter 4). Our research into brightness metrics revolves around the Helmholtz-Kohlrausch (H-K) effect, where visual stimuli appear brighter as the chromatic saturation increases even if the stimuli do not contain more light. This chapter begins with an example of the feedback loop mentioned above, where during our efforts to model the H-K effect, we found a flaw in the current construction of equations for perceptual attributes in CIECAM16 (Section 3.1). In Section 3.2, we then report the results of an experiment that demonstrated that the discrepancy be-

tween perceived brightness and measures of achromatic response are due to cognitive factors related to the experimental task used to measure either brightness or luminance. Sections 3.3 and 3.4 cover our efforts to build a CAM-based model of the H-K effect, first by using existing brightness-matching data (Section 3.3) and then by designing our own experiments (Section 3.4). The use of color appearance modeling in our experimental design and in our data analysis allows us to build a model of the H-K effect with more theoretical grounding than other models, and which performs well on other recent brightness-matching data sets. Finally, in Section 3.5, we test the performance of the model on high-dynamic-range images.

Our chromatic adaptation research focuses on how observers cognitively discount the color of the illumination when they experience partial sensory adaptation. A preliminary study (Section 4.2) measured baseline sensory adaptation to our chosen experimental setup of a heterochromatic viewing environment. Section 4.3 then builds on this baseline through an asymmetric hue matching experiment. The implementation of this experiment relied heavily on color appearance modeling and demonstrated the benefit of using CAMs in experimental design. The results shed light on the degree to which observers cognitively compensate for the illuminant and highlight how further work could determine how best to model cognitive chromatic adaptation with a CAM framework.

Chapter 2

Background

Psychophysics is the study of the psychological effect of physical input. In our case, visual psychophysics involves presenting observers with visual stimuli and asking them to complete a task (e.g. "Adjust these stimuli until they are equally bright.") or make a judgment (e.g. "Which of these stimuli are brighter?"). The observers' performance depends both on the sensory neural calculations performed by the visual processing system in the brain and higher-level cognitive functions that interpret the question and make judgments based on the information generated by the visual system's early stages of visual processing. This thesis proposes that quantitative modeling of these early stages of visual color processing can provide scaffolding on which we can better design psychophysical experiments to measure cognitive color appearance phenomena and better model the results of such experiments. Color appearance models provide the scaffolding that will be implemented in this thesis to study color appearance phenomena related to chromatic adaptation and judgments of brightness.

2.1 Early Visual Processing in Color Appearance Models

Color appearance models are quantitative systems that seek to predict how light will be perceived by human observers. They use information about the physical stimulus (as represented by CIE XYZ tristimulus values) alongside information about the condition in which the stimulus is viewed in order to predict scales of perceptual attributes. To do so, the models tend to mathematically mimic the processing that occurs in the early stages of vision, mostly in our retina. This processing can be summarized in four steps: trichromacy, adaptation, compression, and opponency. This section provides a brief introduction to these

visual processes, using the CIECAM16 color appearance model, which has recently been recommended by the Commission Internationale de l'Eclairage (CIE), as an example [1].

2.1.1 Trichromacy

The first stage of visual processing is the absorption of light by three spectrally-biased classes of cone photoreceptor cells in the retina; this is why we use tristimulus values to specify visual stimuli. CIECAM16, like many other color appearance models, estimates cone signals—R, G, B—from CIE XYZ tristimulus values using a matrix transform:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0.401288 & 0.650173 & -0.051461 \\ -0.250268 & 1.204414 & 0.045854 \\ -0.002079 & 0.048952 & 0.953127 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{2.1}$$

The effective spectral sensitivities of the CAM16 cone values are shown in Figure 2.1 alongside physiological cone fundamentals from Stockman and Sharpe [2, 3]. Note that the CAM16 cone signals are not intended to directly replicate physical cone sensitivities. Rather, the transformation from XYZ to RGB in CIECAM16 was optimized for performance on color matching and color appearance data [4].

2.1.2 Adaptation

The photoreceptor cells and subsequent retinal processing cells adapt to the amount of light in the visual field [5]. This adaptation occurs on a per-channel basis, allowing our visual system to adjust to the color of the illumination and maintain (at least partially) the achromatic appearance of non-spectrally selective surfaces. The white balance of a camera mimics this process, which in human vision is called chromatic adaptation.

In most color appearance models, chromatic adaptation is modeled as a linear gain control that applies to each channel independently. This mathematical simplification is called the Von Kries model of chromatic adaptation. In CIECAM16, this takes the mathematical form:

$$\begin{bmatrix} R_c \\ G_c \\ B_c \end{bmatrix} = \begin{bmatrix} D\frac{Y_W}{R_W} + (1-D)\frac{Y_{WR}}{R_{WR}} & 0 & 0 \\ 0 & D\frac{Y_W}{G_W} + (1-D)\frac{Y_{WR}}{G_{WR}} & 0 \\ 0 & 0 & D\frac{Y_W}{B_W} + (1-D)\frac{Y_{WR}}{B_{WR}} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.2}$$

The chromatic adaptation of each channel is accomplished by division by the respective cone signal generated by a white stimulus (R_W, G_W, B_W), subject to the degree of adaptation, D, (between zero and one) and normalization to the relative luminance of the white point, Y_W

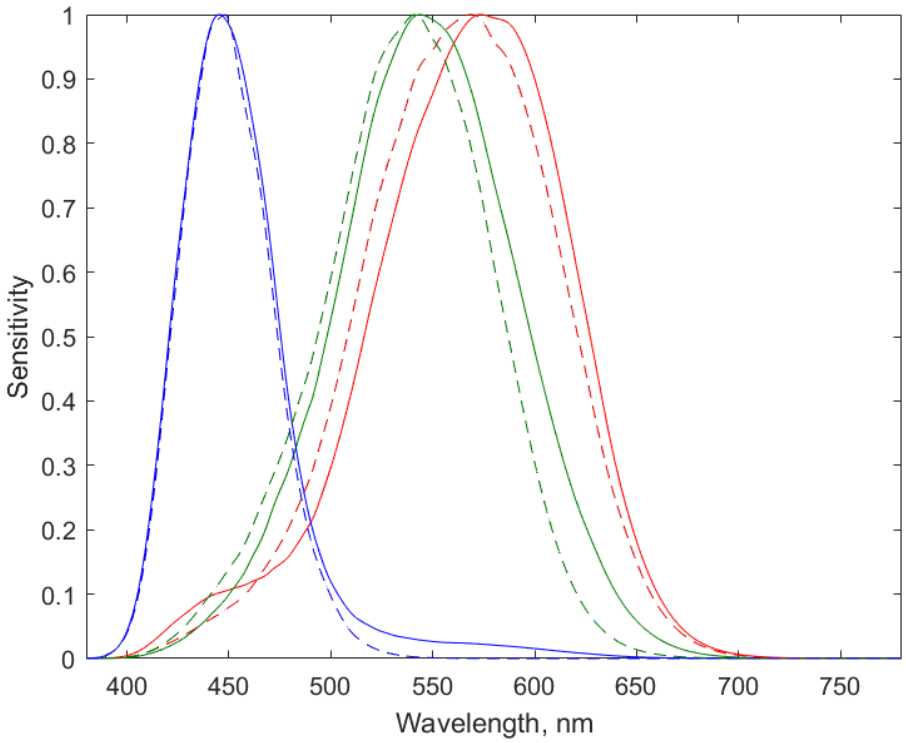


Figure 2.1: (solid lines) CIECAM16 RGB pseudo-cone fundamentals compared to (dashed lines) physiologically-based cone fundamentals derived by Stockman [2, 3]

(typically 100). If the degree of adaptation is less than one, then the observer's white point deviates from the color of the illumination towards their internal resting reference white point (R_{WR}, G_{WR}, B_{WR}), which in CIECAM16 is arbitrarily and incorrectly set to the equal energy illuminant ($R_{WR} = G_{WR} = B_{WR} = 100$) [6]. In CIECAM16, the adaptation of the visual system to the overall level of light in the scene is accomplished simultaneously with compression, which is discussed next.

2.1.3 Compression

We are capable of detecting stimuli across three to four log units of light intensity within a single scene [7, 8]. The human visual system uses a nonlinear response to light to be able to see detail across such scenes. (Local adaptation of retinal cells also contributes to the dynamic range of the human visual system [9].) In CIECAM16, this nonlinear response is modeled by a piecewise function anchored by the following hyperbolic equation:

$$R_a = 400 \cdot \left(\frac{\left(\frac{F_L \cdot R_c}{100}\right)^{0.42}}{\left(\frac{F_L \cdot R_c}{100}\right)^{0.42} + 27.13} \right) + 0.1 \tag{2.3}$$

R_c is the chromatically-adapted cone signal from Equation 2.2 and F_L is a term that nonlinearly depends on the adapting luminance. Below cone values of 0.26 and above 150, linear extensions of the hyperbolic function are used [1]. The relationship between stimulus luminance for an achromatic stimulus and the tone-compressed cone signal, R_a, is shown in Figure 2.2 for four luminance levels.

2.1.4 Opponency

The three chromatic channels of information from the cones are converted to an achromatic response channel and two opponent color channels before visual information is passed from the retina to the rest of the brain. These opponent dimensions form the basis of the higher level perceptual calculations performed by our visual cortex and are represented in CIECAM16 by chromatic dimensions a and b, which roughly correspond with red vs. green and yellow vs. blue color opponency, respectively:

$$a = R_a - \frac{12 \cdot G_a}{11} + \frac{B_a}{11} \tag{2.4}$$

$$b = \frac{(R_a + G_a - 2 \cdot B_a)}{9} \tag{2.5}$$

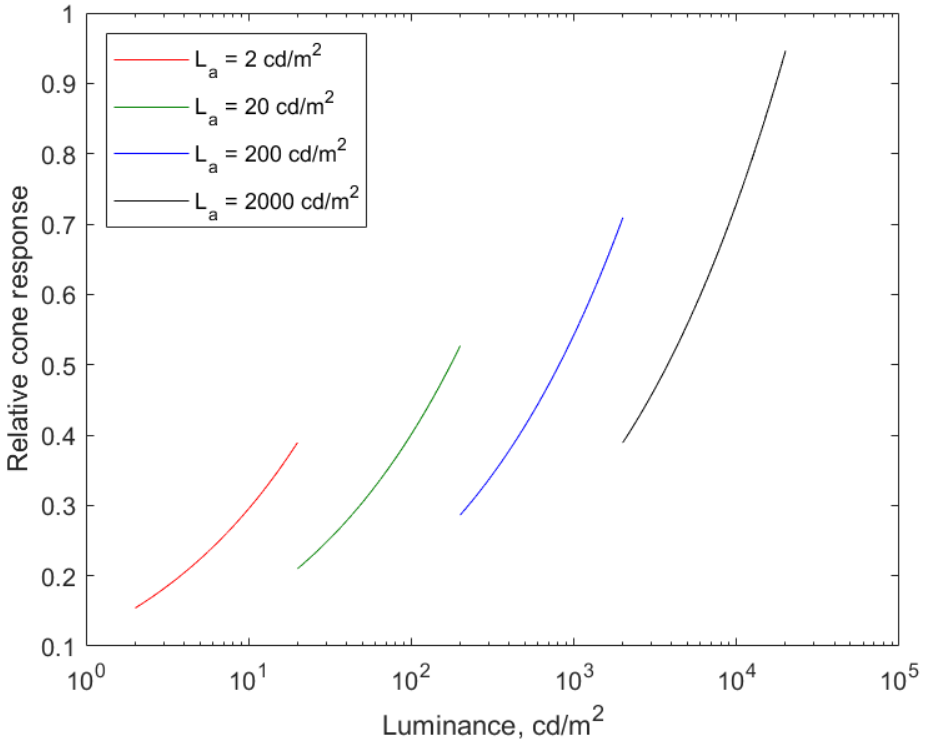


Figure 2.2: Relationship between the luminance of achromatic stimuli and tone-compressed cone response as modeled by CIECAM16 (Equation 2.3) at four levels of adapting luminance. Luminance factors range from near black to scene white (five times adapting luminance).

In CIECAM16, the adapted and tone-compressed cone signals are weighted and summed to derive an achromatic signal term, A:

$$A = \left(2R_a^{'} + G_a^{'} + 0.05B_a^{'} - 0.305\right) \cdot N_{bb} \tag{2.6}$$

where A_W is the achromatic signal for the reference white and N_{bb} is a background dependency defined by:

$$N_{bb} = 0.725 \left(\frac{Y_B}{Y_W}\right)^{-0.2} \tag{2.7}$$

Y_B is the luminance factor of the background and Y_W is the luminance factor of the reference white.

2.2 Perceptual Attributes and Cognition

Our visual cortex takes these three channels of visual information and uses them to generate our perceptual experience. Color scientists describe our perceptual experience of color using the following scales:

- Brightness: the attribute by which a stimulus appears to emit or reflect more or less light.
- Colorfulness: the chromatic intensity of a stimulus.
- Hue: the attribute of color appearance relating to color names such as red, green, blue, purple, or yellow.

The above terms are often considered to be absolute measurements of intensity. There is a lower bound (zero) on brightness and colorfulness, but neither has an inherent upper bound except for the total capacity of our neural system. These absolute perceptual attributes are often expressed relative to the scene conditions:

- Lightness: the brightness of a stimulus relative to the brightness of scene white.
- Chroma: the colorfulness of a stimulus relative to the brightness of scene white.
- Saturation: the colorfulness of a stimulus relative to its own brightness.

In the early stages of visual processing, color appearance models can mimic the well-documented neural structure of our retina. However, we do not have a well-defined biological analog for the calculations of perceptual attributes. Instead, we rely on color appearance data from

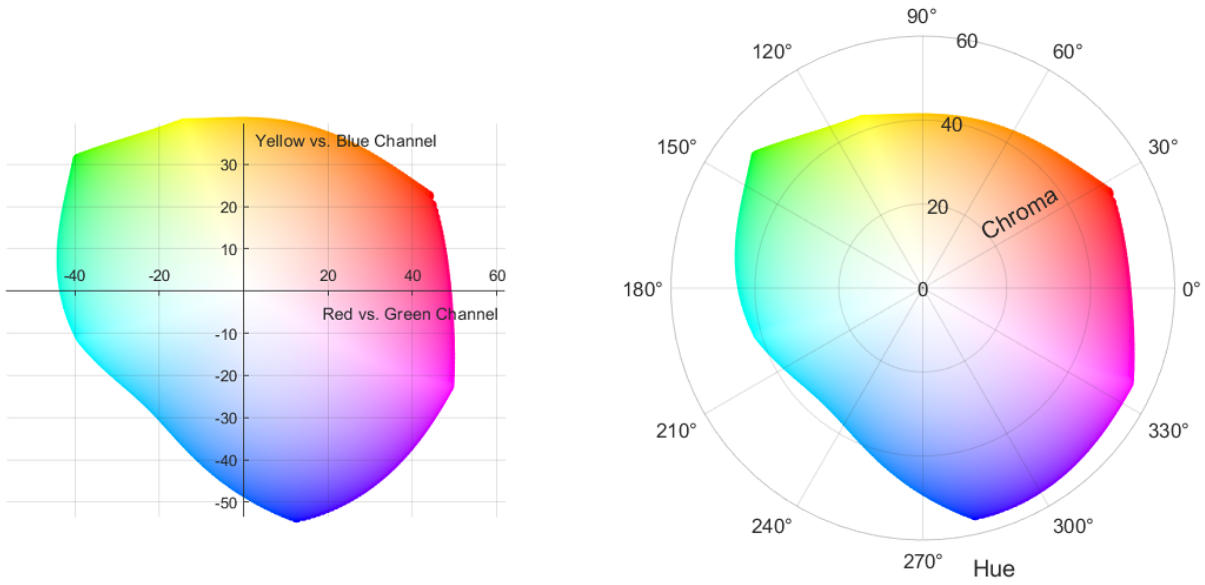


Figure 2.3: A simple method to estimate perceptual attributes from opponent color channels is to plot them on a Cartesian plane (left). Converting from rectangular to polar coordinates (right) leads to a simple measure of chroma (the radius) and hue (the polar angle), although a color appearance model will perform further calculations to improve these metrics' performance.

psychophysical experiments, such as the well-known LUTCHI dataset [10, 11], to try to determine the best mathematical approach. The methods by which CIECAM16 calculates these perceptual attributes will be covered in a later chapter in this dissertation.

In the most basic paradigm used by simple color spaces such as CIELAB, brightness and/or lightness can be derived directly from the achromatic signal sent to the visual cortex. Scales of chromatic intensity and hue can be estimated by plotting the opponent chromatic signals on a Cartesian plane and converting to polar coordinates, where the radius represents the chromatic intensity and the polar angle represents the hue (Figure 2.3). We will see that these simple mathematical approaches do not capture the cognitive processing that occurs on these information channels. This is especially true in the case of brightness, which receives input from both the chromatic and achromatic channels. Nonetheless, we will see that these preliminary scales form a useful foundation to model these cognitive calculations as we explore this phenomenon, known as the Helmholtz-Kohlrausch effect, in a subsequent chapter.

2.3 Cognitive Mechanisms of Chromatic Adaptation

Cognitive mechanisms are important in states of partial chromatic adaptation. Partial chromatic adaptation describes the state in which the ambient illumination does not appear neutral. The human visual system has a natural white point and as the color of the illumination changes from that preferred white point, the visual system does not completely compensate for the color of the illumination [6, 12–17]. Under states of partial chromatic adaptation, non-spectrally selective surfaces no longer appear neutral. For example, a piece of paper that appears white under daylight may look slightly yellowish when illuminated by an incandescent light. However, if an observer is asked the color of the paper, they may answer that the paper is white because they would expect the paper to appear white under illumination to which they fully adapted (which perhaps could also be called neutral illumination). This process of cognitively accounting for the color of the illumination in one's perception of the color of illuminated objects is called "discounting the illuminant" by color scientists [18, 19]. This cognitive "discounting the illuminant" is distinguished from the so-called sensory adaptation that occurs in the retina.

Chromatic adaptation is the first step following trichromacy in color appearance models, so the models are implicitly representations of sensory retinal adaptation. There is no provision later in the models to include cognitive adaptation nor is there any evidence of how it should be modeled. Currently, it is possible in models like CIECAM16 to assume that our cognitive mechanisms would follow the same path of chromatic adaptation as our sensory mechanisms and thus one could still use a von Kries model by setting the degree of adaptation to 1 (Equation 2.2). However, due to the inherent uncertainty in how the color of an object will change under a different color of illumination, it is theoretically impossible for our visual system to predict the color of an object under different illumination. Multiple spectral reflectances (called metamers) can lead to the same cone responses under a given illumination but these metamers will no longer match in color under a second illumination [20]. The potential for metameric illumination–where two different lighting technologies achieve the same color of illumination with different spectral power distributions–only adds to the uncertainty inherent in trying to predict the color of an object under a reference illumination given only its apparent color under a test illumination.

Nonetheless, there is a robust field of study within experimental psychology that explores how and if we maintain stable perceptions of object color appearance across changes in illumination. Psychologists refer to this perceptual phenomenon as color constancy [18, 21]. The degree and accuracy of color constancy reported in the literature varies greatly from experiment to experiment [21] and depends on the experimental design [22]. One

common experiment to test this phenomenon is to ask observers to make color matches across different illumination conditions (examples: [23, 24]; comprehensive list: [21]). At times, though, these studies fail to account for the uncertainty due to metamerism [23] and the term color constancy itself is potentially misleading for the same reasons [20]. However, such experimental designs may potentially provide the opportunity to study discounting the illumination since complete sensory adaptation to heterochromatic illumination is not possible. In a later chapter of this dissertation, we will explore how color appearance models can be employed to overcome the shortcomings of these earlier experiments and provide insight into this cognitive mechanism of chromatic adaptation.

Chapter 3

Brightness

This chapter begins with an analysis of how CIECAM16 predicts scales of perceptual attributes (Section 3.1). We uncover significant flaws in the model and propose improvements that will allow us to use CIECAM16's scales of perceptual attributes as the foundation for the rest of the experimental work in this thesis. In Section 3.2, we investigate how the cognitive effects of different experimental methods of measuring brightness lead to a discrepancy between measures of achromatic brightness and perceived brightness. This section reveals why we need a model that accounts for the chromatic contribution to our perception of brightness, known as the Helmholtz-Kohlrausch (H-K) effect. A model of the H-K effect based on extant experimental data is introduced in Section 3.3. Section 3.4 then follows this with a series of brightness matching experiments that allowed us to fit a new model of the H-K effect with better performance and better theoretical grounding. Finally, we test an application of the H-K model to high dynamic range images in Section 3.5.

3.1 Measures of Brightness, Lightness, and Chromatic Intensity in the CIECAM framework

3.1.1 Background

Brightness is the perceptual attribute by which a light source or reflective surface appears to emit or reflect more or less light [19, 25, 26]. Lightness is the brightness of a stimulus relative to the brightness of a white-appearing stimulus in a similarly illuminated area, also known as the reference white [19, 25, 26].

The perceptual attribute colorfulness describes the absolute chromatic intensity of a visual stimulus. Chroma and saturation are relative measures of colorfulness; chroma is

defined as the colorfulness of a stimulus relative to the brightness of similarly-illuminated white and saturation is defined as the colorfulness of a stimulus relative to its own brightness [19, 25, 26].

Much work has been done over past decades to model brightness, lightness, colorfulness, and chroma. This section analyzes the lineage and current state of the brightness, lightness, colorfulness, and chroma functions in two prominent color appearance models: CIECAM02 [27, 28] and CIECAM16 [1, 4]. It is worth noting that CIECAM02 and CIECAM16 are identical after the chromatic adaptation stage. Thus, they are treated interchangeably in this section.

As outlined in Section 2, widely used color appearance models follow a similar flow. First, given the CIE XYZ tristimulus specification of a stimulus, a model will predict the responses of the three types of cone cells in our retina which are the basis of color vision. (Note that these cone spectral responses are chosen for model performance and are not meant to represent biological cone spectral sensitivities.) The chromatic and luminance adaptation of the cone cells will be modeled using information about the viewing conditions. The adaptation of the cone cells represents the first nonlinearity between signal and light in these models. The adapted and compressed cone signals will then be weighted and summed to derive an achromatic signal, A, and opponent chromatic signals, a and b. For instance, in CIECAM02 and CIECAM16, this relationship is represented by

$$A = [2R_a^{'} + G_a^{'} + 0.05B_a^{'} - 0.305]N_{bb} \tag{3.1}$$

$$a = R_a^{'} - 12G_a^{'}/11 + B_a^{'}/11 \tag{3.2}$$

$$b = \frac{1}{9}(R_a^{'} + G_a^{'} - 2B_a^{'}) \tag{3.3}$$

where $R_a^{'}$, $G_a^{'}$, and $B_a^{'}$ are the adapted signals of the three cone types and N_{bb} is a background dependency [1, 4]. Lightness, J, and brightness, Q, are then derived from the achromatic signal. The chromatic signals are used to calculate chroma, C, saturation, s or t, and colorfulness, M. The equations used to calculate J, Q, C, and M are the subject of this section.

3.1.2 Brightness and Lightness

The equations for brightness, Q, and lightness, J, in CIECAM02 and CIECAM16 originate mostly from the Hunt appearance model, which underwent several iterations from the early 1980s to the mid-1990s [29–33]. Of particular interest is the revision to the equations for J

and Q that Hunt made in the early 1990s [32]. The original model had linear relationships between A, Q, and J [31]:

$$Q = N_1(A + M) - N_2 \tag{3.4}$$

and

$$J = 100\frac{Q}{Q_w} \tag{3.5}$$

where M is the colorfulness of the stimuli, N_1 and N_2 are factors that Hunt used to account for luminance dependencies related to work by Stevens and Bartleson [34], and Q_w is the brightness of white in the scene. Note that the use of M in the equation for Q was an attempt by Hunt to account for the Helmholtz-Kohlrausch effect. This dependency of Q on M was lost in the transition from the Hunt model to CIECAM97s, CIECAM02's precursor, and its presence or absence in the equations discussed in this section does not detract from the overall discussion.

In 1991, Hunt revised the equations for Q and J, introducing nonlinearities in each [32]:

$$Q = N_1\left[7\left(A + \frac{M}{100}\right)\right]^{0.6} - N_2 \tag{3.6}$$

and

$$J = 100\left(\frac{Q}{Q_W}\right)^z \tag{3.7}$$

where

$$z = 1 + \sqrt{\frac{Y_B}{Y_W}} \tag{3.8}$$

with Y_B and Y_W being the luminance factors of the background and white point, respectively. (Note: $M/100$ in Equation 3.6 is equal to M in Equation 3.4. Hunt changed the scaling of that variable between the two papers.)

The first substantial change in the model is the inclusion of a 0.6 power in converting from A to Q. Hunt offers no explicit justification for this modification, merely stating: "The different achromatic signal A in the revised model, leads to the following expression for Q," and then, "These formulae lead to values of Q that, at normal photopic levels, are very similar in the original and revised models"[32]. However, careful examination of the methods for calculating A in each model reveals no differences that would necessitate the inclusion of the 0.6 power nonlinearity. The only apparent nonlinearity prior to this stage in either model is the tone compression function for the cone signals, which is identical in both models [31, 32]:

$$f(I) = 40\frac{I^{0.73}}{I^{0.73} + 2} \tag{3.9}$$

There is neither a clear cause for the inclusion of the 0.6 power nor any evidence of equality in Q values between the two models, contrary to the claims made by Hunt in justification for the equations.

Hunt also introduces a pair of background dependencies in his 1991 model [32]: a multiplicative factor in the formula for A (N_{bb} in Equation 3.1) and a nonlinearity in the formula for deriving J from Q (z in Equation 3.7). The N_{bb} term predicts that the achromatic signal, A, of a given stimulus will increase as the luminance factor of the background decreases. This predicted increase in A is carried through to the brightness of the stimulus, Q, and the brightness of the reference white, Q_W, via Equation 3.6. However, the contribution of N_{bb} is canceled out in the formula for J, Equation 3.7, when Q is divided by Q_W. Thus Hunt needed the z term to also increase the lightness of a stimulus as the background luminance factor, Y_B, decreases. Since Q/Q_W varies from 0 to 1 in Equation 3.7, decreasing z as Y_B decreases (Equation 3.8) causes J to increase as Hunt desired.

The addition of the z exponent to Equation 3.7 provides a possible explanation for Hunt's addition of the 0.6 exponent to Equation 3.6. Hunt may have wanted to maintain the similarity between his previous model (Equations 3.4 and 3.5) and the equation for brightness published by Bartleson [34], which Hunt claims is equivalent to his model (see Appendix II of [31]). By including the 0.6 power in the conversion from A to Q, Hunt partially undoes the nonlinearity introduced in going from Q to J, making the overall conversion from A to J more similar to his original, linear relationships between these (Equations 3.4 and 3.5). However, if this was Hunt's motivation, there was no clear justification for separating the two nonlinearities between two steps of the model (Equations 3.6 and 3.7) instead of just applying them both in a single step, such as in the formula for J (Equation 3.7).

Hunt's decision to separate these two nonlinearities in his model has been propagated through the CIE-approved color appearance models for the past thirty years. The Hunt model was drawn heavily upon and formed the basis for the Q and J equations when CIECAM97s was developed to unify the various competing color appearance models of the 1980s and 1990s [19]. The two nonlinearities that are used in CIECAM97s to calculate J from A seem to be equivalent in function to Hunt's nonlinearities. Since CIECAM97s and the subsequent CIE color appearance models calculate J before Q (as opposed to Q before J in the Hunt model), both nonlinearities from Equations 3.6 and 3.7 are included in a single step:

$$J = 100 \left(\frac{A}{A_W} \right)^{c \cdot z}. \tag{3.10}$$

In CIECAM97s, c is set to either 0.525, 0.59, or 0.69 for dark, dim, or average surrounds, respectively. Thus, c carries similar values to the 0.6 power used in the Hunt model (Equation

3.6). Like in the Hunt model, z depends on the relative background luminance:

$$z = 1 + F_{LL}\frac{Y_B}{Y_W} \tag{3.11}$$

where F_{LL} is one for stimuli smaller than 4° of visual angle and zero otherwise. Then, deriving the calculation for Q from J from the Hunt model, CIECAM97s essentially inverts Equation 3.7, introducing a third nonlinearity to undo the z power that the Hunt model predicted to solely apply to J:

$$Q = \left(\frac{1.24}{c}\right)\left(\frac{J}{100}\right)^{0.67}(A_W + 3)^{0.9}. \tag{3.12}$$

The exponent in Equation 3.12, 0.67, is approximately the multiplicative inverse of z, 0.69, for a typical, mid-gray background ($Y_B = 20$), which seems to confirm our interpretation that the 0.67 exponent is merely an artifact of how the formula for Q was adapted from the Hunt model and was not based on visual data. Importantly, the derivation of these CIECAM97s formulas did not account for the intent behind the placement of these nonlinearities in the Hunt model. As discussed above, the z background dependency was most likely introduced by Hunt into Equation 3.6 to compensate for the fact that J has not been affected by Hunt's other background dependency, N_{bb}, which affected Q. Thus z was required by Hunt to only apply to J. With the order of J and Q calculation reversed in CIECAM97s, z now affects both J and Q, so Hunt's requirement of the placement of z in Equation 3.6 is no longer relevant for CIECAM97s. The focus in the derivation process of CIECAM97s on the mathematics of the Hunt model led to a literal inversion of Equation 3.6 to create Equation 3.12 without considering that the nonlinearity in Equation 3.6 is only present because Q is calculated before J in the Hunt model.

The basic structure of these equations introduced in CIECAM97s—two nonlinearities from A to J and then a third nonlinearity from J to Q—has been carried forward into the model's successors, CIECAM02 and CIECAM16 [1, 27]. In both models, the formula for J matches Equation 3.10, although z is now slightly different:

$$z = 1.48 + \sqrt{\frac{Y_B}{Y_W}}. \tag{3.13}$$

The nonlinearity to calculate Q from J was simplified from a 0.67 power to a square root:

$$Q = \left(\frac{4}{c}\right)\left(\frac{J}{100}\right)^{0.5}(A_W + 4)F_L^{0.25}. \tag{3.14}$$

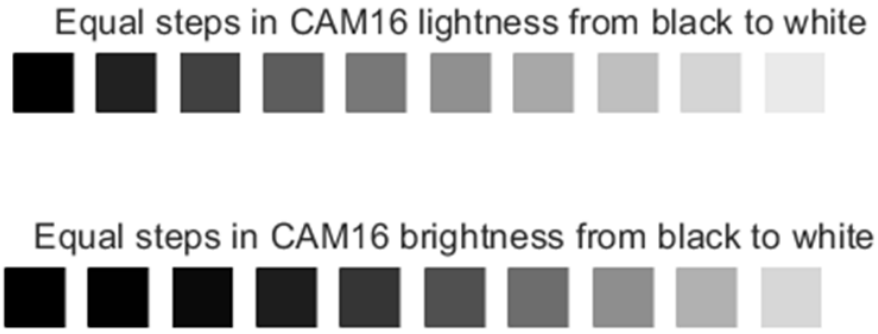


Figure 3.1: Approximate lightness and brightness scales calculated using the CIECAM16 formulas for J and Q (Equation 3.10 and Equation 3.14). The viewing conditions were assumed to be "average" with a reference white of 400 cd/m^2 and a white background.

There is little theoretical justification for the nonlinear relationship between lightness and brightness. No other color appearance model that predicts both brightness and lightness includes a nonlinear relationship between the two. A simple thought experiment highlights the nonlinearity's problematic nature. Imagine being shown an array of gray cards and being asked to choose the card that is halfway between black and white in terms of lightness. Then, you are asked to choose the card that is halfway between black and white in terms of brightness. The two cards chosen would be the same (allowing for some small degree of psychophysical uncertainty). But CIECAM16 claims that the same card will never be chosen, no matter the viewing conditions: the card that CIECAM16 predicts to have middle lightness will always be lighter than the card that CIECAM16 predicts to be middle brightness.

Practically, the nonlinear relationship between lightness and brightness seems to lead to incorrect predictions of brightness. Figure 3.1 shows neutral scales from black to white in equal steps of either lightness or brightness, as predicted by CIECAM16. (The viewing conditions were assumed to be "average" with a reference white of 400 cd/m^2 and a white background.) Figure 3.1 is a direct visual description of the nonlinearities of the lightness and brightness of CIECAM16 and it is clear from these figures that the brightness nonlinearity is faulty.

Fortunately, this error can be simply remedied by removing the nonlinearity in the equation for Q. Additional improvements to the performance of the brightness equation can be made by removing extraneous luminance and background dependencies in the equations for A and Q that duplicate dependencies which already exist in the formulas. The equations for

the achromatic signal, lightness, and brightness become:

$$A = 2R_a^{'} + G_a^{'}0.05B_a^{'} - 0.305 \tag{3.15}$$

$$J = 100\left(\frac{A}{A_W}\right)^{c \cdot z} \tag{3.16}$$

$$Q = \left(\frac{2}{c}\right)\left(\frac{J}{100}\right)(A_W) \tag{3.17}$$

Equation 3.17 restores the linear relationship between J and Q. The removal of N_{bb} from Equation 3.15 compared to Equation 3.1 achieves two ends. First of all, this background-dependent term is redundant, given that Q depends on J, and J depends on the relative background luminance factor via z in Equation 3.16. Hunt originally introduced N_{bb} in his 1991 model (Equations 3.6-3.8), where the z background dependency only applied to J, and thus the N_{bb} term was necessary to give brightness a background dependency. Now that z effects both J and Q, there is no need for N_{bb}. In fact, such a factor is undesirable since it only effects Q and not J. Additionally, N_{bb} behaves impossibly, approaching infinity as the relative background luminance approaches zero and producing clearly unrealistic predictions below a background luminance factor of 6%, which is the darkest background used by Hunt in deriving the term. Removing this explicit background dependency is consistent with the LUTCHI data, where all brightness scaling was done against the same gray background [11]. Thus, removing the N_{bb} factor returns the formulas to being a representation of the LUTCHI data, where brightness has the same background dependency as lightness.

The F_L factor in the CIECAM16 formula for Q (Equation 3.14) was introduced to CIECAM02 via a paper [35] that explored the use of a power function instead of a hyperbolic function to represent the cone dynamic response function in CIECAM97s. The inclusion of this F_L factor was not justified by specific data nor mentioned in the text. Nonetheless, while that paper's main proposal for a power function to serve as the cone response function was not adopted by CIECAM02, CIECAM02 did include this F_L factor in the formula for Q. Its inclusion was not mentioned in the papers that introduced CIECAM02 [28, 36]. It's possible that the factor was introduced to help the Q formula mirror the adapting luminance dependency of the formula for colorfulness, M, so that saturation, which is colorfulness divided by brightness, would remain constant across adapting luminance. However, including the F_L factor actually achieves the exact opposite, making the adapting luminance dependencies of Q and M less similar, because Q contains an additional adapting luminance dependency in the A_W term. So no theoretical or data-based justification for the F_L factor in the formula for Q can be found. Additionally, removing the F_L factor, as proposed here, significantly im-

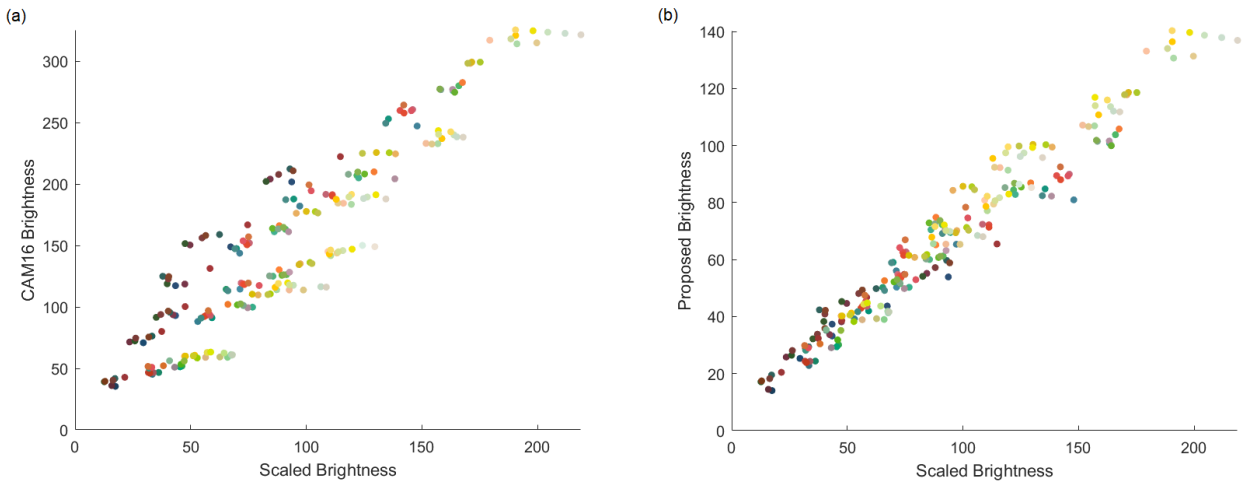


Figure 3.2: LUTCHI brightness scaling data [11] as predicted by (a) Q in CIECAM16 (Equation 3.14) and (b) by the proposed formula for brightness (Equation 3.17). The coefficient of determination (r^2) between the two variables is 0.86 for CIECAM16 and 0.95 for the proposed formula. Colors are approximate. Note that the absolute magnitude of the scales need not match the magnitude of the observed data.

proves the performance of the proposed Q formula (Equation 3.17) on the LUTCHI data (see below). Thus, given the performance benefits and the lack of any data-based or theoretical downside, the F_L factor must be removed.

The inverse c factor, which predicts the overall magnitude of the brightness scale to increase as the surround darkens, is also a candidate for removal, given that the LUTCHI data did not directly test the relationship between the surround conditions and the overall magnitude of the brightness scale. However, we have left that factor in the formula pending further data on the effect of surround conditions on brightness. The overall magnitude of the Q scale has been reduced by half. Originally, Q was scaled to match the arbitrary magnitude of the brightness scale used in the LUTCHI experiment [11]. By rescaling the Q scale, one unit of Q is closer to one unit of reference visual difference as represented by the COMBVD dataset used to derive the CIECAM02-UCS and CAM16-UCS uniform color spaces [4]. Q is now roughly the same magnitude as J when white luminance $L_W = 100$ cd/m^2 since that is the luminance of the reference white for the COMBVD data [16].

The performance of the new formula for brightness (Equation 3.17) was compared to the brightness formula from CIECAM16 (Equation 3.14) using the LUTCHI color appearance data set [11]. These data consist of 36 stimuli whose brightness was scaled at six luminance levels ranging from $L_W = 0.4$ cd/m^2 to $L_W = 842$ cd/m^2 (approximately 11 stops). Hunt relied heavily on these data in his introduction of nonlinearities to the equations for brightness and lightness [9], thus they serve as relevant reference data for the descendants of the Hunt

model, including CIECAM16. The proposed, linear formula for brightness, Equation 3.17, outperforms the CIECAM16 formula for brightness, Equation 3.14, on these data (Figure 3.2). These results lend experimental support to the theoretical justification for the proposed modifications.

3.1.3 Chroma and Colorfulness

In CIECAM02 and CIECAM16, the first step in calculating terms of chromatic intensity is to calculate t, which is similar to saturation, from the opponent chromatic signals a and b:

$$t = \frac{50000 N_c N_{cb} e_t \sqrt{a^2 + b^2}}{13 \left(R'_a + G'_a + \frac{21}{20} B'_a \right)} \tag{3.18}$$

In this formula, N_c is either 1, 0.9, or 0.8 for average, dim, and dark surround conditions, respectively. R'_a, G'_a, and B'_a are the adapted cone signals. The hue-dependent eccentricity factor e_t is included to account for scaling differences between a and b. N_{cb} is a background dependency. These terms will be discussed below. t is then used to calculate chroma, C, colorfulness, M, and saturation, s:

$$C = t^{0.9} \left(1.64 - (0.29)^{\frac{Y_B}{Y_W}} \right) \sqrt{\frac{J}{100}} \tag{3.19}$$

$$M = C \cdot F_L^{0.25} \tag{3.20}$$

$$s = 100 \sqrt{\frac{M}{Q}} \tag{3.21}$$

CIECAM16's formula for chroma has threefold dependence on the background luminance factor, Y_B: via the explicit term in Equation 3.19, via J (see the z term in Equations 3.10 and 3.13), and via the N_{cb} factor in the formula for t (Equation 3.18). Colorfulness, as derived from chroma in CIECAM16, is subject to these three background luminance factor dependencies, plus the additional dependence F_L on the background luminance factor. The threefold dependence of chroma on background was an intentional choice by Hunt in his 1994 model [37]. The desired effect was for a darker background to increase the chroma and colorfulness of medium-dark and dark colors and to decrease the chroma and colorfulness of lighter colors [37]. However, the current background dependencies only achieve this effect when the background luminance factor is greater than 20 (Figure 3.3). Below $Y_B = 20$, predicted the chroma of constant stimuli increases, approaching infinity as Y_B approaches zero. This implausible behavior is due to N_{cb}, which approaches infinity as Y_B approaches

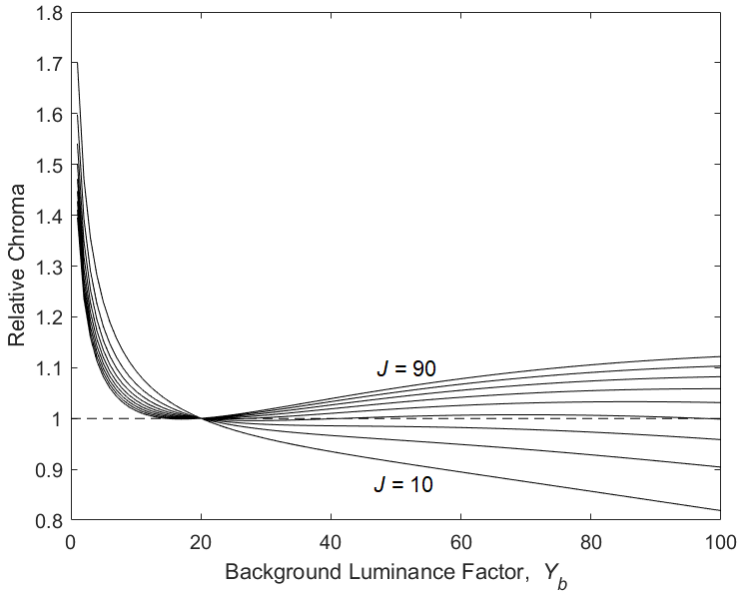


Figure 3.3: Dependence of the chroma of stimuli in CIECAM16 on the background luminance factor relative to their chroma with a mid-gray background, for stimuli with lightness ranging from 10 to 90 in steps of 10. All lines approach infinity as the background luminance factor approaches zero.

zero:

$$N_{cb} = 0.725 \left(\frac{Y_B}{Y_W}\right)^{-0.2} \tag{3.22}$$

Thus, the current background dependencies in the CIECAM16 formula for chroma do not follow Hunt's desired behavior and produce highly implausible values for realistic background levels achieved by modern displays.

Furthermore, careful analysis of the LUTCHI data does not support Hunt's claims about the background dependency of chroma and colorfulness. Key sessions in the LUTCHI experiment involved scaling colorfulness (and other color appearance attributes) against gray and white backgrounds. In line with Hunt's observations, dark colors were scaled with lower colorfulness against the white background than against the gray background. However, the actual colorimetry of the stimuli changed between the background conditions; the stimuli against the white background for which Hunt observed lower colorfulness ratings were, in fact, physically dimmer. Thus, no background dependency is necessary to predict colorfulness ratings across different background luminance factors from the LUTCHI experiment.

The lack of background effects on colorfulness in the LUTCHI experiment seems contradictory to the high performance of CIECAM16—background dependency included—on the LUTCHI colorfulness data. This discrepancy can be explained via the relationship between adapting luminance and background luminance factor. If the adapting luminance is not specified by the user, CIECAM16 recommends using the background luminance factor to

calculate the adapting luminance from the white point luminance. Unfortunately, this causes predicted brightness and colorfulness—color appearance attributes that scale with adapting luminance—to decrease as the background luminance decreases even if the stimulus is held constant. So, the additional background dependencies in the brightness and colorfulness formulas (Figure 3.3) merely offset this unintentional decrease, holding the color appearance attributes constant for constant stimuli.

This combination of deriving adapting luminance from background luminance factor and then undoing the effects of adapting luminance through the three background dependencies in the equations for chroma and colorfulness is confusing for the user, overly complicated, and misrepresents what the color appearance model is doing. A simpler and clearer formula for colorfulness can be derived from the numerator of the formula for t (Equation 3.18):

$$M = 47 N_c e_t \sqrt{a^2 + b^2} \tag{3.23}$$

Chroma is derived by colorfulness by dividing by the achromatic white signal to make chroma invariant to scene luminance:

$$C = 35 \frac{M}{A_w} \tag{3.24}$$

Saturation, s, can be calculated from colorfulness and brightness using a linearized version of the CIECAM16 formula for s:

$$s = 100 \frac{M}{A} \tag{3.25}$$

The achromatic signal, A, is used in the formula for saturation in place of brightness, Q, to help achieve partial chromaticity-constancy of the saturation metric across changes in luminance, which is a desirable property for such metrics [36].

In addition to removing the myriad background dependencies, these formulas make the theoretical improvement of linearizing the formulas. The many nonlinearities in the original formulas (Equations 3.19 to 3.21) appear to have been introduced to improve the performance of the model on the LUTCHI data without theoretical justification. As will be seen below, linearizing the formulas improves the linearity of their chroma and colorfulness predictions. Removing the background dependencies (specifically, the $\sqrt{\frac{J}{100}}$ term in Equation 3.19) requires this restructuring of the formulas for chroma and colorfulness. An additional reason for removing the explicit J factor from the formulas for M and C is that the factor merely and poorly canceled out the denominator of the formula for t (Equation 3.18), leading to incorrect predictions of chroma and colorfulness for stimuli with large values of blue cone signal, B_a'.

For a given reflective object, chroma—as predicted by both CIECAM16 and the proposed

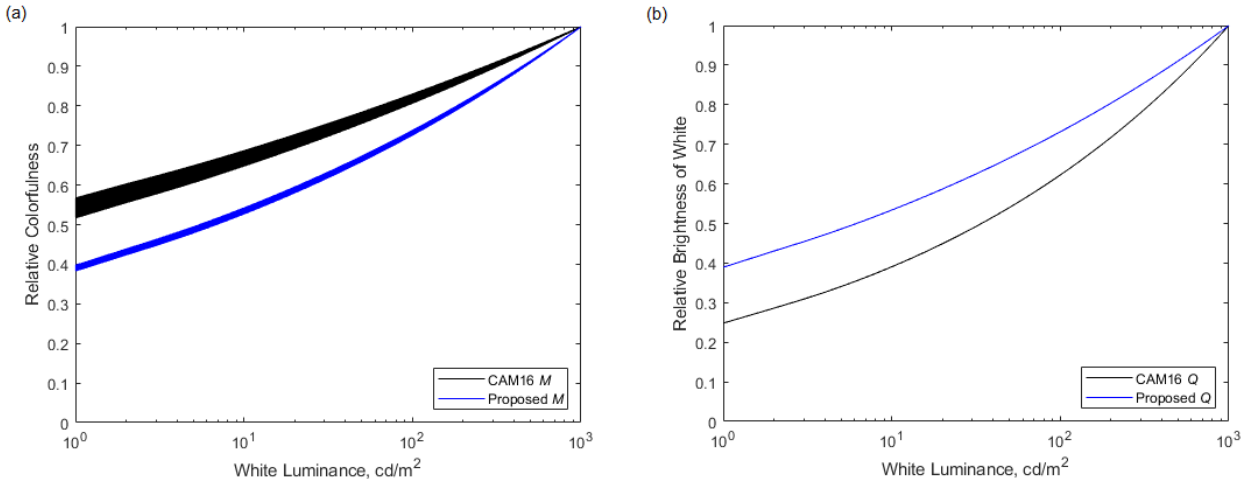


Figure 3.4: Effect of the overall luminance level on (a) colorfulness, M, and (b) brightness, Q, for CIECAM16 and the proposed formulas. Colors from the Munsell color order system were used to measure the relationship between colorfulness and luminance. The proposed formulas for Q and M follow similar trends with regard to luminance, as they both scale proportional to A_W. All values were normalized relative to a white luminance of 1000 cd/m^2.

model—is constant as the scene luminance changes. In both models, the colorfulness of a reflective object increases with increasing scene luminance (Figure 3.4). In CIECAM16, this luminance dependency is proportional to $F_L^{0.25}$. In the proposed model, the colorfulness of a given object has the same relationship with scene luminance as the achromatic white signal, A_W, given that a, b, and A_W are all proportional to the adapted cone signals $R_a^{'}$, $G_a^{'}$, and $B_a^{'}$.

We have modified the overall magnitude of the M and C scales. In CIECAM02 and CIECAM16, M was scaled to match the arbitrary magnitude of the colorfulness scale from the LUTCHI experiments [11]. The proposed M formula, Equation 3.23, is scaled by 0.75 compared to the formula from CIECAM16. This was done to better match the scale of unit visual differences from the COMBVD color difference dataset [38]. Given that the proposed Q scale was scaled by 0.5 relative to its CIECAM16 formula, the M dimension is now 50% larger relative to the Q dimension in the proposed formulas. This was done to minimize STRESS on the COMBVD data [39].

The formula for C was scaled by 0.6 relative to its magnitude in CIECAM16. This scaling provides a more accurate magnitude for chroma relative to the J dimension in the proposed formulas, improving the uniformity of the scales and minimizing STRESS as measured by the COMBVD color difference dataset [38, 39].

We propose to change the eccentricity function, e_t. The opponent chromatic signals, a and b (Equations 3.2 and 3.3), are not guaranteed to be properly scaled in magnitude

relative to each other. Thus, CIECAM02 and CIECAM16 use an eccentricity factor, e_t, to account for differences in the scaling of a and b when calculating chroma and colorfulness. The formula used in CIECAM16 can be traced back to values derived by Hunt for his 1982 color appearance model [29]. Hunt derived his eccentricity factors by drawing loci of constant saturation from the NCS color order system on a $u'v'$ chromaticity diagram. Specifically, he calculated the relative radii of these loci at the four unique hues from the NCS system. While $u'v'$ coordinates have no concrete relationship with a and b, Hunt reasoned that the limits of the relative radii of the loci as the saturation approached zero would be invariant of the color coordinates used. These assumptions and calculations led to eccentricity values of 1.45 for NCS unique blue, 0.65 for unique red, 0.5 for unique yellow, with the eccentricity of unique green set to unity [29].

In the 1985 revision of his color appearance model, Hunt introduced cross-talk between his R, G, and B cone signal values [30]. This cross-talk calculation included an additional square root applied to the cone responses. Hunt also took the square root of the eccentricity values from his 1982 paper since he was now working in square root response space, leading to eccentricity values of 1.2 for blue, 0.8 for red, and 0.7 for yellow, with the eccentricity of green set to unity. These values were used to derive the eccentricity function, e_t, found in CIECAM16 (Figure 3.5) [4]:

$$e_t = \frac{1}{4}\left[\cos\left(\frac{h\pi}{180} + 2\right) + 3.8\right] \tag{3.26}$$

The hue angle, h, is defined as:

$$h = \tan^{-1}\left(\frac{b}{a}\right) \tag{3.27}$$

The formula for eccentricity gives values of approximately 1.198 for blue, 0.774 for red, 0.723 for yellow, and 0.988 for green, closely matching the above values from Hunt's 1985 paper. The hue angles used for the NCS unique hues were also transcribed from Hunt's 1985 model as opposed to measuring the hue angle of the NCS unique hues in the CIECAM16 a-b dimensions.

This eccentricity function is problematic for a number of reasons. First of all, there are several potential flaws in the method used by Hunt to derive the initial eccentricity values in 1982. Chromaticness in the NCS system is relative to the maximum chromatic intensity of each individual hue [40], thus NCS chromaticness and saturation are not meant to be compared in absolute terms across hues as Hunt did by drawing loci of constant saturation. Additionally, there is no self-apparent justification for his assumption that the limit of the loci of constant saturation as saturation approaches zero is invariant across different color

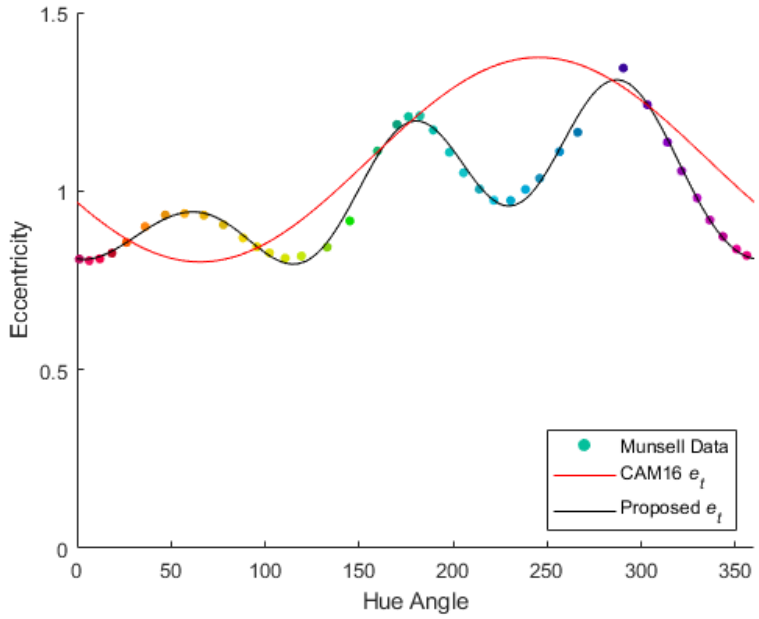


Figure 3.5: Average eccentricity of Munsell colors as a function of CIECAM16 hue angle in comparison to the CIECAM16 e_t function (Equation 3.26) and the proposed formula, which was fit to the Munsell data (Equation 3.28). Colors are approximate.

spaces. Secondly, there is no justification for the use of the same numeric values for Hunt's 1985 model and CIECAM16, given that they have different RGB cone spaces and different tone compression functions. While CIECAM16 has remained true to the values derived by Hunt, it has lost the connection to Hunt's original intent in proposing these values. Finally, CIECAM16 assumes that the proper eccentricity values follow a sinusoidal shape between the target unique hue values. However, no evidence is provided to support such an assumption.

By returning to Hunt's original intent—the scaling of stimuli of each hue by their relative chromatic strength—we can derive an eccentricity function that resolves the problems described above. Unlike the NCS system, which normalizes each hue by its chromatic strength, the Munsell system has a measure of chromatic intensity, chroma, whose magnitude can be compared across hues [40]. To determine the proper eccentricity function, the Munsell chroma of each Munsell color is divided by $\sqrt{a^2+b^2}$, which is proportional to the chroma of the color following our proposed formula (Equations 3.23 and 3.24). The mean dividend for each Munsell hue is shown in Figure 3.5. A new formula for e_t was fit to these data:

$$\begin{aligned} e_t = &-0.0582\cos(h) - 0.0258\cos(2h) - 0.1347\cos(3h) + 0.0289\cos(4h) \\ &-0.1475\sin(h) - 0.0308\sin(2h) + 0.0385\sin(3h) + 0.0096\sin(4h) + 1 \end{aligned} \tag{3.28}$$

The formula was normalized to have an average value of one. While this proposed formula is more complex than the current formula (Equation 3.26), it merely reflects the trend of the Munsell data (Figure 3.5), which appears plausible. We believe that it is better to directly represent the Munsell data rather than choose an ambiguous middle-ground between complexity and basis in data.

While the proposed formulas offer clear theoretical advantages to the current CIECAM16 formulas for chroma and colorfulness, it is important to verify that these proposed formulas also perform well on visual data. Data from the Munsell color order system [41] and the LUTCHI color appearance scaling experiments were used to compare the proposed and current models. These LUTCHI data contain two subsets. The first set of data consists of the scaled colorfulness of 99 stimuli at two luminance levels (252 cd/m^2 and 42 cd/m^2) and three relative background luminance levels (6.2%, 21.5%, and 100%) [10]. The second set of data consists of the scaled colorfulness of 36 stimuli at six luminance levels ranging from $L_w = 0.4$ cd/m^2 to $L_w = 842$ cd/m^2 (approximately 11 stops) against a mid-gray background [11].

Different methods were used to calculate the adapting luminance for the CIECAM16 formulas versus the proposed formulas. Since, as discussed above, the background dependencies in the current CIECAM16 formulas for chroma and colorfulness compensate for background dependency of adapting luminance, the adapting luminance was allowed to vary with background luminance when predicting the LUTCHI data with the current CIECAM16 colorfulness formula. Even though this method of calculating the adapting luminance is problematic (as discussed above) and can easily lead to errors for unaware practitioners, this method was chosen to represent the best possible performance for CIECAM16 on the LUTCHI colorfulness data. On the other hand, since the proposed formulas remove this convoluted set of counteracting background dependencies, the adapting luminance could be held at 20% of the white point luminance for all LUTCHI data calculations.

The models' performance on the Munsell data is shown in Figure 3.6 and their performance on the LUTCHI data is shown in Figure 3.7. The proposed chroma formula shows a clear improvement on the Munsell data compared to the current formula. For the current formula, the plot appears to curve downwards as chroma increases. This nonlinearity is possibly due to the nonlinear relationship between t and C in CIECAM16 (Equations 3.19), where greater values of C are compressed. Figure 3.6 shows the clear advantage in linearizing these formulas: there is no more downward curve at high chromas with the proposed formula. However, when analyzing Munsell chroma predictions within a single hue and chroma, the proposed chroma of the proposed formula decreases with decreasing value. In summary, the proposed formulas appear to be superior at predicting Munsell chroma as chroma and hue

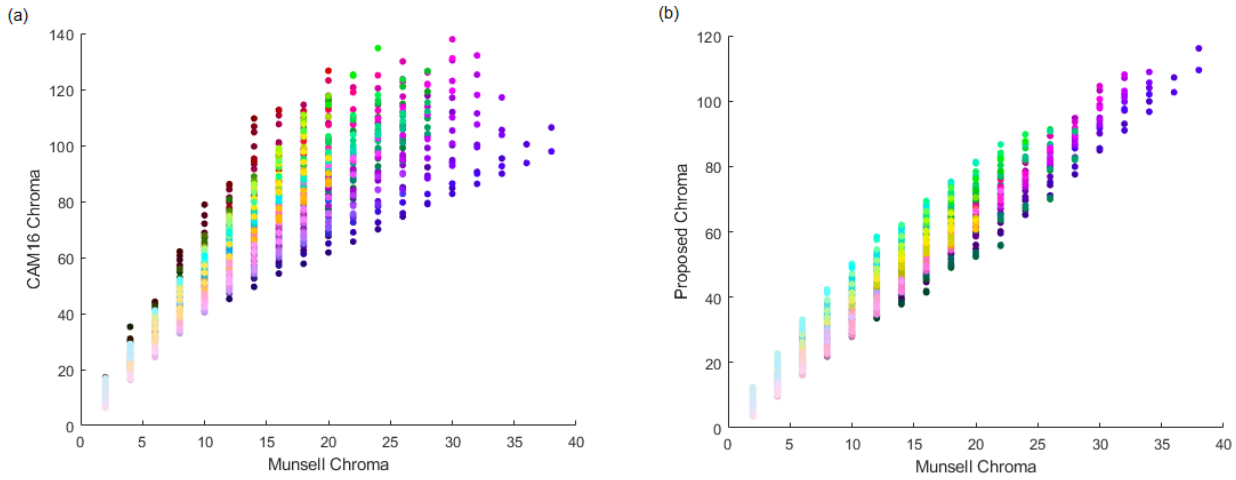


Figure 3.6: The chroma of colors from the Munsell color order system [41] as predicted by chroma in (a) CIECAM16 (Equation 3.19) and (b) the proposed formula (Equation 3.24). The coefficients of determination (r^2) for the data are 0.87 for CIECAM16 and 0.96 for the proposed formula. The proposed chroma attribute also demonstrates improved linearity. Note that colors are approximate and that the scales need not be equal in magnitude.

change, but not as consistent at predicting Munsell chroma as value changes. We believe that this is a worthwhile tradeoff.

The proposed formula for colorfulness (Equation 3.23) underperforms the current CIECAM16 formula (Equation 3.20) on the LUTCHI data (Figure 3.7). This performance advantage for the CIECAM16 formula is due to the difference in luminance dependency of the current and proposed formulas. The current colorfulness formula scales proportional to $F_L^{0.25}$, whereas the proposed colorfulness formula has the same relationship with scene luminance as A_W (Figure 3.4). However, there is an important theoretical argument for the proposed formula's proportionality to A_W: this matches the luminance-dependent behavior of the proposed brightness, Q, formula (Equation 3.17). Thus, as scene luminance increases, the proposed colorfulness and brightness scales remain in proportion to each other. This proportionality is necessary for saturation to remain invariant to scene luminance level. In the current CIECAM16 formulas, brightness increases more quickly than colorfulness with increasing luminance, leading to the poor performance of CIECAM16 on the LUTCHI brightness data (Figure 3.2). Given these theoretical considerations and the importance of the A_W dependency for the proposed Q formula, the worse performance on the LUTCHI data by the proposed colorfulness formulas is permissible.

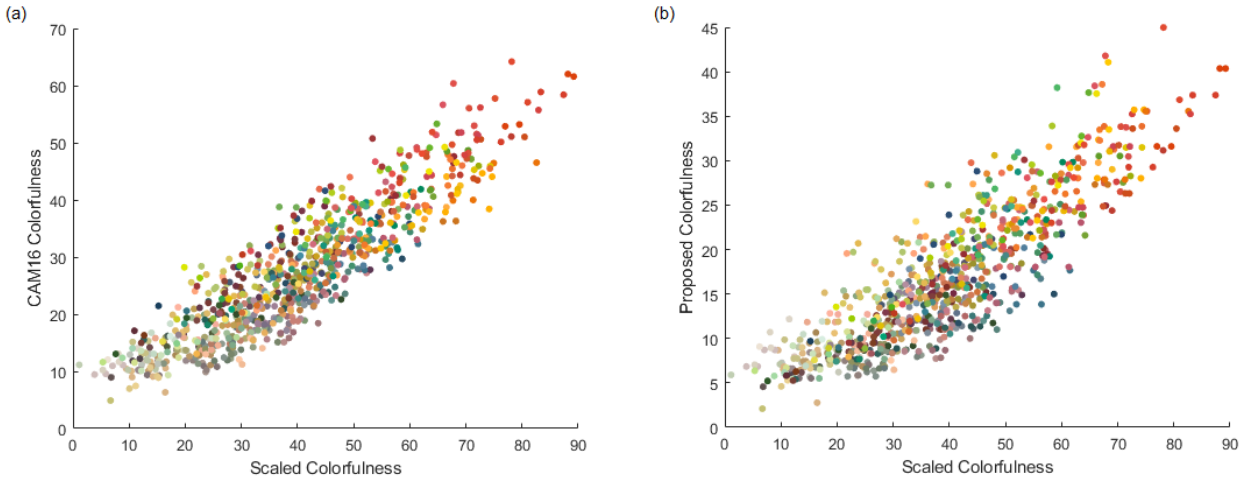


Figure 3.7: Colorfulness data from the LUTCHI scaling experiments [10, 11] as predicted by (a) CIECAM16 (Equation 3.20) and (b) the proposed formula (Equation 3.23). The coefficients of determination (r^2) for the data are 0.81 for CIECAM16 and 0.71 for the proposed formula. Note that colors are approximate and that the scales need not be equal in magnitude.

3.1.4 Discussion

We have introduced important revisions to the CIECAM02 and CIECAM16 formulas for brightness, colorfulness, and chroma. Our goal has been primarily conservative in nature—not to extend CIECAM16 for new applications or datasets, but rather to improve its internal consistency while remaining grounded in the LUTCHI dataset and the principles used to derive the original equations. When needed, data from the Munsell color order system has supplemented the LUTCHI data, allowing us to improve the linearity of the color appearance model. Additionally, simplifications to certain formulas have brought the color appearance model into line with the theoretical definitions of color appearance terminology while also making explicit the effects accounted for by the model.

Analyzing the history of the equation for brightness, Q, in these color appearance models, we found that the nonlinear relationship between lightness, J, and brightness is an artifact of how the Hunt model was transcribed to CIECAM97s. Linearizing the nonlinearity (Equation 3.17) removes a perceptual paradox (Figure 3.1) and improves the performance of the Q equation on brightness scaling data from the LUTCHI experiment (Figure 3.2). Removing redundant dependencies from the CIECAM16 equation for Q (Equation 3.14) simplifies the brightness formula and improves performance. Thus, the proposed changes to the brightness formula are justified and necessitated by both theory and performance and are the most urgent of all changes proposed in this section.

Resemblance between the CIECAM16 formulas for Q and C, chroma, (Equation 3.19)—

specifically, the $\sqrt{\frac{J}{100}}$ term that appears in both and is clearly incorrect in the Q formula—prompted a reevaluation of CIECAM16 formulas for chroma, colorfulness, M, and saturation, s. Subsequent improvements made to the chroma and colorfulness formulas fall into three categories: background dependencies, eccentricity, and linearity.

The current CIECAM16 formulas contain myriad background dependencies that counteract each other. This convoluted formulation hid the fact that the actual background dependency did not follow Hunt's qualitative description of how chroma and colorfulness depend on background luminance factor. Furthermore, close analysis of the LUTCHI data revealed no statistically significant effect of background luminance factor on scaled colorfulness or chroma. Thus, the background dependencies have been removed from the formulas for colorfulness and chroma. Additionally, it is now recommended that the adapting luminance be specified directly by the user, as opposed to being derived from the background luminance. Together, these changes mimic the background-invariance found in the current formulas. Now, this invariance is explicit, as opposed to the current formulas, which claim to be dependent on background luminance but are actually invariant in practice. This new, more honest formulation allows for future addition of background dependencies, if desired by the user. For instance, a background dependency of colorfulness has been reported by Kim, Weyrich, and Kautz [42]. They follow a similar approach to modeling as proposed here; they do not include an explicit background dependency in their formulas for colorfulness and chroma. Instead, they modify the adapting luminance input term to reflect the changing background level while holding the stimulus luminance level constant. The Kim et al. approach may be worth exploring in a future iteration of CIECAM16. In the current model, such an accommodation is not possible because absolute luminance level of the stimulus is derived from the adapting luminance as opposed to being specified independently.

Eccentricity is a key function in CIECAM02 and CIECAM16 that scales colorfulness, chroma, and saturation to be perceptually uniform across hues. The current eccentricity function (Equation 3.26) was fit to four values from an early version of the Hunt model. These values are no longer relevant to the current model given the fundamental differences between Hunt's early model and CIECAM16. Furthermore, their original derivation relied on assumptions that are unsupported and potentially incorrect. We have followed the core principles laid out by Hunt along a more rigorous path to deriving an eccentricity function directly from an analysis of the Munsell color order system using CIECAM16 color coordinates. The directness of the derivation promises to provide a much more reliable measure of eccentricity (Equation 3.28).

The proposed formulas for colorfulness, chroma, and saturation have been linearized in comparison to their current CIECAM16 counterparts. This linearization is more theoreti-

cally grounded in the definitions of these attributes and leads to improved performance on data from the Munsell color order system. However, the linearization of the colorfulness equation contains a significant tradeoff. In the current CIECAM16 colorfulness formula, colorfulness increases in proportion to the adapting-luminance-dependent $F_L^{0.25}$. Colorfulness in the proposed formulas increases with increasing adapting luminance at the same rate as A_W, the achromatic white signal (Figure 3.4). The matching luminance dependencies of A_W and M in the proposed formula ensures that colorfulness remains proportional to brightness as adapting luminance increases, and the A_W dependency of brightness is in turn necessary to correctly predict the LUTCHI brightness data. However, the A_W dependency does hurt the performance of the proposed colorfulness formula on the LUTCHI colorfulness data compared to the current CIECAM16 formula. More work should be done to evaluate the proper relationship between colorfulness and adapting luminance.

3.2 Why Achromatic Response is Not a Good Measure of Brightness

3.2.1 Background

One common metric for the sensitivity of the human visual system in photopic conditions is luminance, which is defined by a spectral sensitivity function symbolized $V(\lambda)$. $V(\lambda)$ was standardized in 1924 by the Commission Internationale de l'Éclairage (CIE) using results from an experimental method called heterochromatic flicker photometry [43]. In flicker photometry, the visual stimulus alternates quickly between a test patch and a reference patch (on the order of 10-30 Hz [44–46], but dependent upon the luminance of the stimuli [47]). The intensity of the reference patch is adjusted until the two patches fuse and the perception of flicker disappears. If the frequency of the flicker is properly set, then there will be a small range of intensities for which flicker fusion occurs [45]. In the case of $V(\lambda)$, flicker photometry was used with spectral test stimuli to determine the visual sensitivity to the entire visible spectrum. (For more detailed discussion of the methods and data used to derive $V(\lambda)$, see [44].) Luminance has been widely adopted, partially because its definition is inherently additive: the luminance of a stimulus made from different sources is equal to the sum of the luminances of its constituent sources.

Brightness is defined as the degree to which a visual stimulus appears to emit or reflect more or less light [25]. Lightness is defined as the brightness of a stimulus relative to the brightness of an equally-illuminated white object in the scene [25]. A common method for

assessing brightness is direct brightness matching [48–52]. In this method, two visual stimuli are shown next to each other with a small gap in between. Observers adjust the intensity of one stimulus until the two stimuli appear equally bright or light. Unlike luminance, brightness is not additive [53]: if chromatic lights (e.g., red, green, and blue lights) are combined to form an achromatic light, the sum of the brightnesses of the constituent lights will be greater than the brightness of their light when combined. In other words, the results of heterochromatic flicker photometry and direct brightness matching do not agree with each other, and the brightness of chromatic stimuli is underestimated by their luminance. This phenomenon is known as the Helmholtz-Kohlrausch effect and has been widely studied [49–51, 54–57].

There are two fundamental explanations for the discrepancy between heterochromatic flicker photometry and direct brightness matching. Firstly, the neural pathways that carry chromatic information—the parvocellular pathways—are less sensitive, and perhaps completely insensitive, to the temporal frequencies at which flicker photometry is performed [46, 58]. Undoubtedly, though, they will be sensitive to the static stimuli in direct brightness matching. This suppression of chromatic information only during flicker photometry but not during direct matching might partially explain the discrepancy between the luminance and brightness of chromatic stimuli [59]. Furthermore, the two experimental methods contain different tasks: in one, to minimize flicker, and in the other, to match brightness. There is no rule of perception that states that the flicker minimization occurs when the stimuli are equally bright. In fact, task-dependency has already been reported in research that compared direct brightness matching to an alternate method called minimally distinct border that also uses static stimuli [51]. In minimally distinct border experiments, a test patch and reference patch are placed directly adjacent, and the reference patch adjusted until the border between the two is minimally visible or distinct. Such experiments appear to produce results that are more similar to luminance-like matches than the results produced by direct brightness matching, showing that task method can have a substantial effect [51, 60].

In this work, we separate the two factors described above which distinguish flicker photometry from direct brightness matching. By slowing down flicker photometry to allow both the magnocellular and parvocellular pathways to respond to the experimental stimuli [61], we can directly measure the effect of the difference between these two experimental tasks. Such work has potential implications for color spaces, such as CIELAB [62], and color appearance models, such as CIECAM16 [1, 4], which seek to predict perceptual attributes including brightness and lightness from information about the stimulus and (in the case of color appearance models) the environment in which the stimulus is viewed. Understanding the difference between methods of measuring brightness is especially relevant for those with

an interest in incorporating the Helmholtz-Kohlrausch effect into such models or color spaces [32, 56, 63, 64]. Conversely, color appearance models provide a baseline within the context of experimental design to investigate how these experimental tasks operate in the perceptual domain, allowing us draw more meaningful conclusions from experimental results than would be possible using CIE XYZ tristimulus values, which have no inherent perceptual meaning.

3.2.2 Methods

Psychophysical experiments were run using two methods of stimulus presentation: direct matching and temporal oscillation/flicker. In direct brightness matching, one chromatic patch and one achromatic patch were shown side by side against a random noise background (Figure 3.8). Each patch filled approximately one degree of visual angle and the two patches were separated by approximately one half of one degree of visual angle. The nine observers were instructed to use a keyboard to adjust the luminance of the achromatic patch until the two patches matched in brightness. Each judgement was repeated, resulting in eighteen total observations. The random noise background had an average CIECAM16 lightness of 50 relative to a 950 cd/m^2 D65 white point. The noise pattern was used to reduce the effect of simultaneous contrast from uniform backgrounds and to prevent bias by removing a fixed reference point for observers (such as would be provided by a uniform background) without changing their overall state of adaptation.

In the flicker/temporal gradient method, a single, one-degree patch was shown to observers against the random noise background (Figure 3.9). The patch oscillated continuously between the test chromatic stimulus and the achromatic stimulus (Figure 3.10). Observers were instructed to adjust the luminance of the achromatic stimulus to minimize their perception of flicker. Five oscillation frequencies were tested: 0.5 Hz, 1.39 Hz, 3.87 Hz, 15 Hz, and 30 Hz. Intermediate stimuli in the oscillations were evenly spaced in CIECAM16 color space between the chromatic and achromatic stimuli. The number of intermediate stimuli was determined by the oscillation frequency and the refresh rate of the monitor (60 Hz). For instance, the 1.39 Hz oscillation had 21 intermediate stimuli, whereas the 15 Hz oscillation had only one, and the 30 Hz oscillation had none.

In a follow-up experiment to test the effect of the intermediate stimuli, the intermediate stimuli were removed. In this case, the patch simply alternated between the achromatic and chromatic endpoints at each test frequency. The results of this follow-up are included in the Results section below.

In an additional follow-up experiment, the intermediate stimuli were adjusted from being evenly spaced between the achromatic and chromatic endpoints to being sinusoidally spaced

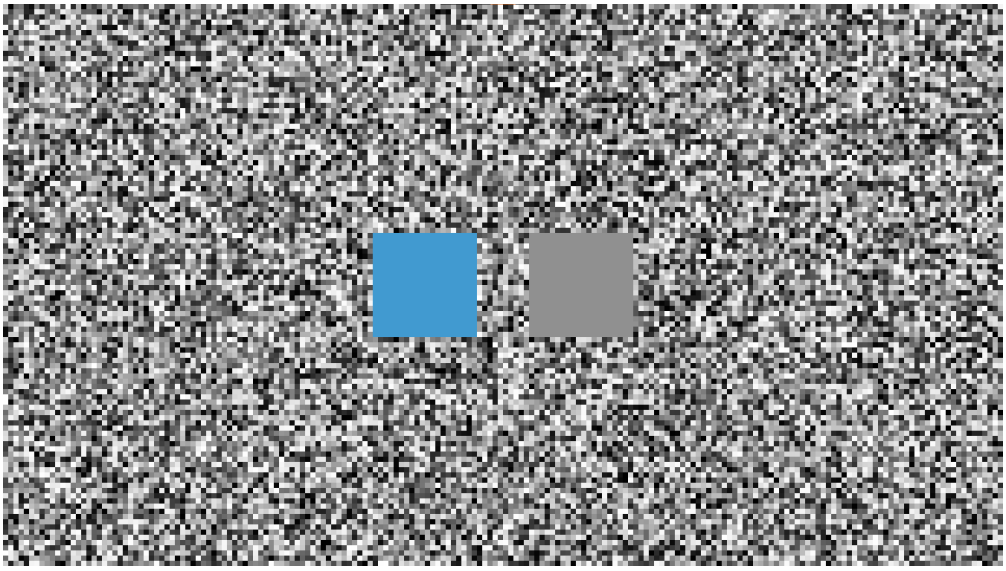


Figure 3.8: Screenshot of example stimulus from the direct matching experiment. Observers adjusted the luminance of the achromatic patch (right) until it matched the brightness of the chromatic patch (left). The left/right orientation of the achromatic and chromatic patches was randomized for each trial. e Each patch occupied approximately 2° of visual angle with a 1° gap between them. Colors are approximate.

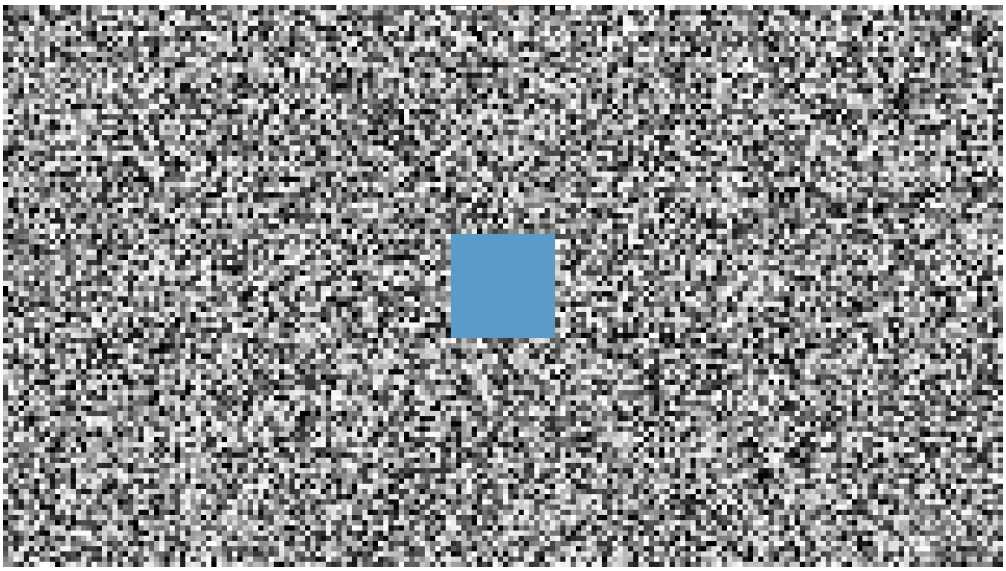


Figure 3.9: Screenshot of example stimulus from the flicker/temporal oscillation experiment. The patch occupied approximately 2° of visual angle. Color is approximate.

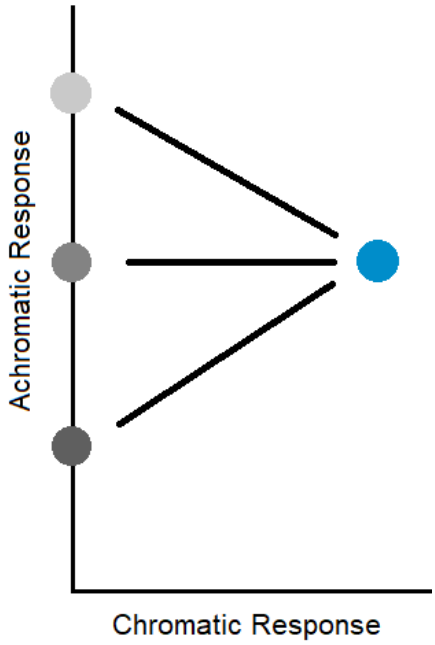


Figure 3.10: Schematic diagram of stimuli used for temporal oscillation experiment. The single patch stimulus (Figure 3.9) oscillated between the fixed chromatic endpoint and the adjustable achromatic endpoint. CIECAM16 lightness and chroma (with the revisions proposed in Section 3.1) were used as the achromatic and chromatic response dimensions, respectively, to calculate the intermediate stimuli between the endpoints. Observers adjusted the luminance of the achromatic endpoint until a perceptual minimum was reached. This point indicates the achromatic endpoint which has the same achromatic response as the chromatic endpoint in the observer's internal perceptual color space (see Discussion).

between the endpoints. This modification had no statistically significant effect on the results and thus is excluded from further analysis.

Experimental stimuli were displayed on an Asus ProArt PA32UCS monitor, controlled using Psychtoolbox-3 [65] and MATLAB on a Windows computer with an Nvidia Quadro P400 video card. Three chromatic test patches with CIECAM16 hue angles of 12°, 110°, and 242° were tested, roughly corresponding to red, yellow, and blue, respectively. The red, yellow, and blue patches had CIECAM16 lightness of 47.1, 50.7, and 50.7, and CIECAM16 chroma of 20.5, 26.3, and 23.2, respectively, calculated using our proposed corrections to CIECAM16 (Section 3.1). A dark surround and a degree of adaptation of one were used for all CIECAM16 calculations. The slight variation in lightness and chroma values between hues were due to differences between the Rec. 2100 color space [66], which was used to generate the code values for the experimental stimuli, and the performance of the Asus ProArt display in matching that standard. These variations do not impact the conclusions of this paper.

3.2.3 Results

The results of the experiment are quantified by the lightness of the achromatic patch matched to the test chromatic patch by the observers in each viewing situation. The mean lightnesses of the achromatic patches are shown in Figure 3.11 along with estimated 95% confidence intervals. The statistical significance of the differences between mean values was calculated using Welch's two-sample t-test at an α level of 0.05 with equal variances not assumed.

First, the results from the 15 Hz and 30 Hz flicker observations were compared to determine which frequency best represented traditional heterochromatic flicker photometry. The mean values from the two frequencies were not significantly different for any individual hue (Table 3.1), but the variance of the 30 Hz observations was substantially greater. Observers reported that there was a wide range of lightnesses for which their perception of flicker disappeared at this frequency of oscillation, explaining the high variance of their responses. Thus it was decided that the 15 Hz oscillation should serve as the representative sample of the traditional heterochromatic flicker method.

Figure 4 shows that at lower frequency oscillations, the observers chose a lighter achromatic patch to minimize their perception of flicker with the same chromatic patches. Values that were significantly different from the mean value at 15 Hz are circled in Figure 3.11 (p values in Table 3.1). The difference between the 15 Hz oscillation and slower frequencies could be due to the increased sensitivity of the parvocellular neural pathway—carrying color information—at lower frequencies.

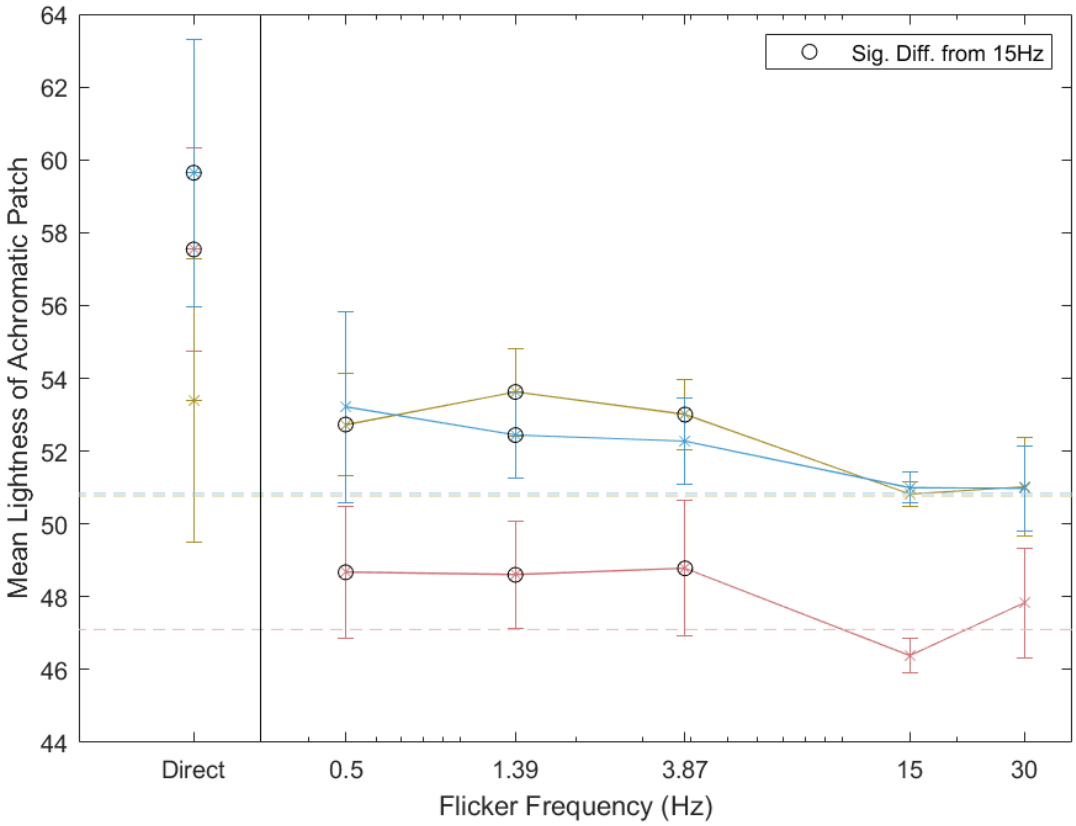


Figure 3.11: Results of the direct matching experiment (left) and the oscillation method with variable frequency (right) for the three tested hues (red, yellow, and blue). The y-axis represents the mean lightness of the achromatic patch that was adjusted by observers in each experimental condition. Dashed lines indicate the lightnesses of the fixed chromatic patches. Values that are significantly different ($p < 0.05$ on a two-sample t-test with unequal variances) from the luminance-like match at 15 Hz are circled.

Color	Flicker Frequency (Hz)				
	Direct	0.5	1.39	3.87	30
Red	$\underline{4.2\times10^{-7}}$	$\underline{0.027}$	$\underline{0.011}$	$\underline{0.024}$	0.085
Yellow	0.22	$\underline{0.018}$	$\underline{2.8\times10^{-4}}$	$\underline{4.1\times10^{-4}}$	0.79
Blue	$\underline{2.6\times10^{-4}}$	0.12	$\underline{0.036}$	0.063	0.96

Table 3.1: p values for Welch's two-sample t-test with equal variance not assumed for the data represented in Figure 3.11 compared to mean achromatic lightness for the 15 Hz oscillation. Statistically significant values are underlined.

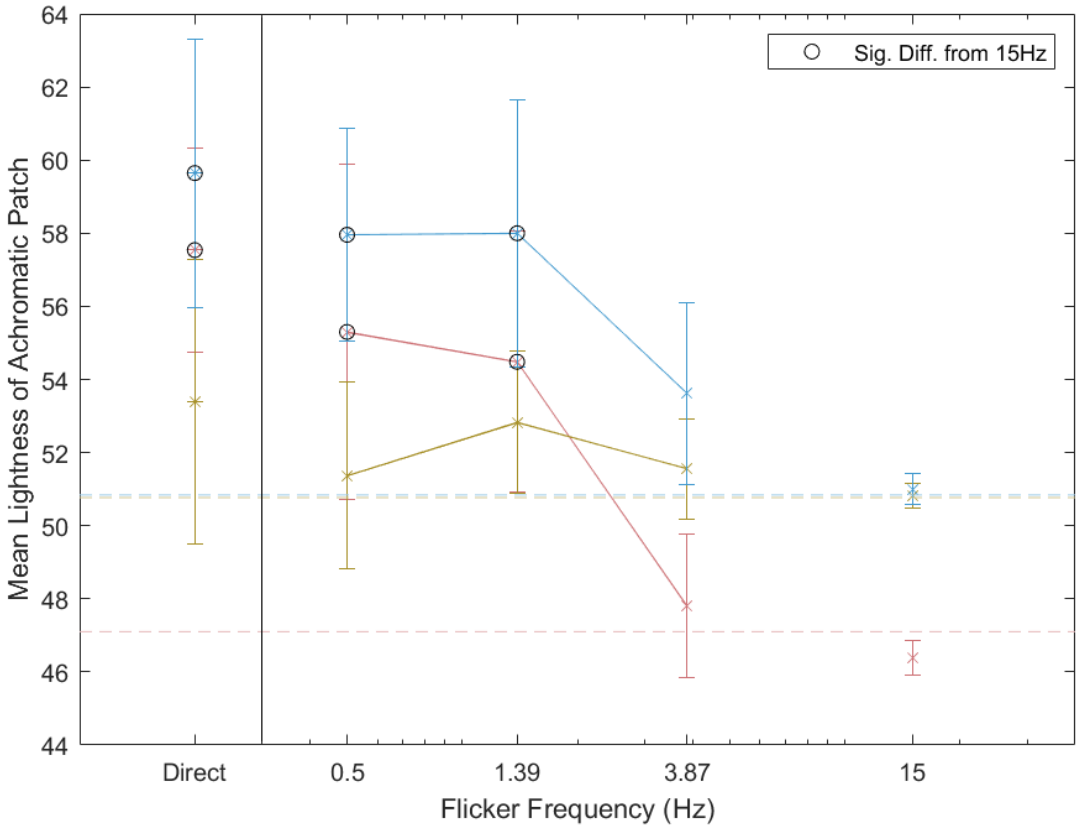


Figure 3.12: Results of the square-wave oscillation experiment (flicker frequencies 0.5 Hz, 1.36 Hz, and 3.87 Hz) compared to the direct matching and 15 Hz results from Figure 3.11. The y-axis represents the mean lightness of the achromatic patch that was adjusted by observers in each experimental condition. Dashed lines indicate the lightnesses of the fixed chromatic patches. Values that are significantly different ($p < 0.05$ on two-sample t-test with unequal variances) from the luminance-like match at 15 Hz are circled

When the pattern of the slow frequency oscillations was changed from an even gradient to simply alternating between the two endpoints, the matched lightness of the achromatic patch increases even further (Figure 3.12) for the red and blue patches. The difference between the two methods was statistically significant using test described above for the red and blue patches, but not for the yellow patches. An explanation for this result is given in the Discussion section.

The results of the direct matching experiment are also shown in Figure 3.11. For the red and blue test patches, the mean lightnesses of the matched achromatic patches were significantly greater than their value from the smooth oscillation method (Table 3.2).

Color	Flicker Frequency (Hz)				
	Direct	0.5	1.39	3.87	30
Red	$\underline{1.4\times10^{-5}}$	$\underline{8.4\times10^{-6}}$	$\underline{1.8\times10^{-5}}$	$\underline{4.2\times10^{-7}}$	$\underline{2.5\times10^{-6}}$
Yellow	0.75	0.91	0.85	0.22	0.27
Blue	$\underline{2.6\times10^{-4}}$	$\underline{2.4\times10^{-4}}$	$\underline{0.0012}$	$\underline{0.0015}$	$\underline{0.0089}$

Table 3.2: p values for Welch's two-sample t-test with equal variance not assumed for mean results of oscillation method compared to mean results for the direct matching achromatic. Statistically significant values are underlined.

3.2.4 Discussion

The experimental results clearly show three distinct regimes of observer behavior corresponding with the direct matching, slow oscillation, and fast oscillation methods of stimulus presentation. The fast oscillation presentation leads to a luminance-like match solely based on the achromatic information present in the stimuli [58]. In direct brightness matching, chromatic information is incorporated into the stimulus judgement, leading chromatic stimuli to be judged as brighter in direct matching than in luminance-like matching [59]. This is known as the Helmholtz-Kohlrausch (H-K) effect and is represented in the results of our experiment. That we observed a stronger H-K effect for the red and blue stimuli than for the yellow stimulus is in agreement with other studies [56, 67]. Given that observers in the direct matching stimulus presentation were asked to match the patches in brightness, we can conclude that patches directly matched have equal (perceived) brightness. Therefore, the patches that were matched in the fast oscillation method do not have equal brightness.

The results from the novel slow oscillation presentation require more detailed explanation. Given the increased chromatic contrast sensitivity at the low frequencies tested in this experiment, we expect the chromatic (parvocellular) channel of our visual system to respond to the oscillation, as in direct brightness matching [58, 61]. However, the observers' task is different for the slow oscillation matching than the direct brightness matching: minimizing their perception of flicker versus matching brightness.

Unlike with fast flicker, where the observers' perception of flicker could disappear when the chromatic and achromatic endpoints were matched, the perception of oscillation never disappeared as the patch oscillated between its achromatic and chromatic endpoints. Nevertheless, observers could adjust the achromatic patch to find a clear perceptual minimum in the slow oscillation. At that point, observers experienced a minimum in the perceived speed of the oscillation. The perception of a minimum in speed can be understood by conceptualizing the observers' perception of the experimental stimuli as existing in an internal color

space (Figure 3.10). As the achromatic endpoint moves up and down the achromatic axis as it is adjusted by the observer, it moves farther from and closer to the chromatic endpoint. Thus, the oscillation covers more or less distance in our internal color space over the same amount of time and appears to speed up or slow down. A minimum in perceived speed occurs when the distance between the achromatic and chromatic endpoints is at a minimum in the observers' internal color space. This minimum distance occurs when the achromatic and chromatic endpoints have an equal position along the dimension of achromatic response. Thus, unlike other psychophysical methods of luminance or brightness measurement, this slow oscillation is a direct measure of this dimension of our internal color space.

This interpretation of the results of the slow oscillation matching is confirmed by the effect of switching from a colorimetrically smooth oscillation between the achromatic and chromatic endpoints (Figure 3.11) to a temporal square-wave pattern that simply alternated between the two endpoints without any intermediate stimuli (Figure 3.12). For the smooth oscillation, the mean lightnesses of the achromatic patches matched to the red and blue chromatic patches were significantly less than the lightnesses of the achromatic patches matched to the same chromatic patches via direct brightness matching (Figure 3.11). However, when the intermediate stimuli were removed, the lightnesses of the matched achromatic patches increased as the frequency decreased, trending towards the directly matched lightnesses. This equivalence between directly matching two patches and viewing them in alternation on a single location is coherent with the idea that the observer must move their gaze between the two patches when directly comparing, so both situations generate similar temporal patterns in the visual system. More importantly, that the square wave oscillation results match the direct matching results confirms that the difference between the smooth oscillation results and the direct matching results is not due to temporal effects in the visual system's response to the low-frequency oscillation which are not accounted for in our above explanation. Put simply, this result supports our assumption that the same neural pathways in our visual system are active during the direct matching and slow oscillation stimulus presentations and that the difference in the results is primarily due to the difference in task.

The statistically significant difference between the slow oscillation method and the direct matching method leads us to the conclusion that an achromatic color and a chromatic color with equal (perceived) brightness do not have the same position along this measurable dimension of our internal color space. Instead, our results show that an achromatic color and a chromatic color are closest to each other in our internal color space when they produce a similar achromatic response in our visual system. Thus, we can conclude that this dimension in our internal color space is the achromatic response, not brightness or lightness. These results have profound implications for one-dimensional scales of achromatic response, such

as the L^* dimension in CIELAB or the Q and J dimensions in CIECAM02 and CIECAM16. We have presented direct experimental evidence that a one-dimensional scale of achromatic response exists in our internal color space and that such a dimension is not scale of brightness or lightness and cannot be so.

A common aim of models of the H-K effect has been to "correct" the achromatic response scale (e.g., L^*, Q, J) so that scale values of chromatic stimuli match the scale values of equally-bright achromatic stimuli. Our experiment demonstrates that brightness (and therefore lightness) are dependent on multiple dimensions and thus should not be modeled using a one-dimensional scale based on a single physical metric such as luminance. This conclusion does not mean that the one-dimensional scales are unneeded or incorrect; our experiment actually demonstrates the opposite: that such a scale does exist in our internal color space. Rather, we have demonstrated that the terms "brightness" and "lightness" should not be used to label any single dimension in color spaces, such as CIELAB or CAM16-UCS, which attempt to match our internal representation of color in three-dimensional space. "Value" could be a better term for achromatic response scales (e.g., L^*, Q, J) to avoid the misconception that colors with equal achromatic response scale values have equal brightness and lightness. "Value" comes from the achromatic scale of the Munsell color order system [68] and has a clear meaning without the connotation that colors with equal value has equal perceived brightness.

The other consequence of this experiment is to show that models of the H-K effect should be multidimensional, combining both the achromatic response dimension and a dimension or dimensions related to chromatic intensity. Examples of such scales of brightness include vector brightness in the color appearance models of Guth [57, 69] and vividness as proposed by Berns [70]. Recent work by Xie psychophysically measuring the zero-grayness threshold also holds potential for developing a measure of brightness that accounts for the colorfulness of stimuli [71, 72], building on previous work on zero-grayness and brilliance by Evans [73, 74]. Fairchild and Heckaman have discussed whether brightness and lightness should even be mapped in three-dimensional space or simply be modeled as an independent scale [75].

The next sections of this dissertation outline our efforts to combine achromatic and chromatic response dimensions to generate a two-dimensional brightness scale. Future work is planned to use the slowly oscillating gradient method to investigate the chromatic dimension of our internal perceptual color space as was accomplished for the achromatic dimension in this study.

3.3 Building a model of the Helmholtz-Kohlrausch effect from past data

3.3.1 Background

As the chromatic intensity (colorfulness) of a stimulus increases, so does its brightness, even if its luminance is held constant. This phenomenal crosstalk between our perceptions of chroma and brightness is called the Helmholtz-Kohlrausch (H-K) Effect. While the H-K Effect is well-known in the color science community, there does not exist a standard model of the effect [76]. Developing such a model is important for accurate prediction of brightness and lightness (brightness relative to the brightness of scene white).

Tracing the history of models of the H-K effect reveals the need for a new model. Early models (pre-1990s) focused on using chromaticity diagrams to describe the H-K effect. Sanders & Wyszecki (1963) [49] and Wyszecki (1967) [50] are examples of such models, which were fit to experimental brightness-matching data. An H-K model from Wyszecki (1967) [50] is shown in Figure 3.13. In such a model, contours on a chromaticity diagram are used to predict the "B/Y ratio" of a stimulus of given chromaticity. The B/Y ratio is the ratio of the luminance of an equally bright neutral patch to the luminance of the stimulus in question. For example, if a stimulus has a luminance of 100 cd/m^2 and a B/Y ratio of 1.35, the model predicts that it will appear equally bright to a neutral patch with luminance of 135 cd/m^2.

As can be seen from Figure 3.13, these models predict that the intensity of the H-K effect depends on both hue and chromatic intensity. An advantage of these models is their simplicity and their dependence on both hue and chroma. However, their predictions are limited by the limited gamut of stimuli used in the experimental data upon which the model was built. Another shortcoming of such models is that xy chromaticity coordinates are a two-dimensional representation of color that does not account for the brightness of stimuli. Thus, B/Y ratios predicted by Wyszecki-style models are invariant to stimulus brightness and luminance. This invariance may be problematic, especially given that stimulus appearance varies greatly with luminance at constant chromaticity.

Yoshinobu Nayatani's models of the H-K effect in the 1990s represent a shift from the diagram-based models described above to formula-based models [54, 55, 67, 77, 78]. For example, his 1997 model for object colors has the form [67]:

$$\frac{B}{Y} = 1 + \left[-0.1340 \cdot q(\theta) + 0.0872 \cdot K_{Br}\right] s_{uv}, \tag{3.29}$$

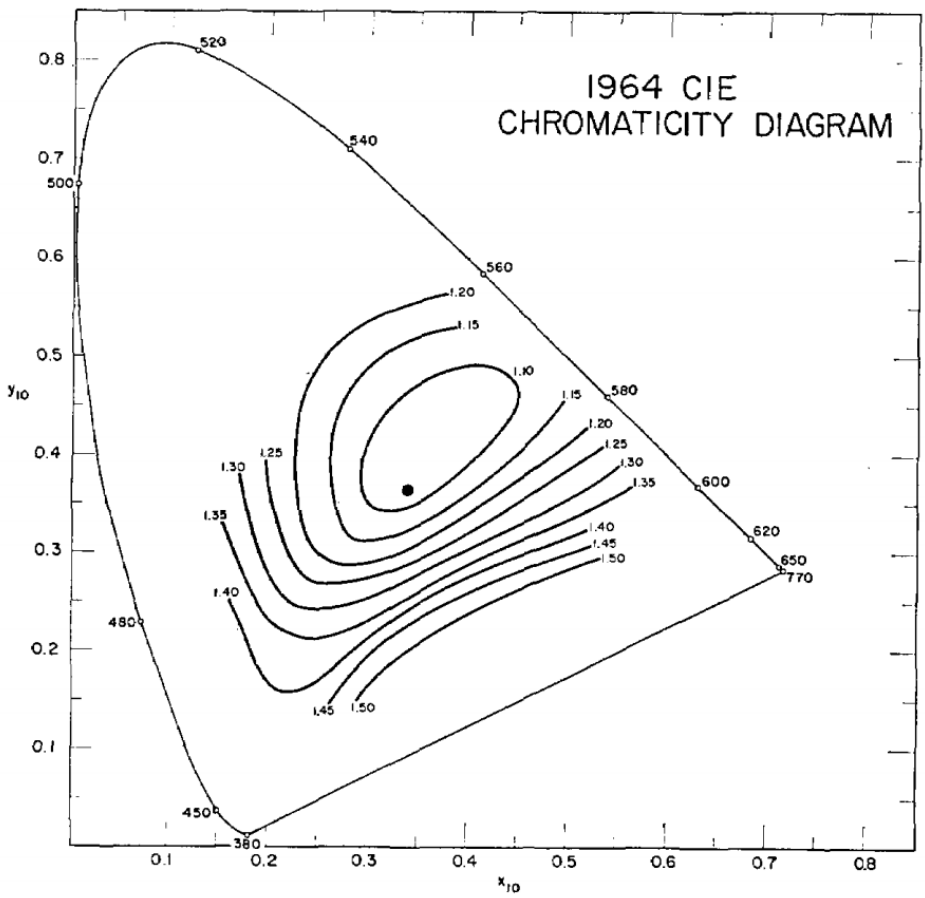


Figure 3.13: An example of a B/Y ratio contour diagram in xy chromaticity space from Wyszecki (1967). Reprinted with permission from [50] © The Optical Society.

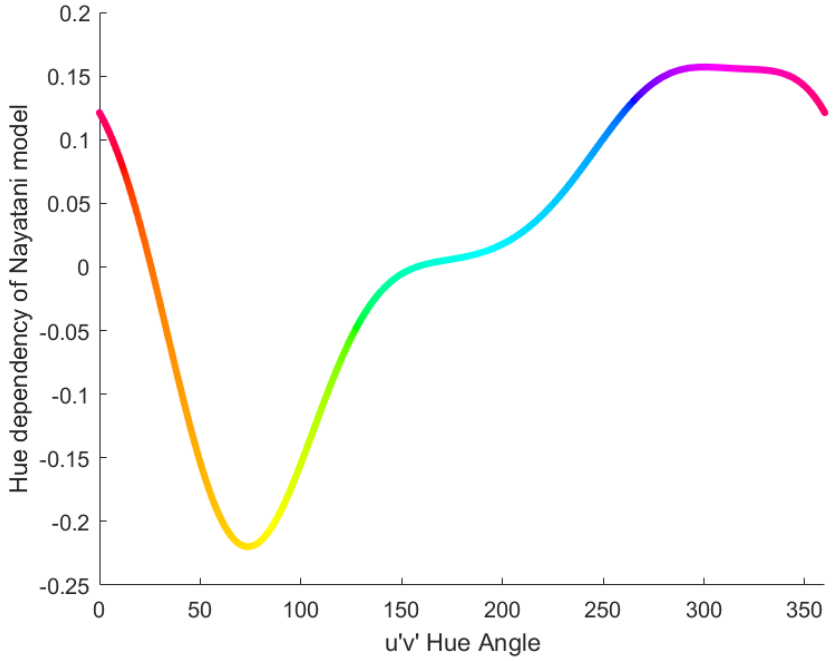


Figure 3.14: Hue angle dependency, $-q(\theta)$, of the 1997 Nayatani model as a function of $u'v'$ hue angle [67]. Colors are approximate.

where K_{Br} depends on the adapting luminance of the scene, s_{uv} is the saturation of the stimulus on the $u'v'$ chromaticity diagram and $q(\theta)$ is a function of the hue angle, θ, in $u'v'$. Like Wyszecki, Nayatani's model uses chromaticity coordinates to predict the strength of the H-K effect, this time in $u'v'$ chromaticity space as opposed to xy chromaticity space. Importantly, Nayatani's B/Y ratio is linearly related to saturation (s_{uv}) and has a hue dependency (Figure 3.14).

Unfortunately, several factors make it difficult to apply Nayatani's H-K model. First, his model did not consist of just a single equation (as presented in Equation 3.29) but rather four different equations that depended upon the application [67]. It is difficult to determine which equation is correct given a specific situation in which one may want to estimate the H-K effect. (For an example of a paper in which the researchers appeared unclear about which of Nayatani's models to use, see Liao et al. (2009) [79].) Specifically, Nayatani's decision to develop two separate models—Variable Achromatic Color (VAC) and Variable Chromatic Color (VCC)—was motivated by the results of a single study that did not follow standard brightness-matching procedures [77]. Furthermore, the quasi-mathematical derivation of the VCC equations from the VAC equations relied on semantic sleight-of-hand and circular reasoning [78]. The confusion about model choice and the faulty mathematics underlying some of Nayatani's derivations limit the usefulness of Nayatani's models.

One further limitation of the Nayatani models was outside of his control. In the late 1990s, the CIE approved the CIECAM97s color appearance model [80], which was then followed up by CIECAM02 and CAM16 [28, 36]. These color appearance models include equations for brightness and lightness that take input from all three CIEXYZ tristimulus values, not just from luminance (Y). This is in contrast to CIELAB, where lightness, L^*, is solely calculated from Y. Thus, simply multiplying the luminance of a stimulus by its B/Y ratio to account for the H-K effect is not compatible with the color appearance models that have become the state-of-the-art in predicting color appearance attributes. Furthermore, applying a B/Y ratio to a stimulus's CIE XYZ values before inputting into a color appearance model would also lead to unintended changes to hue and chroma predicted by the color appearance model. So, the H-K effect must be directly built into any color appearance model that seeks to account for the effect.

An example of an early effort to incorporate a model of the H-K effect was a CIELAB extension developed by Fairchild & Pirrotta (1991) [56]. Fit to experimental brightness-matching data collected by the investigators, their model avoided the incompatibility of the B/Y ratio described above by directly adjusting the CIELAB lightness (L^*) function to create a new lightness scale, L^{**}:

$$L^{**} = L^* + f_1(L^*)f_2(h^\circ)C^* \tag{3.30}$$

C^* is the CIELAB chroma of the stimulus. f_1 and f_2 are functions of CIELAB lightness and hue angle (Figure 3.15 and Figure 3.16). Like the Nayatani model, the Fairchild-Pirrotta model includes hue and chroma dependencies. Additionally, because the model does not use chromaticity coordinates as the independent variable, the Fairchild-Pirrotta model is also able to include a lightness dependency.

A key development in the Fairchild-Pirrotta was the transition from the multiplicative-factor based design of previous models to an additive model. While multiplicative compensation for the H-K effect works well in the luminance domain, which is linear to the amount of light, additive models are the norm for H-K models when operating directly on appearance scales, which are non-linearly compressed relative to the amount of light.

Recently, the Fairchild-Pirrotta model inspired the H-K extension to CIECAM02 proposed by Kim, Jo, Park, and Lee (2019) [63]. In fact, the Kim model directly transcribed the Fairchild-Pirrotta model [56] (including values) into CIECAM02. Unfortunately, given that CIECAM02 and CIELAB are distinct color spaces with different methods for calculating lightness, chroma, and hue, a simple transcription of values from one color space to another is not appropriate.

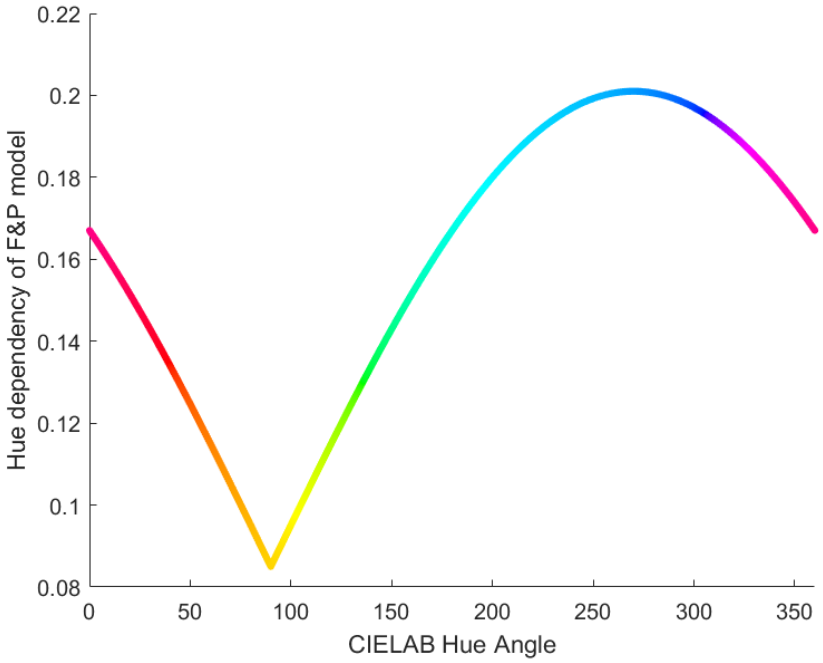


Figure 3.15: Hue angle dependency of the H-K effect in the Fairchild & Pirrotta CIELAB model [56]. Colors are approximate.

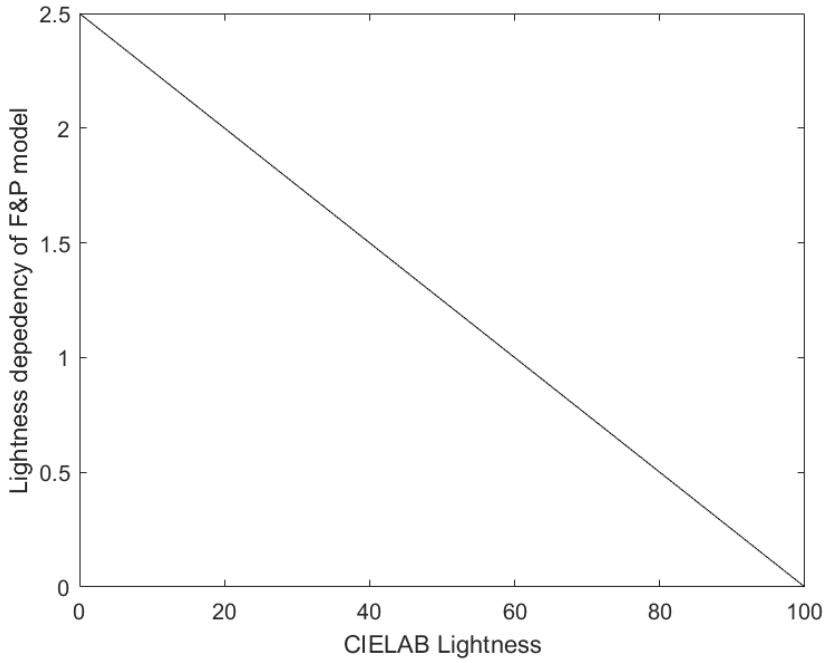


Figure 3.16: Lightness dependency of the H-K effect in the Fairchild & Pirrotta CIELAB model [56]. Theirs is the sole model with an explicit lightness dependency; they found that the strength of the H-K effect decreased with increasing lightness.

Several notable color appearance models account for the H-K effect through an additive contribution of colorfulness, M, to their equations for brightness, Q. The Hunt model from the late 1980s and early 1990s uses a simple addition of colorfulness to the achromatic signal, A [31]:

$$Q = (A + M)N_1 - N_2 \tag{3.31}$$

where N_1 and N_2 are constants based on viewing conditions and are irrelevant to the present discussion. Indeed, one of the strengths of the Hunt model was its accounting for the H-K effect in this way. However, this contribution of colorfulness was lost in the simplification of the Hunt model as it was translated to CIECAM97s. It was understandably problematic for the developers of CIECAM97s to use colorfulness in the calculation of lightness and brightness given that lightness was also used to calculate colorfulness (Equations 3.19 and 3.20).

Two recent color appearance models for unrelated and self-luminous stimuli, CAM15u [81] and CAM18sl [64], also account for the H-K effect through a contribution of colorfulness to brightness. For instance, in CAM15u, this H-K effect compensation takes the form:

$$Q = A + 2.559M^{0.561} \tag{3.32}$$

The QUGR glare model refined this H-K effect model of the CAM15u brightness formula by combining it with the Fairchild-Pirrotta H-K hue dependency function (Figure 3.15) [82]. As mentioned above regarding the Kim model [63], this method of simply transcribing an H-K hue dependency function from one context to another is unlikely to accurately predict the H-K effect.

CAM15u, CAM18sl, and QUGR were specifically designed for stimuli that appear self-luminous, such as light sources. CIECAM02 and CIECAM16 can also be applied to self-luminous stimuli, such as those generated by displays, when a reasonable white point can be inferred. In fact, the LUTCHI dataset, which forms the experimental foundation of these models, was conducted using displays. In contrast, CAM15u and CAM18sl are designed for situations where there is no meaningful white point, such as when viewing a traffic light at night. Thus, these models occupy a domain separate from the applications of CIECAM02 or CIECAM16. Nonetheless, their incorporation of additive colorfulness contributions to the equation for brightness reinforces the validity of this model structure.

From this review of past models of the H-K effect, it can be seen that there remains a lack of a model that fulfills all of the following attributes:

- Implemented directly in CIECAM02 and CIECAM16.

- Fit to experimental H-K effect data.
- Includes additive contribution of colorfulness to lightness or brightness.
- Dependent upon hue.

3.3.2 A model based on past studies

The H-K effect is typically measured by asking observers to directly match achromatic and chromatic stimuli so that they appear equally bright. Four sources of published experimental brightness-matching data on the H-K effect were assessed for incorporation in the fitting of a CIECAM02/CIECAM16 extension for the H-K effect: Sanders & Wyszecki [49], Wyszecki [50], Fairchild & Pirrotta [56], and Nayatani, Sobagaki, & Hashimoto [77]. Data used to derive the CAM15u and CAM18sl models were not considered because these data do not include direct brightness matching and these models cover a separate domain from CIECAM02 and CIECAM16.

Observers in the Sanders & Wyszecki experiment [49] performed direct brightness matching on a bipartite field between an achromatic stimulus and 96 stimuli of luminance 20 cd/m^2 and varying chromaticity. The luminance of the achromatic stimulus was adjusted by the observer to match the brightness of each chromatic stimulus. While the stimuli were self-luminous, they still fell into the category of related stimuli (as opposed to unrelated) as they were viewed against a well-lit background in a scene from which a plausible white point could be inferred. Twenty observers participated in the experiment. One flaw in the experimental design is that the color of the achromatic background was much bluer than the color of the achromatic reference stimulus (approximately 7500K versus 5000K). Thus, it is likely that the reference stimulus appeared yellowish to the observers due to chromatic adaptation to the bluer background. (Those interested in using these data should note that there is a typo in the y chromaticity coordinate for stimulus number 583/4. The correct y chromaticity, converted from the provided 1960 uv coordinates, is 0.415.)

The Wyszecki experiment [50] is notable in that chromatic tiles were used instead of the chromatic light of the Sanders & Wyszecki experiment. 76 observers scaled the lightness of 43 chromatic tiles against a scale of neutral tiles. Matches were made under daylight-colored illumination (approximately 6500K) with an adapting luminance of approximately 30 cd/m^2 (500 lux at the surface).

The Fairchild & Pirrotta experiment [56] also used physical samples. In this case, 11 observers matched the lightness of 36 glossy Munsell papers to an adjustable gray sample generated by a spinning Maxwell disc. The experiment was performed under a fluorescent

daylight simulator (6700K) with an adapting luminance of around 75 cd/m^2 (1200 lux). Unfortunately, the color of the chromatic samples under the experimental illumination was not reported in the paper. Instead, the paper lists the CIELAB coordinates of the Munsell papers under standard illuminant D65. To estimate the color of the samples in the experimental condition for use in this section, the CIE XYZ values of each tested Munsell paper were recalculated using CIE standard illuminant F7 (6500K) [62] and reflectance measurements of the glossy Munsell papers from the UEF Spectral Database.

Munsell papers were also used in the Nayatani, Sobagaki, & Hashimoto experiment [77]. However, as opposed to the other experiments listed here, where the chromatic stimulus had a fixed lightness and the achromatic reference stimulus varied in lightness, this experiment had its 4 observers estimate the equivalent lightness of a fixed reference Munsell paper on a lightness scale of each of 10 chromatic stimuli. Unusually, the reference stimulus was a chromatic yellow as opposed to the typical neutral. Due to the small number of observers and the atypical choice of reference stimulus, this experiment's data were excluded from our study.

The data from the three selected studies were input in CIECAM02 and CIECAM16 using the corrections to these color appearance models proposed in Section 3.1. (Note: Due to the similarity between CIECAM02 and CIECAM16, I will now focus on the CIECAM16 model. The model fit to CIECAM02 can be found in the Appendix of [83].) Based on the descriptions of the experimental viewing conditions, it was determined that the experiments were performed under "average" conditions for the purposes of CIECAM16. Additionally, the degree of adaptation was set to unity in CIECAM16 calculations. Using these parameters, CIECAM16 lightness, J, was estimated for each set of chromatic stimuli and matched neutral achromatic stimuli using white points derived from the description of each experiment. These data indicate that CIECAM16 systematically underestimates the lightness of stimuli as their chroma increases, meaning that CIECAM16 does not compensate for the H-K effect.

Based on our analysis of past models of the H-K effect (see Section 3.3.1), we hypothesized that an additive contribution of colorfulness or chroma to either lightness or brightness would properly account for the H-K effect. Unlike the Hunt model, which calculates brightness first and then derives lightness from brightness, CIECAM16 first derives lightness and then brightness. (CAM15u and CAM18sl do not predict lightness.) Thus, for CIECAM16, it is simplest to account for the H-K effect in the equation for lightness and then derive an H-K-compensated brightness from lightness. Thus, our hypothesized model has the form:

$$J_{HK} = J + f(h)C^{\gamma} \tag{3.33}$$

where J is lightness, $f(h)$ is the hue angle dependency and C is the chroma as calculated using our proposed chroma correction to CIECAM16 (Section 3.1). H-K-effect-compensated brightness, Q_{HK} is calculated directly from J_{HK}, following the proposed brightness correction to CIECAM16 (Section 3.1):

$$Q_{HK} = \left(\frac{2}{c}\right)\left(\frac{J_{HK}}{100}\right)(A_w) \tag{3.34}$$

This linear relationship between lightness and brightness ensures that the additive contribution of chroma to lightness is mathematically equivalent to adding colorfulness to brightness, à la the Hunt model (Equation 3.31).

The form of the hue angle dependency, $f(h)$, was hypothesized to be a simplified version of the hue dependency in the Nayatani model:

$$f(h) = \alpha_1 \cos h + \alpha_2 \cos 2h + \alpha_3 \sin h + \alpha_4 \sin 2h + \alpha_5 \tag{3.35}$$

γ and α_1 to α_5 were parameters to be set by least squares fitting to the experimental data, minimizing the difference in J_{HK} between the chromatic stimuli and the experimentally-matched reference stimuli. The results of such fitting leads to the following values:

$$J_{HK} = J + f(h)C^{0.587} \tag{3.36}$$

$$f(h) = -0.160 \cos h + 0.132 \cos 2h - 0.405 \sin h + 0.080 \sin 2h + 0.79 \tag{3.37}$$

The hue dependency function, $f(h)$, is shown in Figure 5. The hue dependency follows a similar trend to the hue dependencies in the Nayatani and Fairchild & Pirrotta models (Figure 3.14 and Figure 3.15). However, direct comparison of the hue dependencies is not possible because the current model operates on CIECAM16 lightness whereas the previous models operate on luminance and CIELAB lightness, respectively.

The performance of this model is evaluated alongside other models of the H-K effect on these and other data sets in Section 3.4.5.

3.4 A model of the H-K effect built from new experimental data

In this section, we present a subtle innovation in the method of direct brightness matching by fixing the chroma difference between the reference and test patches. In the traditional

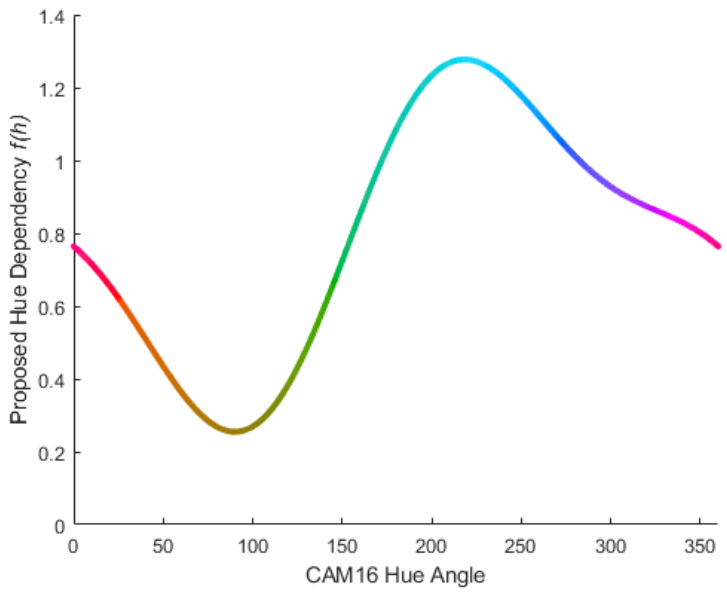


Figure 3.17: Hue dependency of the proposed H-K effect model (Equation 3.37). Colors are approximate.

design of brightness matching experiments, the difference in chroma between the achromatic test patch and the chromatic reference patch fluctuates as the reference patches change in chroma. By removing the restriction that the test patch be achromatic, we could fix the chroma difference between the two patches, with the formerly achromatic test patch increasing in chroma as the reference patch increases in chroma. The data collected using this method measure not just the strength of the H-K effect, but how the change in the strength of the H-K effect varies with the lightness, chroma, and hue of the stimuli.

Measuring the brightness difference between two chroma levels is not possible using traditional methods that always involve an achromatic patch. One could try to use achromatic-chromatic patch matches to infer brightness differences between chromatic patches, but using the common variable-achromatic method, one would only be able to infer chromatic patch differences when observers were lucky enough to match two chromatic patches to the same achromatic patch. If one instead fixes the achromatic patch and has observers adjust the chromatic patch, one then becomes constrained by the gamut of the display, and the limited lightness range at higher chromas leads to bias in the results. Thus, our method of constant chroma differences between the actual stimuli is the only feasible method for deriving this type of data in emissive viewing modes.

Our calculus-based approach to modeling these data sets rules that models of the H-K effect must follow. We present here a new model of the H-K effect built on top of the CIECAM16 color appearance model [4] (following key corrections to CIECAM16 proposed by

Hellwig & Fairchild [83]) and based on the results of a series of three experiments conducted using our refined method of direct brightness matching.

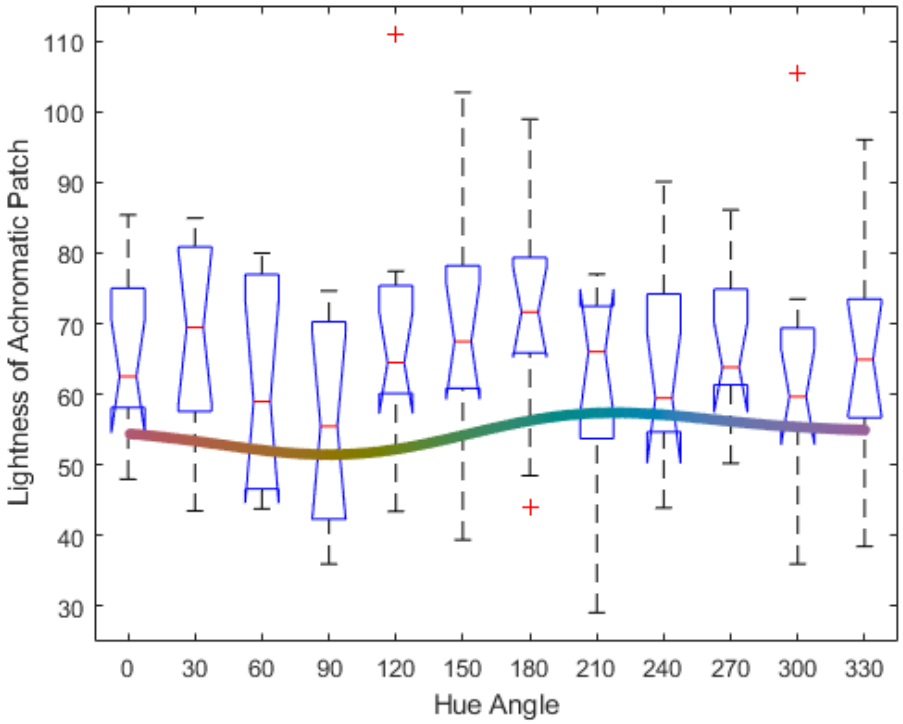
3.4.1 Experiment 1: Preliminary study

Methods

Direct brightness matching was performed on 37 pairs of stimuli. For each stimulus presentation, the observer was presented with two patches, each subtending approximately 2° of visual angle and separated by 1° of visual angle, against a random noise background. The observer was able to adjust the luminance of the lower-chroma test patch and was asked to match the brightness of the test patch to the brightness of the higher-chroma reference patch. There was no time limit for observers to complete each match; the median response time was 15 seconds. Stimuli were presented in random order.

The colors of the stimuli were chosen in an attempt to independently measure the dependence of the H-K effect on hue, lightness, and chroma. The hue dependency was tested by 12 pairs of stimuli evenly spaced in CIECAM16 hue angle, with lightness values of 50, and with chromas of 0 and 20. The lightness and chroma dependencies of the H-K effect were measured by sampling a range of lightnesses and chromas in yellow ($h = 90°$) or purple ($h = 330°$) hue planes. The CIECAM16 achromatic lightness of the reference patch was varied from 30 to 90 in steps of 20 and the CIECAM16 chroma of the reference patch was varied from 20 to 60 in steps of 10 as allowed by the gamut of the display. In each case, the chroma difference between the reference and test patches was 20. All CIECAM16 calculations were performed following key corrections proposed in Section 3.1. The CIECAM16 degree of adaptation was set to 1 and the surround condition was set to dark for all calculations.

A D65 white point at 500 cd/m^2 was used for all stimuli and the background. A random noise background was used instead of a uniform background to reduce the effects of simultaneous contrast and crispening, which are not accounted for by CIECAM16. The background was a pixelated semi-uniform noise distribution with 0.01° pixels. The range of background pixel luminances ranged from 5.5 cd/m^2 to 500 cd/m^2, with the uniformity of the random luminance distribution modified so that the average background luminance was mid-gray (94.1 cd/m^2). The background subtended approximately 20° of visual angle horizontally and 10° vertically. The experiment was run in a dark room. The stimuli were displayed on a 32" Asus ProArt PA32UCX monitor driven by an NVIDIA Quaddro P400 graphics card on a Windows computer. The stimuli were generated using MATLAB and Psychtoolbox [65]. 11 expert observers participated in the experiment.



.

Figure 3.18: Partial results of Experiment 1, with the box plots showing the achromatic lightness of achromatic patches matched to the chromatic patches with achromatic lightness 50 and chroma 20 at various hue angles. ANOVA indicated that there was no statistically significant difference between these hues. The chromatic line is the hue dependency from the H-K model fit to previously published experimental data proposed in Section 3.3

Results

The H-K effect is measured by comparing the achromatic lightness (CIECAM16 J) of the lower-chroma test patch to the fixed achromatic lightness of the higher-chroma reference patch. The lower-chroma patch usually has a higher achromatic lightness than the matched higher-chroma patch because the increased chroma typically increases the observers' perception of brightness.

The box plot in Figure 3.18 displays the results of the attempt to measure the hue dependency of the H-K effect using the 12 evenly-spaced stimuli. One-way ANOVA was performed on these data and indicated that there was not a statistically significant effect of hue on the results ($p = 0.57$).

In a later experiment, we discovered that a few of the matches made by a few observers were limited by the gamut of the display. The resulting clipping of the range of responses could influence the mean value of the responses. Thus the median was calculated for all

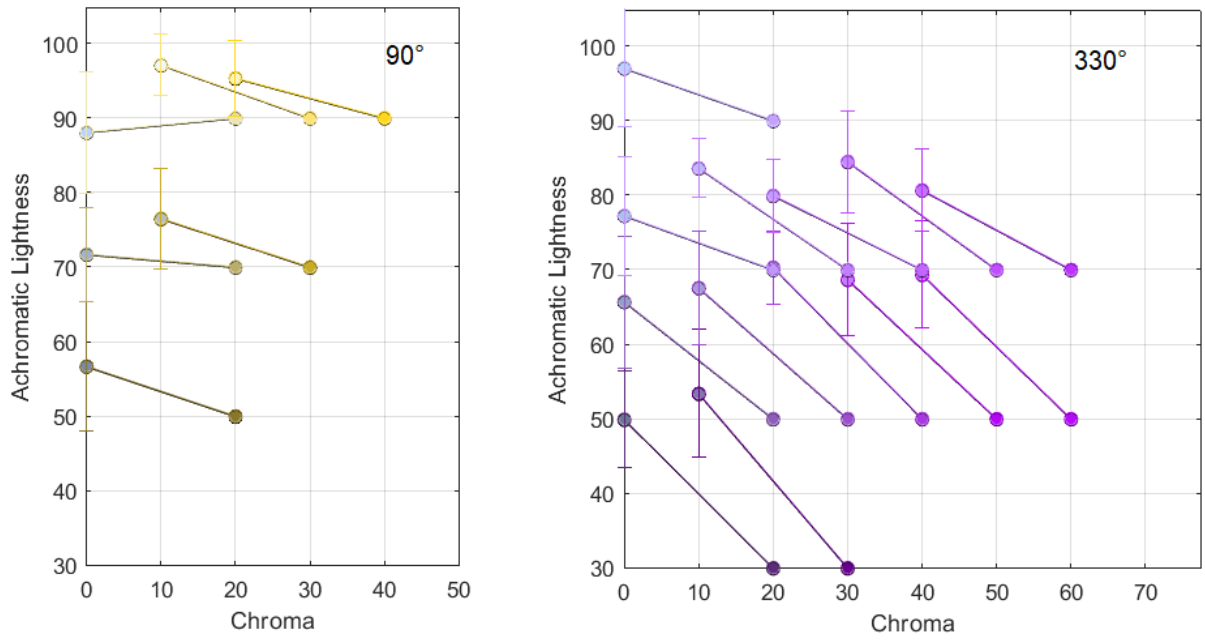


Figure 3.19: Partial results of Experiment 1, showing pairs of brightness-matched stimuli in the revised CIECAM16 achromatic lightness-chroma plane.[83] Steeper lines indicate a stronger H-K effect. The results for the 330° hue plane demonstrate the trends found throughout our experiments: the strength of the H-K effect is constant with chroma and decreases with increasing achromatic lightness.

results to eliminate this potential source of bias. Additionally, using the median helps avoid questions about whether the mean should be calculated using perceptually-linear or light-linear scales.

Figure 3.19 shows the results of the lightness and chroma sampling in the purple and yellow hue planes. The median achromatic lightness set by observers was calculated for each pair of stimuli. A larger difference between the achromatic lightness of the matched stimuli indicates a larger contribution of chroma to perceived brightness, which we refer to as the strength of the H-K effect. Two-way ANOVA performed on the data in the purple hue plane (which had the most data) indicated a significant effect of lightness on H-K effect strength ($p < 10^{-6}$), with increasing lightness leading to weaker H-K strength. ANOVA also indicated that there was not a significant effect of chroma on H-K strength ($p = 0.27$). Furthermore, ANOVA found that the H-K effect was significantly stronger in the purple hue plane than the yellow hue plane ($p = 0.008$), although this conclusion may be compromised by the fact that it was not possible to generate lower-lightness stimuli in the yellow hue plane due to the shape of the display's gamut.

Overall, the results indicated that more data needed to be collected to build a complete model of the H-K effect. While our explicit measurement of the hue dependency of the H-K effect in CIECAM16 space did not find a significant effect, there was a significant difference in the data between the better-sampled purple and yellow hue planes. Complete sampling of additional hue planes would help resolve this discrepancy in the results.

3.4.2 Experiment 2: Measuring the full gamut

Methods

A second experiment was performed to completely sample the dependence of the strength of the H-K effect in 12 hue planes in CIECAM16 color space. An identical brightness matching technique was used as in Experiment 1. The CIECAM16 achromatic lightness of the reference stimulus was varied from 30 to 90 in steps of 20 and its CIECAM16 chroma was varied from 20 to the edge of the gamut (up to 80) in steps of 10 at each hue. This even sampling in color space led to uneven numbers of stimuli at different hues due to the difference in the size of the display's color gamut at each hue. Thus, for hues with fewer pairs, the lightness sampling step size was decreased from 20 to 10, and for hues with too many pairs of stimuli, the chroma sampling step size was increased from 10 to 20. This sampling led to 134 total stimuli at the 12 hues, providing about 11 pairs per hue (ranging from 7 to 15).

In Experiment 2, a Samsung S95B QD OLED display was chosen for the experiment to take advantage of the display's wide color gamut. The luminance of the white point for the displayed stimuli was 400 cd/m^2, although observers were allowed to adjust the test patch up to approximately 690 cd/m^2. A white point with a correlated color temperature of 12000 K was used to better match the resting white point of an observer in a dark room [84]. The background subtended approximately 20° of visual angle horizontally and 10° vertically. Repeated stimuli between Experiments 1 and 2 indicated that this change in white point did not affect the results. 25 observers participated in the experiment. The observers included a mix of expert and naïve observers, although no significant difference was found in the results between the two groups. Median response time remained similar at 12 seconds.

Results

The results for the 12 hue planes tested in Experiment 2 are shown in Figure 3.20. Once again, ANOVA revealed a significant effect of reference patch achromatic lightness on the strength of the H-K effect ($p < 0.0005$) and no significant effect of reference patch chroma on the strength of the H-K effect ($p = 0.83$), which is represented by the slope of the lines in

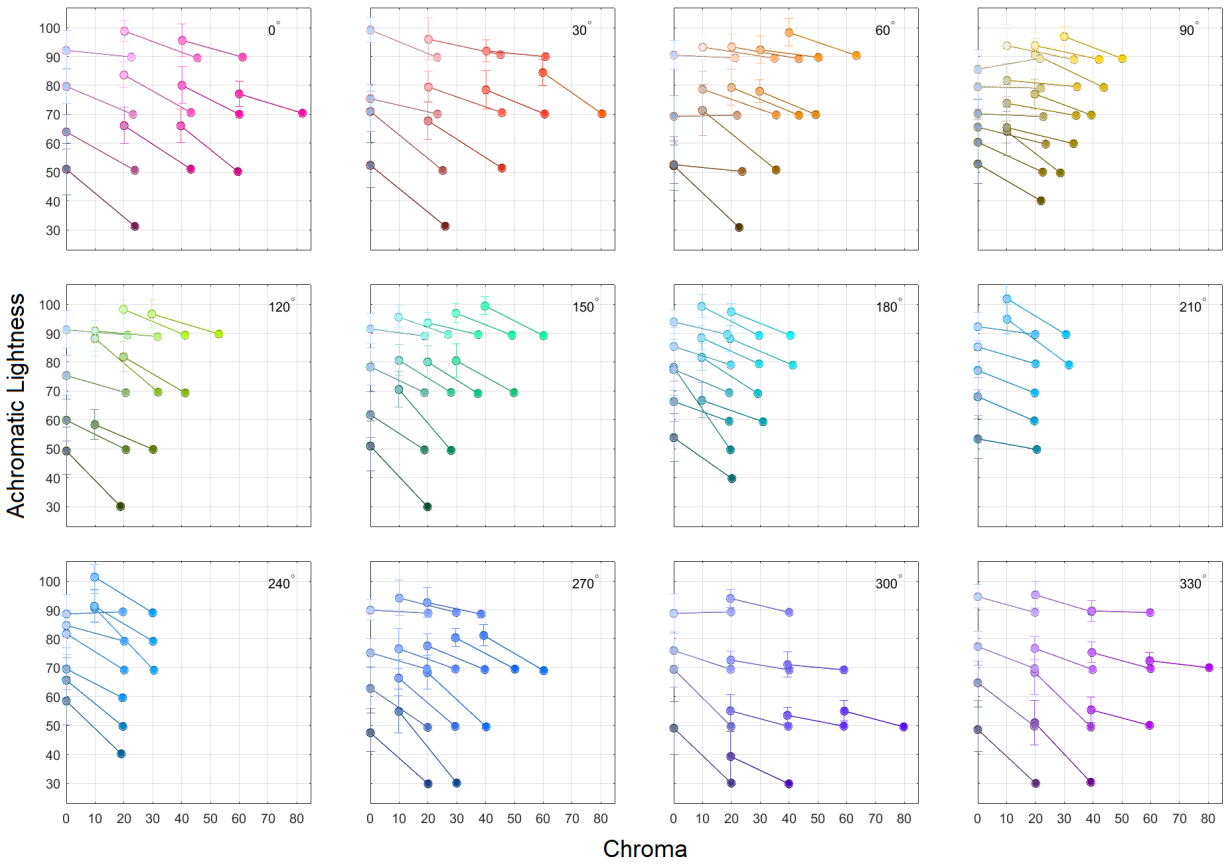


Figure 3.20: The results of Experiment 2, showing pairs of brightness-matched stimuli in the revised CIECAM16 achromatic lightness-chroma plane. [83]

Figure 3.20. The strength of the H-K effect did seem to fluctuate slightly with hue, especially for stimulus pairs where the test patch was achromatic. However, when hue was viewed as a continuous variable with one degree of freedom, ANOVA indicated that there was not a statistically significant effect of hue on the strength of the H-K effect ($p = 0.83$), although it would be unlikely for such analysis to detect a periodic hue dependency. There will be additional discussion on the hue dependency of our experimental data in Section 3.4.4.

3.4.3 Experiment 3: The transitivity of brightness matches

Methods

The use of constant chroma differences between the reference and test patches in Experiments 1 and 2 was a key innovation of our research that will allow for a new form of data analysis (Section 3.4.4). However, it was important to confirm that this method of brightness

matching is transitive—that if Patch A with a chroma of 40 is matched to Patch B with a chroma of 20, and Patch B with a chroma of 20 is matched to Patch C with a chroma of 0, then Patch A would also be matched to Patch C. The transitivity of brightness matches would confirm that our method of stimuli choice is equally valid to the traditional method of only using achromatic adjustment patches.

65 pairs of stimuli were chosen at CIECAM16 lightnesses and hues where, in Experiment 2, multiple chroma levels had been tested. In this experiment, larger chroma differences (ranging from 30 to 60 CIECAM16 chroma units) were used to test whether the match made across larger chroma differences could be predicted by integrating the smaller chroma differences tested in Experiment 2. The same physical setup was used in Experiment 3 as in Experiment 2. 12 observers (only 1 naïve) participated in the experiment. Median response time remained similar at 11 seconds.

Results

The transitivity of direct brightness matches was tested by seeing if the matches made in Experiment 3 could be predicted by integrating the matches made in Experiment 2. Two-sample t-tests were used to statistically test this hypothesis, using the Šidák correction to maintain a family-wise error rate of 0.05. The null hypothesis was not rejected for all 65 trials with all p-values exceeding the Šidák-corrected α value of 0.0008. Only four tests returned p-values less than 0.05 and only one p-value was less than 0.01, quantities which would be expected from running 65 t-tests. Thus, we can conclude that brightness matches are transitive and our method of stimulus choice with a constant chroma difference is valid.

3.4.4 Building a model

The novel use of equal chroma differences between test and reference patches in direct brightness matching allows for a calculus-based approach to building a model, which was unattainable with data derived from previous methods of brightness matching. Specifically, we can quantify how the slope of the equal brightness lines connecting brightness-matched stimuli (Figures 1 and 2) varies with chroma, lightness, and hue. As mentioned above, we found that the chroma of the reference patch did not have a significant effect on the slope of the equal brightness lines, but that the negative slope of the equal brightness lines increased (became less steep) with increasing achromatic lightness.

If we express the equal brightness line as a function in the achromatic lightness and chroma plane, the function's derivative should be negative, invariant to chroma (C), and increase (become less steep) with increasing achromatic lightness (J). We assessed three

potential forms for the equation for the derivative of the equal brightness line.

$$\frac{dJ}{dC} = aJ - b \tag{3.38}$$

$$\frac{dJ}{dC} = -ae^{-bJ} \tag{3.39}$$

$$\frac{dJ}{dC} = \frac{-a}{J-b} \tag{3.40}$$

In these formulas, $\frac{dJ}{dC}$ represents the strength of the H-K effect, with a more negative value meaning a stronger H-K effect. We found that the formulas for H-K compensated lightness, J_{HK} derived from each of these potential formulas for $\frac{dJ}{dC}$ could equally fit our experimental data since they meet the criteria discussed above. However, their behavior differed significantly at higher achromatic lightness. The linear derivative, Equation 3.38, predicts that above a certain lightness, we will see an inverted H-K effect, where more chromatic stimuli appear darker than less chromatic stimuli! This implausible prediction allowed us to eliminate Equation 3.38.

Two other factors lead us to choose the inverse-form derivative, Equation 3.40, over the exponential-form derivative, Equation 3.39. First of all, the slower decrease in predicted strength of the H-K effect at high lightness by Equation 3.40 is aligned with our belief that the H-K effect should still be present for stimuli above diffuse white [64]. The inverse-form derivative nicely predicts a constant strength of the H-K effect at constant saturation as lightness increases. Secondly, we will see that the formula for H-K compensated lightness that results from solving Equation 3.40 more closely resembles some previously published formulas for H-K compensated brightness [57, 85]. These reasons serve as sufficient justification to proceed with the inverse-form derivative, Equation 3.40, outside of empirical evidence to separate the two potential models.

The differential Equation 3.40 has the general solution for equal-brightness lines in the achromatic lightness and chroma plane:

$$J = \sqrt{k - \mu C} + \beta, \tag{3.41}$$

where μ and β are positive constants. The constraint that the H-K compensated lightness, J_{HK}, is equal to the achromatic lightness, J, when C is zero allows us to solve for the constant of integration, k:

$$J = \sqrt{(J_{HK} - \beta)^2 - \mu C} + \beta \tag{3.42}$$

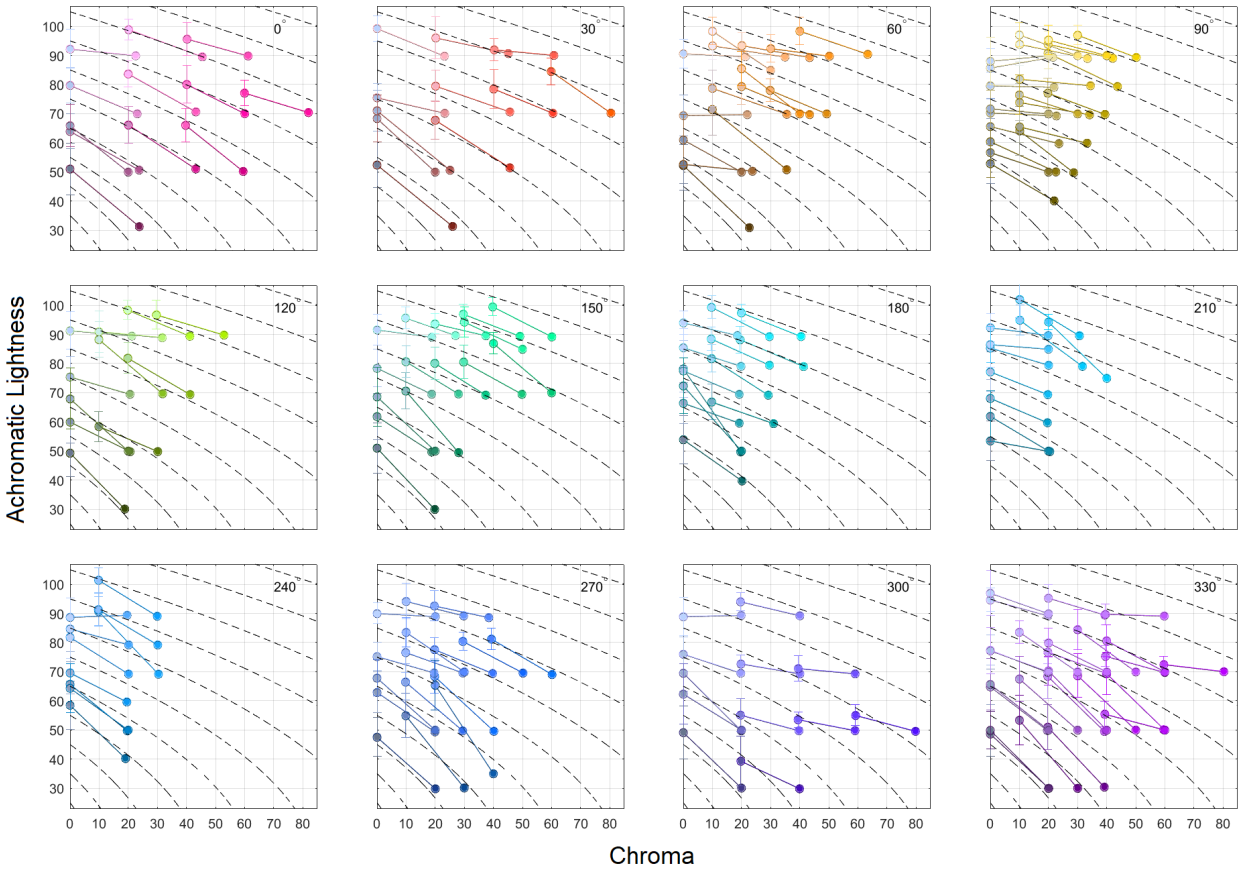


Figure 3.21: Our proposed model (Equation 3.44) compared to the data from Experiments 1 and 2

When solved for H-K compensated lightness, this formula becomes:

$$J_{HK} = \sqrt{(J-\beta)^2 + \mu C} + \beta \tag{3.43}$$

This formula, however, has the domain restriction that $J > \beta$. Fitted values of β were typically very small, near 1, and we found that the β could be set to zero with minimal loss in performance on our data. When fit to all data that we collected across the three experiments, we found that a μ value of 66 minimized the root-mean-square error. Thus, our complete formula for H-K compensated lightness is:

$$J_{HK} = \sqrt{J^2 + 66C} \tag{3.44}$$

This model is shown alongside the data from Experiments 1 and 2 in Figure 3.21.

The revised, linear relationship between lightness and brightness can be used to predict

H-K compensated brightness, Q_{HK}, from H-K compensated lightness:

$$Q_{HK} = \left(\frac{2}{c}\right)\left(\frac{J_{HK}}{100}\right) A_w, \tag{3.45}$$

where A_w is the achromatic signal for the reference white and c is a term that depends on the surround viewing conditions.

A surprising aspect of the proposed model is that it does not contain an explicit hue dependency. The revised version of CIECAM16 used for this model includes a correction to the eccentricity function that improves the uniformity of the chroma scales relative to hue (see Equation 3.28 and Figure 3.5 in Section 3.1.3). Thus, we hypothesize that the improved uniformity of our chroma values may obsolete the need for an explicit hue-dependency in our formula for H-K compensated lightness. The necessity of a hue dependency in many other models of the H-K effect may be simply due to nonuniformities in the chroma or saturation scale used in those models relative to hue. Indeed, when our data is converted from CIECAM16 to CIELAB, the data become less uniform relative to hue, and to fit a model that matches our data, a hue dependency is once again required.

Following our hypothesis that the H-K hue dependency is due to chroma non-uniformity, we simply followed the same procedure that was used to fix the eccentricity function in CIECAM16 to calculate a hue dependency for our CIELAB H-K function. Data from the Munsell system [41] was converted to CIELAB and the average ratio of Munsell chroma to CIELAB chroma was calculated for each Munsell hue. These ratios represent a correction factor for CIELAB chroma, to which we fit the following function (Figure 3.22):

$$\begin{aligned} f(h) = {} & 0.0258\cos(h) + 0.0648\cos(2h) + 0.0138\cos(3h) \\ & +0.0107\cos(4h) - 0.1872\sin(h) + 0.0339\sin(2h) \\ & +0.0152\sin(3h) - 0.0121\sin(4h) + 1 \end{aligned} \tag{3.46}$$

Shockingly, not only does this function—fitted solely to Munsell chroma data—closely resemble the hue dependencies of other models of the H-K effect (Figure 3.22), but it also allows the CIELAB version of our H-K model to almost match the performance of our CEICAM16 H-K model. Thus, we come to the important conclusion that chroma, when properly scaled, equally feeds into our perception of lightness and brightness at all hues. Our full CIELAB model, fit to our experimental data, then takes the form:

$$L^*_{HK} = \sqrt{L^* + 32 f(h) C^*}, \tag{3.47}$$

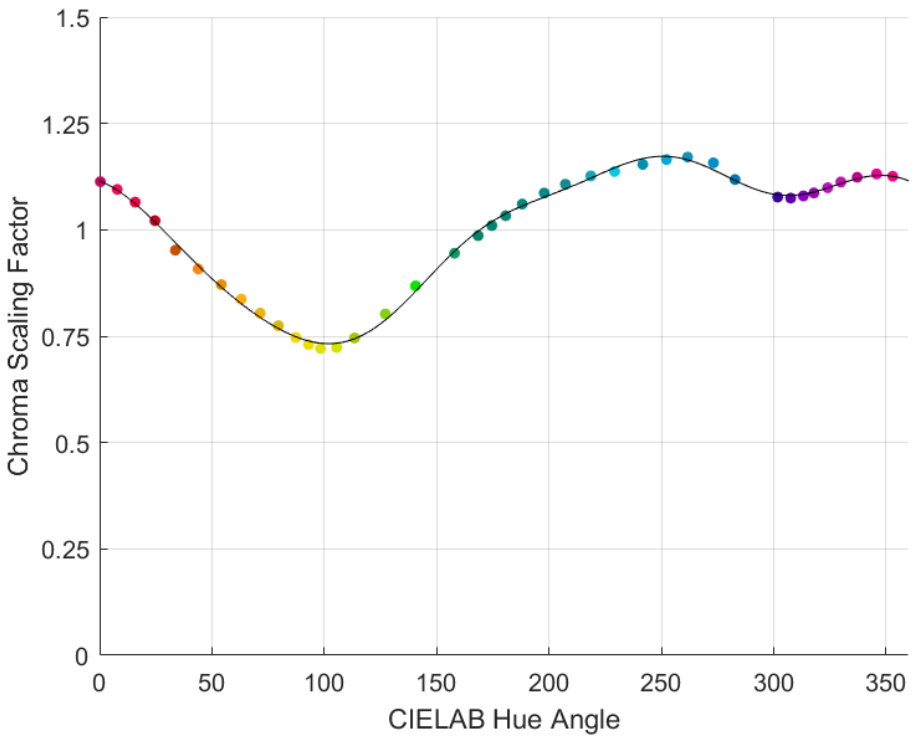


Figure 3.22: The eccentricity function (Equation 3.46) fit to increase the chroma uniformity of CIELAB using data (colored points) from the Munsell color order system.[41] Despite being fit only for chroma uniformity, the function closely resembles the hue dependency of models of the H-K effect fit to brightness-matching data.[56, 86]

where L^*_{HK} is H-K compensated lightness, L^* is CIELAB (achromatic) lightness, C^* is CIELAB chroma, and $f(h)$ is the hue angle dependency (Equation 3.46).

Despite the lack of an explicit hue-dependency in our CIECAM16 model, our data do show some differences between hues (Figure 3.21). The first step from achromatic to chromatic sample for yellow and green hues (hue angle 90° to 150°) at high lightness experiences an extremely weak H-K effect. However, this weakness disappears when two chromatic samples are compared and the data better follow our model. These results suggest an unaccounted-for effect when comparing achromatic and yellow stimuli, which could have also encouraged other models to include the common hue dependency that weakens the H-K effect for yellow. However, since the overall behavior for yellow and green follows our proposed trends, we omit that achromatic oddity from our model.

Additionally, a weak H-K effect can be observed for high chroma purple stimuli (hue angles 300° and 330°) in Experiment 2 (Figure 3.20), which seems to conflict with our model. However, no such weakness was observed for the same hue (330°) in Experiment

1 (Figure 3.19), which does match the model (Figure 3.21). Thus, this isolated weak H-K effect is most likely due to some isolated effect in the experimental setup, to which it would not be valid to fit a hue dependency.

3.4.5 Model Performance

We compared the proposed models (Equations 3.44 and 3.47) to four other models of the H-K effect [56, 63, 86, 87] alongside default CIECAM16 and CIELAB on our data and five other brightness matching datasets [49, 50, 56, 71, 86, 88]. (Hellwig '22 refers to the model described in Section 3.3.) These data consist of CIE XYZ values for pairs of stimuli that were matched in brightness or lightness by observers. An ideal model will produce equal lightness and brightness predictions for each pair of stimuli; the difference between predicted lightness values represents the error of the models. Root-mean-square error values are shown in Table 3.3, with the lowest errors underlined. Our proposed models perform very well on the three recent datasets but poorly on the three older datasets.

Looking at the mean error (Table 3.4) can provide more context for comparing the models. These values represent the average difference in predicted lightness values between the more chromatic and less chromatic stimuli, so a negative mean error means that a model underpredicts the strength of the H-K effect, and a positive mean error means that a model overpredicts the strength of the H-K effect. For example, as expected, CIECAM16 and CIELAB have universally negative mean errors, as they do not explicitly compensate for the H-K effect.

Universally, the analyzed models of the H-K effect have lesser mean errors (either smaller positive or greater negative values) on the four newer datasets—our dataset, the High dataset [86], the Seong dataset [88], and the Xie dataset [71]—than on the three older datasets. This indicates that the strength of the H-K effect is greater in the four newer datasets than in the three older datasets. This could be due to differences in experimental design (Table 3.5). The use of physical, reflective samples is unique to the Fairchild [56] and Wyszecki [50] datasets and there could be unknown differences in the H-K effect due to this difference in medium. Additionally, the Wyszecki experiment [50] and the Sanders experiment [49] both did not separate the test and reference stimuli. Presenting adjacent samples can lead observers to make minimally-distinct-border judgments, which are distinct from heterochromatic brightness matches and tend to lead to luminance-like matches [51].

These experimental differences can help explain why our models and others perform well on some of the datasets but not on others. Our proposed models appear to be well-tuned to the three newer datasets, with low RMSE values and mean errors close to zero. Meanwhile,

Dataset	CIECAM16-based models			
	Proposed	Hellwig '22	Kim '19	CIECAM16
Current	<u>6.0</u>	11.3	<u>5.6</u>	14.3
High '23	2.3	3.1	4.5	6.0
Seong '23	<u>6.3</u>	12.8	26.6	20.1
Xie '21	<u>6.7</u>	12.6	<u>7.0</u>	17.5
Fairchild '91	7.3	<u>3.4</u>	6.8	<u>4.6</u>
Wyszecki '67	4.5	1.6	4.2	4.6
Sanders '63	6.2	<u>1.8</u>	7.3	4.9
Dataset	CIELAB-based models			
	Proposed	High '23	Fairchild '91	CIELAB
Current	<u>5.9</u>	7.4	8.7	14.6
High '23	2.2	<u>1.7</u>	2.3	5.2
Seong '23	13.2	9.0	12.2	26.6
Xie '21	<u>10.4</u>	<u>9.7</u>	13.7	21.7
Fairchild '91	7.0	4.9	<u>4.6</u>	<u>4.9</u>
Wyszecki '67	3.7	1.9	<u>1.2</u>	4.8
Sanders '63	5.5	6.5	<u>2.1</u>	5.9

Table 3.3: RMSE for 8 models of brightness and lightness on brightness matching data from 7 experiments. All data is scaled from 0 to 100 (the traditional lightness scale) to allow for comparisons between datasets and models. Underlined values indicate the best performance or lack of statistical difference from the best performance amongst all eight models (two-sample t-test with $\alpha = 0.05$)

Dataset	CIECAM16-based models			
	Proposed	Hellwig ’22	Kim ’19	CIECAM16
Current	-0.37	-8.4	-2.3	-12
High ’23	-0.087	-1.1	-3.4	-4.8
Seong ’23	-4.7	-11.2	-21.5	-18.1
Xie ’21	-3.0	-8.4	0.030	-14
Fairchild ’91	5.6	1.2	5.4	-1.9
Wyszecki ’67	3.7	-1.3	3.6	-4.3
Sanders ’63	3.3	-1.6	4.4	-4.3
Dataset	CIELAB-based models			
	Proposed	High ’23	Fairchild ’91	CIELAB
Current	-0.72	-2.2	-4.4	-12
High ’23	-0.48	0.22	-0.80	-4.1
Seong ’23	-11.5	-4.4	-10.5	-23.5
Xie ’21	-6.6	-5.8	-8.7	-17
Fairchild ’91	5.2	2.8	1.9	-2.0
Wyszecki ’67	3.0	0.64	-0.60	-4.6
Sanders ’63	3.4	2.8	0.10	-4.8

Table 3.4: Mean error for 8 models of brightness and lightness on brightness matching data from 7 experiments. Negative values indicate the model underpredicts the strength of the Helmholtz-Kohlrausch effect in the data and positive values indicate that the model overpredicts the H-K effect.

Data set	No. Stimuli	No. Observers	Viewing Mode	Match Criteria	Sample Separation
Current	236	11-25	Emissive	Brightness	Yes
High '23	120	19-21	Emissive	Lightness	Yes
Seong '23	285	20	Emissive	Brightness	Yes
Xie '21	15	12	Emissive	Zero-grayness	Yes
Fairchild '91	36	11	Reflective	Lightness	Yes
Wyszecki '67	43	37-39	Reflective	Lightness	No
Sanders '63	96	20	Emissive	Brightness	No

Dataset	Lightness Range	L_W (cd/m^2)	Y_{BG}	Surround
Current	30-90	400, 500	18	Dark
High '23	80-100	120	20	Dim
Seong '23	100	30, 95, 300	0.0005	Dark
Xie '21	100	50, 100, 200	18	Dark
Fairchild '91	30-70	382	20	Average
Wyszecki '67	50	159	30	Average
Sanders '63	50	100	20	Unknown

Table 3.5: Experimental parameters for brightness matching datasets.

positive net errors for the proposed models on the three older datasets indicate that our models overpredict the brightness of more chromatic colors in these data, which follows the above discussion on the relative weakness of the H-K effect in these data. A similar trend of well-predicting the strength of the H-K effect in the newer data but overpredicting it in the older data can be seen for the Kim and High models (Table 3.4). Meanwhile, the Hellwig '22 model and the Fairchild model match the strength of the H-K effect in the older data (to which they were both trained) while underpredicting the H-K effect when applied to the newer datasets (negative mean errors in Table 3.4). Furthermore, analysis of the individual errors for each model suggests that the Hellwig '22, High, and Fairchild models underpredict the H-K effect for darker colors in comparison to the other models.

3.4.6 Dicussion

Our unique method of brightness matching—choosing pairs of stimuli with constant differences in chroma—revealed two important features of the H-K effect:

1. The change in strength of the H-K effect is invariant to starting chroma; that is, you see the same gain in brightness from increasing the chroma from 10 to 15 as from 5 to

10. (The stimulus with chroma 15 is still brighter than the other two.)

2. The strength of the H-K effect decreases with increasing achromatic lightness; that is, the rate at which increasing chroma contributes to perceived brightness and lightness is lower at higher achromatic lightness.

The models proposed in this section follow these laws of the H-K effect; no other considered model does so. The strengths of the H-K effect in the Hellwig '22 model, the Kim '19 model, and the Fairchild '91 model are not invariant to chroma, which can be seen by taking the derivative of these models' formulas for equal-brightness lines. Additionally, the Hellwig '22 and High '23 models do not predict the strength of the H-K effect to decrease with increasing achromatic lightness.

In addition to its high performance, our proposed CIECAM16-based model is much simpler than others; it only uses one fitted parameter versus six parameters for the Hellwig model [87] and five parameters for the High [86], Kim [63] and Fairchild [56] models. Although the proposed CIELAB model does contain more parameters (Equation 3.46), only one parameter was fit to brightness matching data.

Brightness matching is an inherently noisy method of psychophysics, as can be seen from the disordered lines of Figure 3.20. Our ability to draw deeper conclusions about the relationship of the H-K effect with brightness and lightness despite this noisiness is due to our refinement in the method of brightness matching. Hopefully, future experiments will adopt the same constraint of equal chroma differences between stimuli pairs on a chroma-uniform scale, such as the proposed revision to the CIECAM16 chroma scale. Using even smaller chroma differences may lead to further discoveries. In the meantime, our model represents a conservative—in that it does not overfit the noise in the data we collected—yet accurate model of the brightness of chromatic visual stimuli.

The key aspects of the proposed model that differentiate it from other models of the H-K effect also act as evidence for the hypothesis of this dissertation, that higher-level perceptual attributes are best modeled using an accurate color appearance model as a base. Almost all other models of the H-K effect require a hue dependency. However, because we already corrected the hue-dependent eccentricity function (Equation 3.28 in Section 3.1.3), we no longer needed a specific hue dependency in our CIECAM16-based model. The fact that the function made to improve perceptual chroma uniformity also solved this problem in the H-K effect is evidence that the H-K effect is occurring at this higher, cognitive level.

The simplicity of our model also depends on our conclusion that the strength (rate-of-change) of the H-K effect is invariant to chroma, which is in turn reliant upon an accurate scale of chroma. Our proposed formula for chroma (Equation 3.24 in Section 3.1.3) improved

the perceptual linearity of the chroma scale (Figure 3.6). This chroma invariance would only have occurred on such a perceptually linear scale and is further evidence that the H-K effect is a cognitive perceptual phenomenon.

Finally, our logical improvements to the CIECAM16 formulas allow for this H-K model to even exist. Our model would not be invertible without our removal of the explicit lightness dependency in the CIECAM16 formula for chroma (Equation 3.19). Additionally, our newly linearized relationship between lightness and brightness (Equation 3.17) is necessary to generate a model of the H-K effect that can predict both H-K-compensated lightness and H-K-compensated brightness.

3.5 Testing the Helmholtz-Kohlrausch effect model on HDR images

3.5.1 Introduction

In this study, we sought to test whether the model of brightness proposed in Section 3.4 could be applied to predict how observers would perceive modulations in the color of HDR, WCG imagery, simulating how differences in display color gamut and peak luminance may affect ratings of brightness and saturation. Due to spatial effects, color appearance in images is more complex when compared to the simplified geometric stimuli used in the direct brightness matching studies for the H-K effect [9, 89–92]. However, directly applying this model to images serves as a useful first study of how the Helmholtz-Kohlrausch effect functions for images.

3.5.2 Methods

Image Modulation

Nine frames of HDR video were selected from a set of publicly available HDR videos developed by the Video Electronics Standards Association for visual experiments (Figure 3.23). Each frame was mapped to the gamut of the Sony PVM-X3200 reference monitor used in the experiment with the diffuse white point set to D65 at 100 cd/m^2 and the peak luminance of the images clipped to the white point of the display, 790 cd/m^2. The CIE XYZ values of each image pixel were then used as input to the revised CIECAM16 color appearance model. The diffuse white point was used as the reference white point for the model with the degree of adaptation set to 1 and the surround specified as “dark.”

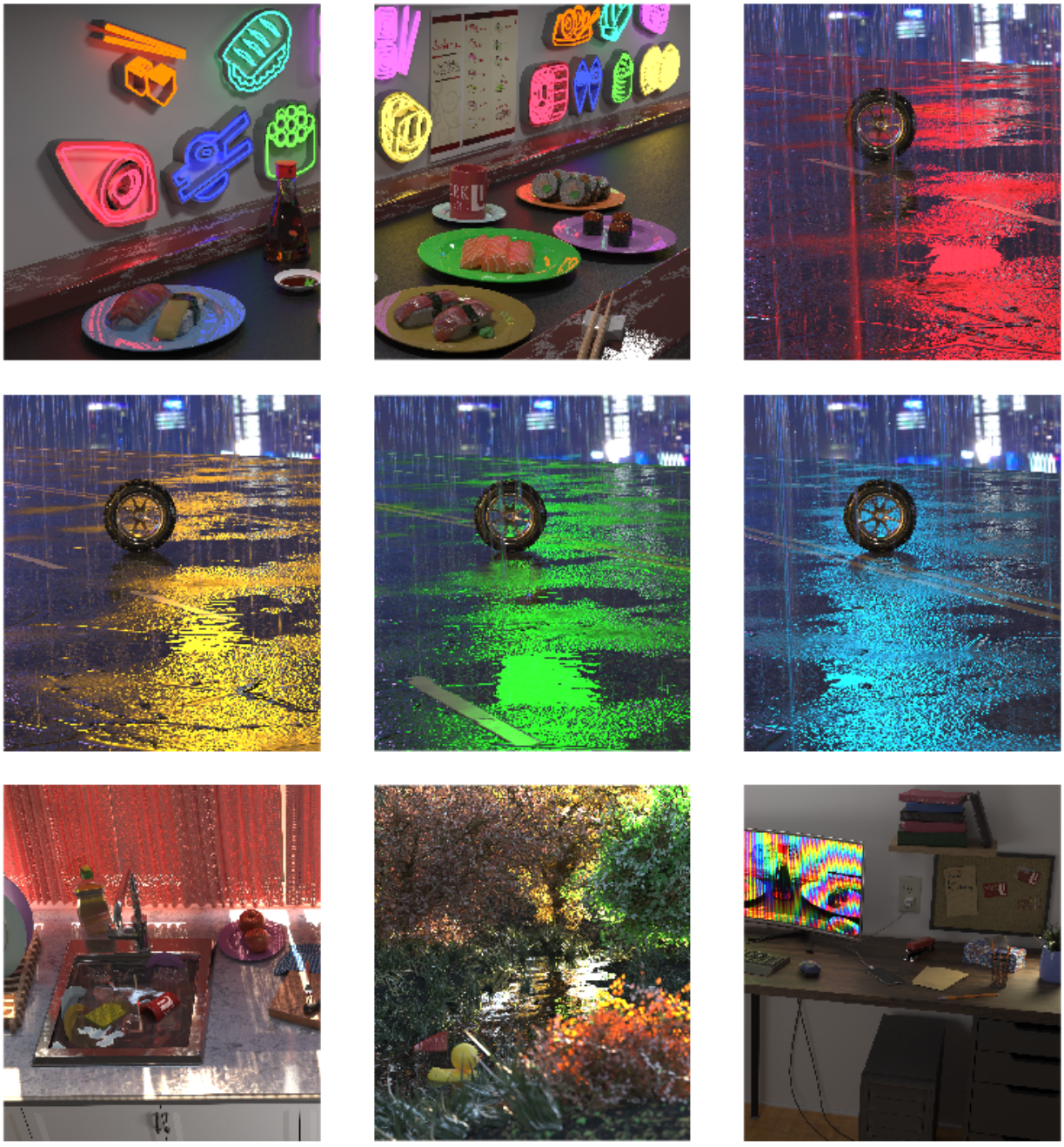


Figure 3.23: SDR approximation of HDR images used in the experiment. Images were used under a CC by 4.0 license from VESA and York University.

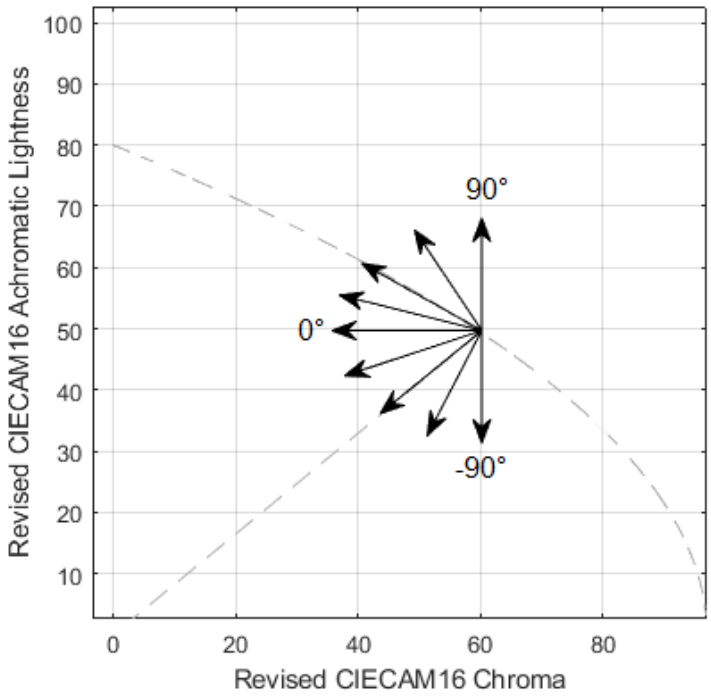


Figure 3.24: Example directions of image pixel color modulation in the achromatic lightness-chroma plane. The upper dashed line represents the equal-perceived-brightness line predicted by our model of the H-K effect (Equation 3.44) and the lower dashed line represents the equal saturation line.

The color of each image pixel was then modulated in one of nine directions in the CIECAM16 lightness-chroma plane to generate nine new images for each start image. The directions of modulation were as follows (ordered counter-clockwise) (Figure 3.24):

1. The positive lightness direction (+90° in Figure 3.24).
2. The angle halfway between directions 1 and 3.
3. The direction of the equal brightness line as predicted by Equation 3.44.
4. The angle halfway between directions 3 and 5.
5. The negative chroma direction (0° in Figure 3.24).
6. The angle halfway between directions 5 and 7.
7. The direction of the equal saturation line (towards the origin of the lightness-chroma plane).
8. The angle halfway between directions 7 and 9.

9. The negative lightness direction ($-90°$ in Figure 3.24).

Given that the derivative of the equal brightness line and the direction of the equal saturation line depend on achromatic lightness and chroma, these directions and the intermediate directions were calculated for each pixel. The magnitude of the modulation was set to either 5% or 10% of the chroma of each pixel, generating 18 modulated images for each reference image. No pixels with chroma below 10 were modulated so as to encourage the viewer to only use the chromatic areas of the image to judge brightness. This expectation was confirmed via post-experiment interviews with the participants.

Experimental Design

The experiment followed a two-alternative forced-choice method where observers were presented with one of the 18 modulated images next to the corresponding reference image (separated by a vertical bar) and asked one of two questions:

- Which image is brighter?
- Which image is more saturated?

This was repeated in random order for all modulated images, all reference images, and both brightness and saturation, leading to 324 total judgments per observer. 18 observers participated in the experiment, 13 of which had previously received formal color science education and 5 of which were naïve. They were instructed in the definition of saturation by demonstration of pages from the DIN color order system.

Stimuli were displayed on a Sony PVM-X3200 reference monitor in a dark room. Colorimetric characterization was performed with a Colorimetry Research CR-100 colorimeter. A $10 \times 10 \times 10$ test grid in Rec. 2020 R′G′B′ space was used to test the color accuracy of the display and optimize a 3×3 transform matrix, which improved the color accuracy of the display from 2.3 ΔE_{00} to 1.3 ΔE_{00}.

3.5.3 Results

General Trends

The direction of image pixel modulation in color space should predict the response of observers. When the angle of modulation is in the positive achromatic lightness dimension ($90°$ in the notation from Figure 3.24), the expectation is that the observers will always perceive the modulated image as brighter and less saturated than the reference image. The converse

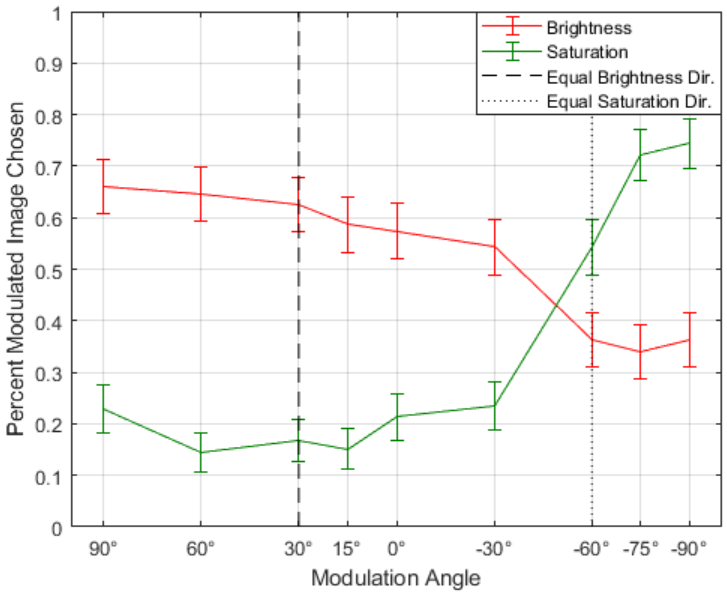


Figure 3.25: Percent of trials in which the modulated image was chosen as brighter or more saturated as a function of modulation angle, averaged across all observers and images.

is expected for the $-90°$ dimension: the modulated image should now appear less bright but more saturated than the reference image. Figure 3.25 shows that these predictions hold when our data is averaged across all observers, images, and modulation intensities.

As the modulation angle decreases in our notation (Figure 3.24), observers should be less likely to rate the modulated image as brighter, as can be seen in Figure 3.25. For modulated images between 90° and 0°, chroma is reduced in the modulated images and achromatic lightness is increased in the modulated images. In our model of H-K-compensated brightness, both chroma and achromatic lightness contribute to perceived brightness. Thus, according to the prediction of our model, reducing the chroma should make the image appear less bright

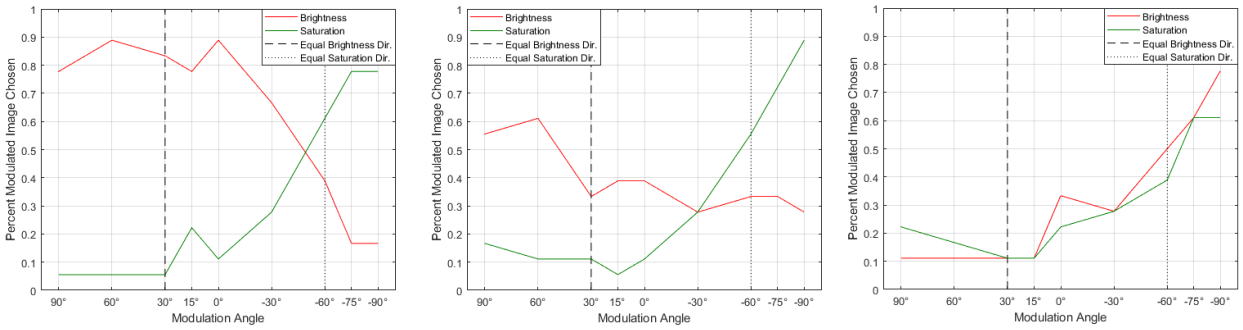


Figure 3.26: Representative examples of brightness curves generated by individual observers in the experiment.

to observers but increasing the achromatic lightness should make the image appear brighter to observers. So, there is an inherent brightness tradeoff in this quadrant. If the achromatic lightness is boosted much more than the chroma is decreased (e.g., modulation angle 2), then we would expect that the small loss in perceived brightness due to decreased chroma to be offset by the increase in achromatic lightness in the modulated image, which would make the modulated image still appear brighter than the reference image. Conversely, if the chroma is decreased much more than the achromatic lightness is boosted (e.g., modulation angle 4), we would expect that the decrease in chroma to cause the modulated image to appear less bright than the reference image, even though there was a small boost to the (physical) achromatic lightness of the image. The point at which the loss in chroma is exactly offset by a gain in achromatic lightness should be when observers were equally likely to choose either the modulated image or the reference image as brighter. Surprisingly, though, when the only modulation was to reduce the chroma (modulation angle of 0°), observers did not perceive the modulated images as less bright! Even when the achromatic lightness was slightly reduced along with the chroma (the first negative modulation angle), observers were equally likely to choose the modulated image as the reference image. Possible explanations for this unexpected result are discussed below.

Observers should be more likely to rate the modulated image as more saturated as the modulation angle decreases in our notation (Figure 3.24). Saturation is typically defined as chroma relative to lightness, so decreasing the chroma or increasing the lightness should cause the perceived saturation to decrease. Thus, for the positive angles of modulation where both chroma is decreased and lightness is increased, the perception of the observers that the reference image was more saturated agrees with our expectations.

In the case of negative angles of modulation, the chroma is still being decreased, which would cause saturation to decrease, but the lightness is now also being decreased, which would cause saturation to increase. The modulation angle at which the decrease in lightness offsets the decrease in chroma and observers are equally likely to rate either image as more saturated should fall along the line of equal saturation which points from the stimulus to the origin of the lightness-chroma plane. In our results, the average percentage of saturation ratings did cross the 50% point at the angle that pointed to the origin, in agreement with the expectations outlined above.

Differences between observers

The 18 observers who participated in the experiment were consistent in their judgments of saturation, with all but two following the trend from the average results. However, there was

significantly more variation between observers in their ratings of brightness. The majority (12) of observers' results followed the average trend from Figure 3.25. However, several observers had results curves that were relatively flat or lacked a clear trend relative to modulation angle (Figure 3.26). Additionally, two observers appeared to conflate brightness and saturation, with their brightness results following the same trend as saturation (Figure 3.26). Interestingly, all five of the naïve observers fell into the first category of observer; the observers that conflated the meaning of brightness and saturation had all received formal color science training. These severe differences between observers that we observed are a cautionary tale for researchers of brightness. Given the fluid linguistic meaning of the term "brightness," we cannot assume that the interpretation of brightness by observers exactly aligns with the technical color science definition [25].

3.5.4 Discussion

The perceptual effect on saturation of our pixel-by-pixel modulation of images was well predicted by the direction of modulation in the revised CIECAM16 achromatic lightness-chroma plane. This indicates that our color appearance modeling pipeline was accurate in predicting the perceptual color attributes of image content and that observers were consistent in their ratings of saturation.

However, our results for observers' ratings of brightness were more mixed. Observers had differing interpretations of brightness or differing implementations in how they made brightness comparisons (Figure 3.26). Thus, brightness judgments may be less reliable in image psychophysics than saturation judgments.

Furthermore, the overall trend in observers' brightness comparisons did not follow the expected transition point predicted by the Helmholtz-Kohlrausch effect as the direction of image pixel color modulation in the lightness-chroma plane changed. This failure of our color appearance modeling pipeline to predict brightness judgments could be partially due to our exclusion of spatial effects in our calculations. For example, in many images, the background of the image was not modulated because its chroma fell below our chroma cutoff. So, the interaction between the constant background and the shifting image content could have created spatial brightness effects for which we did not account.

Spatial effects alone, though, would not be able to account for the most surprising result, which is that when images were modulated to only reduce chroma and keep achromatic lightness constant, observers on average rated the modulated image as brighter than the reference image. Even if the Helmholtz-Kohlrausch effect did not exist, then observers should still have only been equally likely to rate either image as brighter, but with the well-established validity

of the Helmholtz-Kohlrausch effect (more colorful stimuli are brighter at equal luminance), this result is completely unexpected.

A failure of the achromatic lightness dimension in CIECAM16 to be completely isoluminant is one possible explanation for why the less colorful image was perceived as brighter by observers, for in this case changing the chroma could have also led to an increase in luminance that was not accounted for. Additionally, the image content could have played a role in this surprising result. In post-experiment interviews, observers focused on the brightest and most colorful elements in the images when making their judgments, which in six of the nine images were colored lights. It is possible that decreasing the chroma of the colored lights made them look whiter, which observers may have interpreted as brightness in the context of lighting. Perhaps the use of a color appearance model tailor-made for self-luminous stimuli, such as CAM18sl [64], may be more appropriate for this type of image content than CIECAM16.

The color modulations (Figure 3.24) performed in this study could approximate the differences between displays with different peak luminance or color gamut. As expected, increasing the simulated color gamut of an image by increasing its chroma led the observers to view the image as more saturated. While they did not necessarily view such images as brighter, another recent study indicated that wider color gamut and more saturated colors are more important to the brightness and vividness of HDR images than peak luminance [93]. Thus, further exploration of this topic is necessary to address the potential shortcomings in our modeling described above, including the incorporation of spatial effects on the perception of HDR content.

Brightness is not typically referred to as a cognitive color appearance phenomenon. Both the success of our cognitive approach to modeling the H-K effect and the failure of our modeling to directly translate to image brightness are evidence that brightness perception ought to be treated as a cognitive mechanism. We have also shown how accurate color appearance models should form the basis for models of such cognitive perceptual attributes and can provide more useful data when incorporated into experimental design as well.

Chapter 4

Chromatic Adaptation

This chapter investigates cognitive mechanisms of chromatic adaptation, specifically cognitive discounting of the illuminant. Heterochromatic viewing environments, where an observer is presented with two colors of illumination in different parts of the scene, form the basis for our experimental methods and are introduced in Section 4.1. Our first experiment, covered in Section 4.2, measured the sensory adaptation to the heterochromatic lighting environment to establish a baseline for the second experiment. The same heterochromatic viewing condition is used for hue matching in Section 4.3 as we seek to quantify the degree to which observers cognitively discount the illuminant.

4.1 Background

The inherent challenge of studying chromatic adaptation is that we seek to compare stimuli that cannot coexist at the same time and place. Studies of chromatic adaptation then involve either temporal or spatial separation of illumination conditions. In these experiments, we explore simultaneously-presented, spatially-separated heterochromatic viewing environments. Such environments have been implemented across many types of visual experiments.

One application of heterochromatic viewing conditions is to generate data sets of corresponding colors—stimuli which have the same color appearance under different colors of illumination/states of adaptation. Such experiments [73, 94] form the experimental foundation for modern chromatic adaptation transforms and were central to the development of such transforms in the 1980s and 1990s. Importantly, these experiments implemented haploscopic viewing conditions, where each eye is presented with a separated field with its own color of illumination. Each eye then independently undergoes chromatic adaptation to the color of the illumination and the resulting corresponding colors provide useful informa-

tion specifically for models of sensory adaptation, such as that implemented by CIECAM16 (Equation 2.2).

If we are instead to explore cognitive mechanisms of discounting the illumination color, we can allow both eyes to view both scenes/areas of illumination. Such viewing environments were used in the 1990s in color science studies of cross-media image comparison where the white points of the hard copies and soft copies varied [95, 96]. These studies referred to such viewing conditions as "simultaneous-binocular viewing" to differentiate them from the haplosopic conditions discussed above. Similar viewing environments, where two areas with different color of illumination were presented to observers, were used in the same era by psychologists studying "color constancy" and were referred to as "simultaneous asymmetic matching" [23, 97]. While similar in experimental design, these studies implemented a different theoretical approach to the question of how the visual system deals with the color of the illumination. The color scientists tended to think about experiments in terms of finding appearance matches, whereas the "color constancy" researchers used the paradigm of the visual system trying to recover surface properties of objects across viewing conditions [98]. However, such a surface-oriented paradigm often fell prey to the fallacy that there is one "correct" color that an object should appear under a second illumination given its color under a first illumination, a mindset which ignores the phenomenon of metamerism [20].

While binocular viewing conditions offer more natural viewing environments, it is more difficult to define the exact method of adaptation compared to the haploscopic condition, as discussed by both color scientists [95] and psychologists [97]. Our two-part approach to studying chromatic adaptation in heterochromatic environments allows us to resolve some of the uncertainty of the location of chromatic adaptation between sensory and cognitive sites. By first employing a experimental task and stimulus arrangement which forces observers to make achromatic matches based on sensory adaptation, we can separate out the mechanisms of adaptation when we analyze the results of the more natural viewing environment that allows for both sensory and cognitive adaptation. Revisiting this type of experimental setup a quarter century later also provides us the benefit of the a quarter century of maturing of color appearance models, where we have transitioned from a plethora of color appearance models and chromatic adaptation transforms to a mature, CIE-recommended model [1]. (Although research does continue in the development of and improvements to color appearance models [64, 83, 99].)

The key consideration when observers make appearance matches across viewing conditions is the criteria used by the observer to determine a match. The common conceptualization of match criteria consists of two types of matches: light-like matches and surface-like matches [100]. For light-like matches, one is essentially asking an observer to use their visual

system as a colorimeter and ignore cognitive effects as much as possible. This type of match is also referred to as an appearance match and is often cued as such to observers. If instead one asks an observer a question like "Find the two objects that appear to be made from the same paint," the experimenter would expect for the observer to make a surface-like match by cognitively discounting the color of the illuminant.

Certain experimental setups tend to lend themselves to one type of match or the other. Making color matches across two displays in a dark room is going to lead to light-like matches due to the lack of scene cues. Asking an observer to find the piece of paper from which a provided sample was cut is going to lead to surface-like matches. However, in many cases, such as in Experiment 2 (Section 4.3), the type of match made by the observer is not so clear. In an early pilot of this experiment, I specifically instructed observers to make surface-like matches, but this instruction proved difficult for observers to implement in practice. Furthermore, when I tested out the final version of the experiment, I specifically tried to make light-like matches, but found that such matches actually indicated I was strongly cognitively discounting the illuminant on both sides of the light booth. Thus, while experimenters may wish to divide sensory and cognitive mechanisms into, respectively, what we "see" and what we "think," this anecdote indicates that cognitive mechanisms are indivisible at some level from our perception of pure appearance.

4.2 Experiment 1: Sensory adaptation to heterochromatic illumination

4.2.1 Methods

A large bipartite light booth was used to generate heterochromatic lighting environments within which the chromatic adaptation of observers would be measured (Figure 4.1). Each of the two adjacent cells of the light booth were approximately 23 inches wide, 23.5 inches deep, and 24 inches tall. The walls of the booth were painted with neutral gray paint with approximate Munsell Value 8 ($Y_w = 59$). Seven-channel ETC Source Four LED theater lights installed above each light booth cell provided the lighting. The spectral power distributions of the seven LED channels are shown in Figure 4.2. Diffusive panels installed below the lights helped the uniformity of the illumination.

Three colors of light were used as adapting stimuli: CIE Illuminant A, CIE Illuminant D65, and 12000 K daylight calculated using the CIE method. Since there are many combinations of the seven LED channels that can be used to match the color (CIE XYZ values) of

Figure 4.1: Bipartite light booth used in chromatic adaptation experiments.

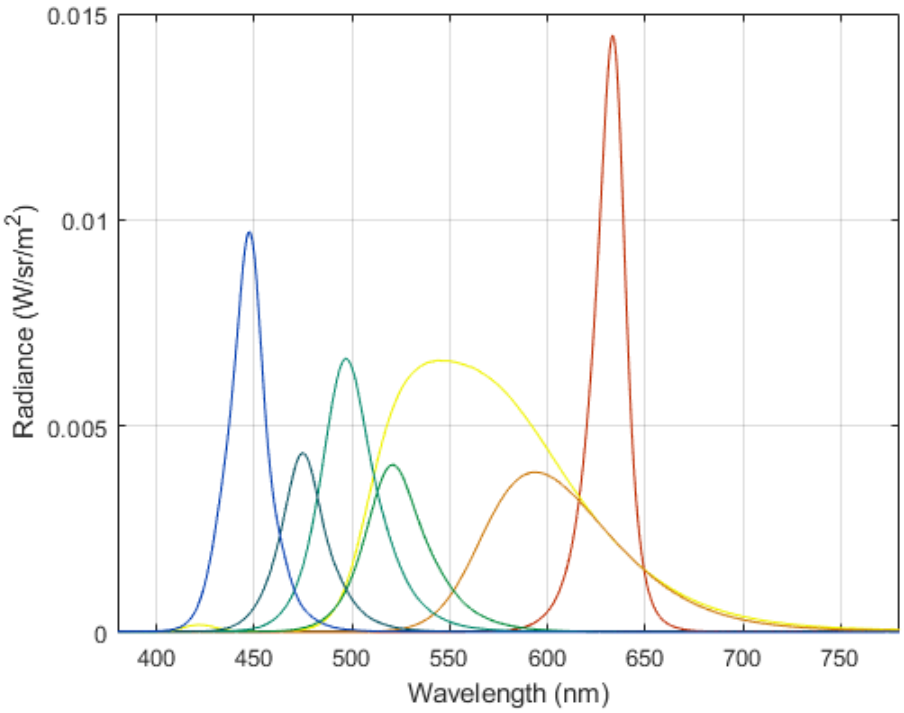


Figure 4.2: Spectral power distributions of the 7-channel LEDs used to illuminate the light booth.

each of these standard illuminants, the LED channel intensities for each color were optimized to minimize the spectral difference between the light booth spectral power distribution and the spectra of the standard illuminants. Spectral differences were minimized in order to reduce the effect of inter-observer differences in spectral sensitivity. Each pairwise combination (three) of these three illuminations were used as adapting stimuli. Three uniform illumination settings, where both sides of the light booth were set to the same color, were also used, bringing the total number of adapting environments to six.

A psychometric staircase method was used to assess each observer's neutral point when adapted to each heterochromatic or uniform lighting environment. At the beginning of each staircase, each observer was adapted to the lighting for two minutes. This length of time was determined to achieve a sufficient degree of adaptation based on the lack of drift in observer results as they completed the staircase and continued to adapt to the adapting environment. Observers were visually cued at short, random intervals to direct their gaze to one side of the light booth or the other in order to ensure that they spent equal time looking at both sides of the light booth. The direction of gaze was randomized by the MATLAB program used to run the experiment, but the exact location of their gaze within each light booth was directed towards the center of each section's back wall. This specific method of adaptation and gaze control was chosen to mimic the likely behavior of observers in Experiment 2, where they would be comparing the color of objects across the two halves of the light booth.

After the adaptation period, each trial stimulus presentation followed the following flow (Figure 4.3):

1. Re-adaptation for 10 seconds following the method described above.

2. A brief, 1 second period where all lights turned off, to indicate to the observer that the trial stimulus was about to be presented.

3. The trial stimulus presentation: a single color of light across both halves of the light booth presented for 0.1 seconds.

4. All the lights off again while the observer enter their judgment as to whether the trial stimulus appears more yellowish or bluish.

The set of available trial stimuli consisted of 121 illuminations along the Planckian and daylight loci. The stimuli were evenly spaced in the CIE $u'v'$ chromaticity diagram. The method recommended by the IES TM-30 standard was used to blend the two loci, with the weighted averaging of the corresponding Planckian and daylight spectra from 4000 K to 5000 K.

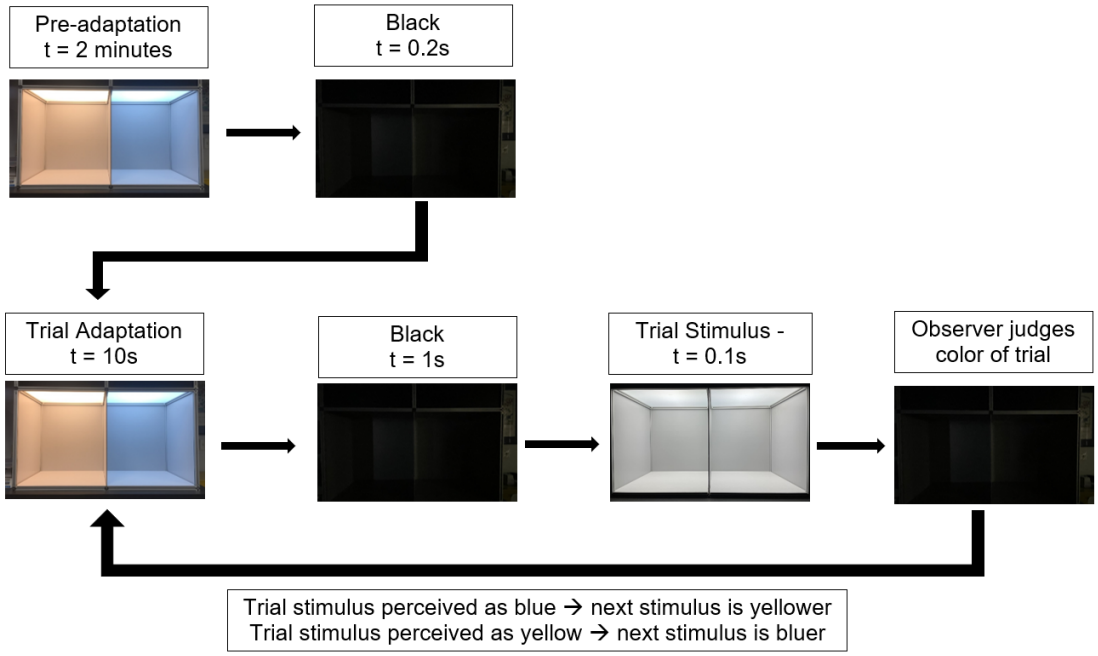


Figure 4.3: Block diagram demonstrating experimental method for Experiment 1.

For each stimulus presentation, the observers judged whether the stimulus appeared more yellowish or bluish. If they responded bluish, the next stimulus would increment down in correlated color temperature (in the yellow direction) and vice versa if they responded yellowish. The turning point, when an observer switches from responding bluish to responding yellowish, is a measure of their reference white appearance, and taking the mean of many turning points provides a measure of their state of chromatic adaptation. It is important to note that this experimental design and set-up specifically measured observers' sensory adaptation without any discounting of the illuminant. This was achieved by removing any objects from the light booth and simply having them judge the color of the illumination. Since they are judging the pure stimulation of the light, there is no additional illumination to discount.

Each staircase began randomly at one of seven starting stimuli that were evenly spaced among the 121 trial stimuli. Three stages of progressively smaller increments (20 steps, 5 steps, and 1 step) through the 121 stimuli were used in order to more quickly hone in on the white-appearing stimuli for each adapting stimulus. The staircase moved from the first stage to the second stage after two turning points and from the second stage to the third stage after three turning points. The first stimulus of the next stage was the median of the turning points from the previous stage. These criteria were chosen experientially in order to minimize the number of trials: requiring more turning points increases the number of trials

required while having too few turning points would lead to the next stage starting farther away from the observer's white point and require additional trials at a smaller increment to get back close to the white point. The lack of drift in median turning point between the different stages across most trials throughout the experiment indicated that these criteria were well chosen.

The third, final stage with the finest increment was stopped after four turning points, which would then be averaged to calculate the white point of the observers. There was concern that observers might detect that their responses directly controlled the next stimulus (responding blue causes the next stimulus to be yellower, and vice versa). So, two simultaneous staircases with different starting positions were run with their stimulus presentations interleaved so as to mask the logic of the experiment, resulting in eight total stage-three turning points.

There was also concern that the small increment size during stage three would lead to observer frustration or distraction because sequential stimuli may appear very similar. So, during stage three, there was a one-in-eight chance of being shown a randomly chosen "dummy" stimulus.

Nine observers (ages 18-35, median 19) participated in the experiment. The observers did not have previous experience with psychophysics but received a training session before participating in the experiment.

4.2.2 Results

The results of the experiment are shown in Figure 4.4. The CCT of the adapting illuminant or illuminants is indicated by the vertical black boxes. The mean of the final stage turning points for each observer is indicated by a vertical black line on each row and 95% confidence intervals for those values are indicated by the green boxes. Confidence intervals were estimated using a Monte Carlo simulation in which each staircase was simulated 1000 times using the observed percentages for each trial stimulus for that specific observer and adapting stimuli.

Alignment of the confidence intervals with the adapting illumination in Figure 4.4 indicates complete chromatic adaptation. For instance, most observers experienced complete chromatic adaptation to D65 (left column, center row) but only partial chromatic adaptation to Illuminant A (left column, top row) and 12,000 K daylight (left column, bottom row). Each observer's results from the single-illuminant condition were used to predict their results in the heterochromatic illumination condition (brown vertical lines in the right column). These predictions were generated by taking the average of the positions of the two white points along the Planckian-to-daylight IES TM-30 reference illuminant curve in the

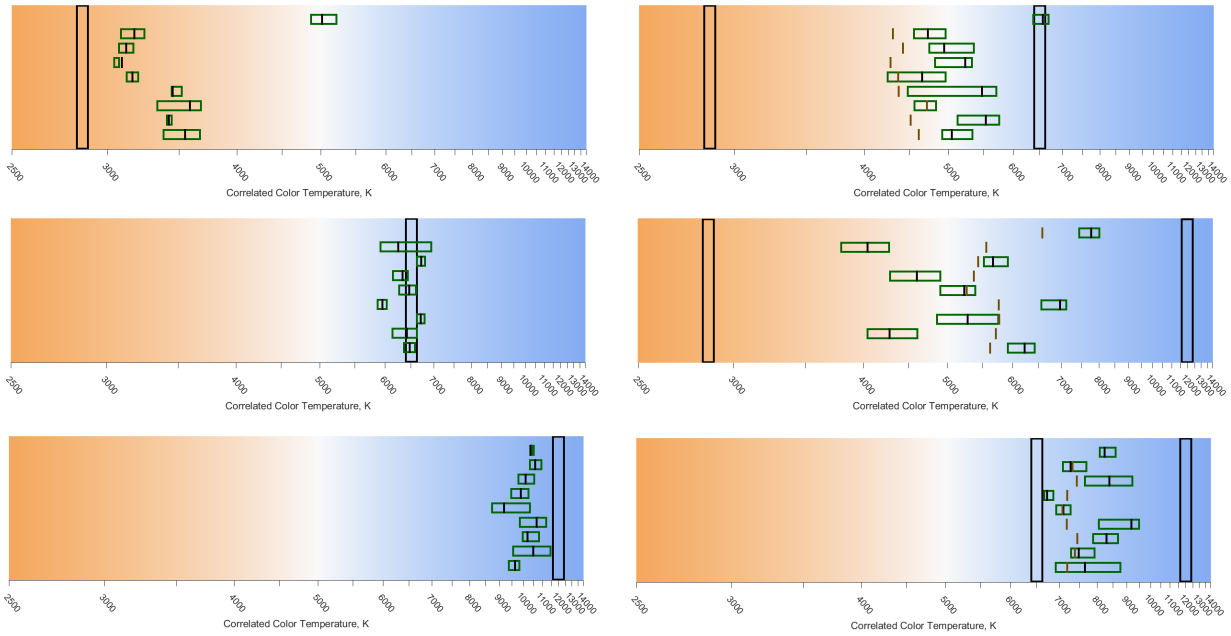


Figure 4.4: Results of Experiment 1. Each sub-figure represents one illumination condition, with the correlated color temperature of the illuminant or illuminants indicated by the vertical black boxes. Within a set, each of the vertical black lines indicates the experimentally determined neutral point of each observer. The green boxes indicate the 95% confidence intervals for each result. The brown vertical lines in the right column indicate the neutral point predicted from the average of the results from relevant single illumination adaptation points in the left column.

	Ill. A	D65	12k K
Ill. A	3450	5179	5402
D65	-	6407	7752
12k K	-	-	10109

Table 4.1: Mean neutral point across observers for each combination of illumination colors.

$u'v'$ chromaticity diagram.

Interestingly, the measured adaptation point for observers in the heterochromatic conditions do not match these values predicted by averaging the single illuminant results. For the D65 and Illuminant A heterochromatic condition (top right in Figure 4.4), the observers' chromatic adaptation was biased towards D65 compared to the average of their adaptation points to the individual illuminants. A similar towards the bluer illuminant can be observed in the results of the D65 versus 12,000 K heterochromatic condition (bottom right). The vastly different colors of the illuminations in the Illuminant A versus 12,000 K condition appears to have led to a wide dispersion of adaptation points to that lighting compared to the relatively narrow dispersion of results from average their adaptation to individual illuminants (middle right). This suggests that judging the achromatic point was more difficult for observers in this condition. Nevertheless, it is useful to record observers' mean white point in these conditions, which are reported in Table 4.1.

4.3 Experiment 2: Hue matching across heterochromatic lighting

4.3.1 Methods

Experimental Design

Experiment 2 sought to expand upon the preliminary results of Experiment 1 by introducing objects into the heterochromatic lighting environments to assess how observers compensated for the color of the illumination beyond their sensory chromatic adaptation in judging the appearance of chromatic stimuli. The first step in the process was creating the objects. 4" hard foam cubes were used as the substrate to which heavy-body acrylic paints would be applied. Roy Berns had previously generated a data set of absorption and scattering spectra for 19 Golden Heavy-Body Acrylics [101]. Two-constant Kubelka-Munk theory was used to predict the reflectance spectra of mixtures of these paints from their scattering and

absorption spectra [102].

Five color centers were targeted in this experiment with CIECAM16 lightness 50, chroma 20, and five evenly spaced hue angles from 18° to 306° in steps of 72°. These hue angles roughly align with red, yellow, green, blue, and purple. The MATLAB Optimization Toolbox was used to find all possible mixtures of three paints plus white which would match the target appearance attributes under the light booth's spectral simulation of D65. Three paints plus white are sufficient to match any target color within their gamut (three degrees of freedom for three dimensions of CIE XYZ values). The more accurate, physiologically-based CIE 2012 10° standard observer was used for all colorimetric calculations in order to improve the quality of the metamers generated by this method. CIECAM16 calculations were performed using revisions proposed by Hellwig and Fairchild [83]. Additionally, all CIECAM16 calculations used a degree of adaptation factor of 1, assumed "average" viewing conditions, and set the relative background reflectance factor to 59.10, the adapting luminance to $40\,\mathrm{cd/m^2}$, and the white luminance to $200\,\mathrm{cd/m^2}$.

For each color center, the set of all matching mixtures under D65 was then simulated as being illuminated by the light booth's spectral match of Illuminant A. It was observed that these metamers under Illuminant A appeared to have one primary degree of variation in the CIECAM16 *a*-*b* plane. Principle component analysis was performed to quantify this variation. From each set of metamers, three specific mixtures were chosen: the two most extreme metamers along the principal component of variation and a third metamer closest to the middle of the variance (Figure 4.5).

A scale with 0.01-gram precision was used to measure out the concentration by weight of each paint for each paint mixture. 20% concentration of Golden acrylic matte medium was added to each mixture, and then the mixture was applied in two layers to the foam cubes. Due to the approximations inherent in Kubelka-Munk theory and the limits of the accuracy of both the absorption and scattering database and the manual process of measuring and mixing the paint, the resulting paint mixtures were not perfect color matches under D65, with a median ΔE_{2000} of 2.0. If need be, though, this color error could easily have been corrected by adjusting the spectrum of the D65 simulation generated by the 7-channel LED lighting to illuminate each cube.

The spectral differences in the painted cubes along with the spectral tunability of the 7-channel LEDs provided the opportunity to present cubes with a broad variety of apparent color without changing the color of the illumination. A significant shortcoming in previous studies of color "constancy" is that one-to-one mappings of object colors between illumination colors were assumed either by the experimenters or implicitly by the lack of spectral tunability and variablity in the physical media used to present physical stimuli.

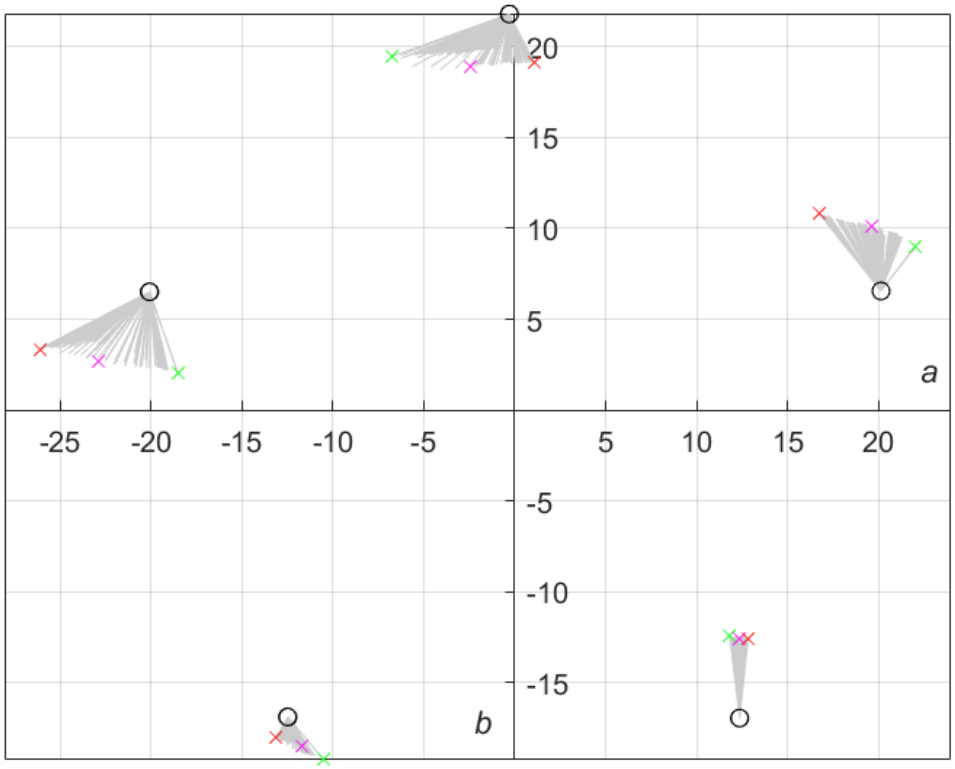


Figure 4.5: Gray lines represent the change in color of each metamer between its color under D65 (circles, the same for each metamer) and its color under Illuminant A in the revised CIECAM16 a-b plane. x's indicate the chosen metamers for painting.

Our spectrally-flexible methods overcome these shortcomings and assumptions, allowing us to control the color of each box via the illumination but independently of the color of the illumination.

We considered two experimental modalities to test how observers account for the color of the illumination in the color appearance of objects. The first modality would have been an adjustment-based method. In this experimental paradigm, observers would have been shown a reference box under one illumination (in practice, D65-colored light), and then would have used a keyboard to adjust the color of a similar box under a second illumination (in practice, Illuminant-A-colored light). Their hue- and chroma-based adjustment would have been translated by an algorithm into adjustments to the spectrum of light that would adjust the color rendering to achieve the target color appearance for that box. The color of the box adjusted by the observer and the reference box color would then form a pair of corresponding colors.

The second candidate experimental modality was a pass/fail acceptability judgment. In this paradigm, observers would simply be shown the reference box color under a D65-colored light and one of many test box colors and asked to judge whether the two boxes match in color. This methodology is more inclusive of the uncertainty of observers' judgments of colors across illumination colors—there is not one single color that observers will perceive as matching (as assumed by the first experimental paradigm), but rather a range of colors that will appear acceptably similar [103]. The color pairs that observers are most likely to approve as matching can be considered corresponding color pairs.

The second methodology was chosen for this experiment to simplify both the experimental design and the task for observers. The experimental task was further simplified and improved by asking observers to simply judge whether the two boxes had the same hue. Removing chroma from the judgment criteria allowed us to spend more time testing hue, to which observers tend to have more stringent criteria. Lightness was never considered because adjusting the spectrum of the illumination had little effect on lightness. Furthermore, object-pair metamers infrequently differ greatly in lightness [104].

Nonlinear optimization in MATLAB was then used to find the range of hues that each box could be made to appear under optimized metamers of Illuminant A via the 7-channel LED light source. For each of the five color centers, then, the one box of the three options that could generate the widest range of hues was chosen for use in our experiment. Then, to generate the exact stimuli for the experiment, nonlinear optimization was used to find LED channel intensities to make an Illuminant A metamer that would cause the boxes to match a given set of CIECAM16 hue angles and revised chroma [83]. The target hue angles ranged from the lowest to greatest hue angle that could be generated by this cube and light source

		Color Center			
		Red	Yellow	Green	Blue
Ill. A	h_{min}	11°	46°	160°	220°
	h_{max}	35°	108°	192°	262°
	Increment	0.5°	1°	0.5°	1°
D65	h_1	18°	66°	146°	229°
	h_2	24°	78°	152°	241°

Table 4.2: Hue angle information for the experimental stimuli used in Experiment 2

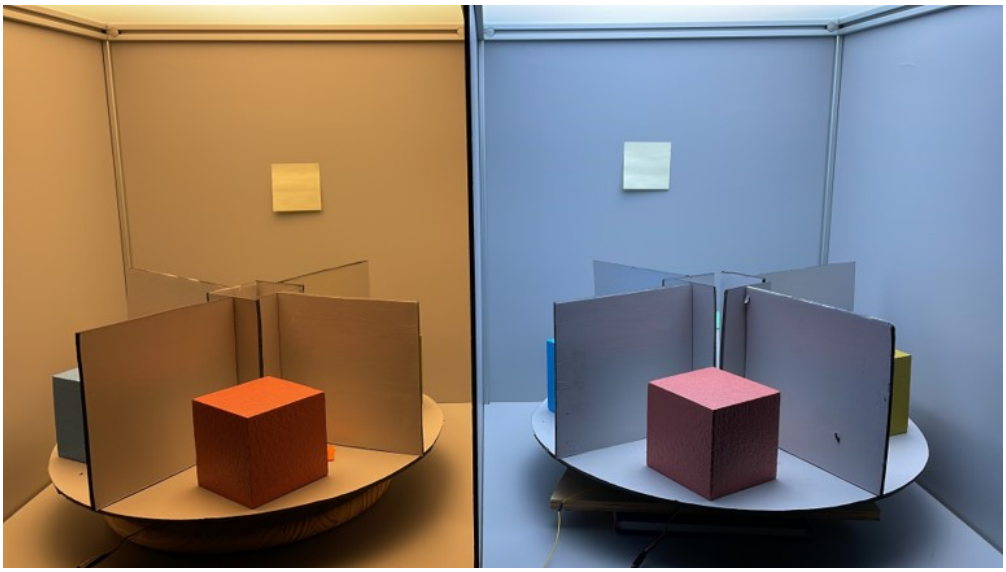


Figure 4.6: Experimental setup.

combination in increments of either 0.5° for the red, green, and purple color centers and 1° for the yellow and blue color centers due to the larger range of hue angles that could be covered by these two color centers. The target chroma was 20, but the chroma was allowed to fluctuate if needed to match the required hue angle. Additionally, for each color center, two sets of LED channel intensities were optimized to find D65-metamer spectral power distributions that would produce two target hue angles for each cube. These hue angles were separated by twelve increments.

At this point, it was discovered that the purple boxes under Illuminant-A-colored light never appeared to have equal hue to the same boxes under D65-colored light, regardless of spectral power distribution. Thus, the purple color center was eliminated from this experiment. Table 4.2 contains the details of the experimental hue angles for the other color centers.

A turntable was built to display the cubes within the light booth (Figure 4.6). Servo motors with positional feedback and a pair of Arduino microcontrollers were used to control the position of the turntables with MATLAB. All surfaces of the turntable were painted with Golden Neutral 8 paint to match the color of the light booth, and vertical barriers obstructed observers' view of the other cubes when they were not being shown.

Experimental Procedure

For each color center, the boxes on either side of the light booth were compared by the observers under each combination of the many Illuminant-A-metamer test lighting settings and the two D65-metamer reference light settings. As discussed above, the colors of the Illuminant A metamers were all identical but caused the boxes to appear a different hue. All stimuli were shown in random order with left-right randomization of Illuminant A and D65. Additionally, it was ensured that adjacent trials never tested the same box. Observers were asked to judge whether the two boxes matched in hue. Before participating, observers were educated on the definition of hue and were shown Munsell constant-hue pages. Instructions to the observers tried to get them to still make light-like matches following the matching paradigm introduced in Section 4.1. This education was important because the boxes sometimes differed in chroma, especially for the blue color center. 17 color normal observers—ranging in age from 20 to 59 with a median age of 25—participated in the experiment.

4.3.2 Results

Figure 4.7 shows the percentage of the 17 observers who responded that the two boxes matched in hue as a function of the CIECAM16 hue angle of the box under each Illuminant A metamer, using Illuminant A as the white point for the CIECAM16 calculations (unity degree of adaptation). Vertical lines of the same color as these lines indicate the CIECAM16 hue angles of the box under each of the two D65 metamers, using D65 as the white point for the CIECAM16 calculations (unity degree of adaptation).

The expected shape of the curves are demonstrated by the results for the yellow boxes (top right). At the hue angle where the Illuminant A metamer causes the box to appear the same hue as the reference D65 box, nearly all observers responded that the two cubes matched in hue (value closest to one). Then, as the hue angle deviates from the best match, the frequency of "match" responses from observer decreases until it approaches zero. These characteristic curves can be approximated as normal probability distribution functions (Figure 4.8), with the abscissas of the peaks of the curves representing the box under Illuminant A that best

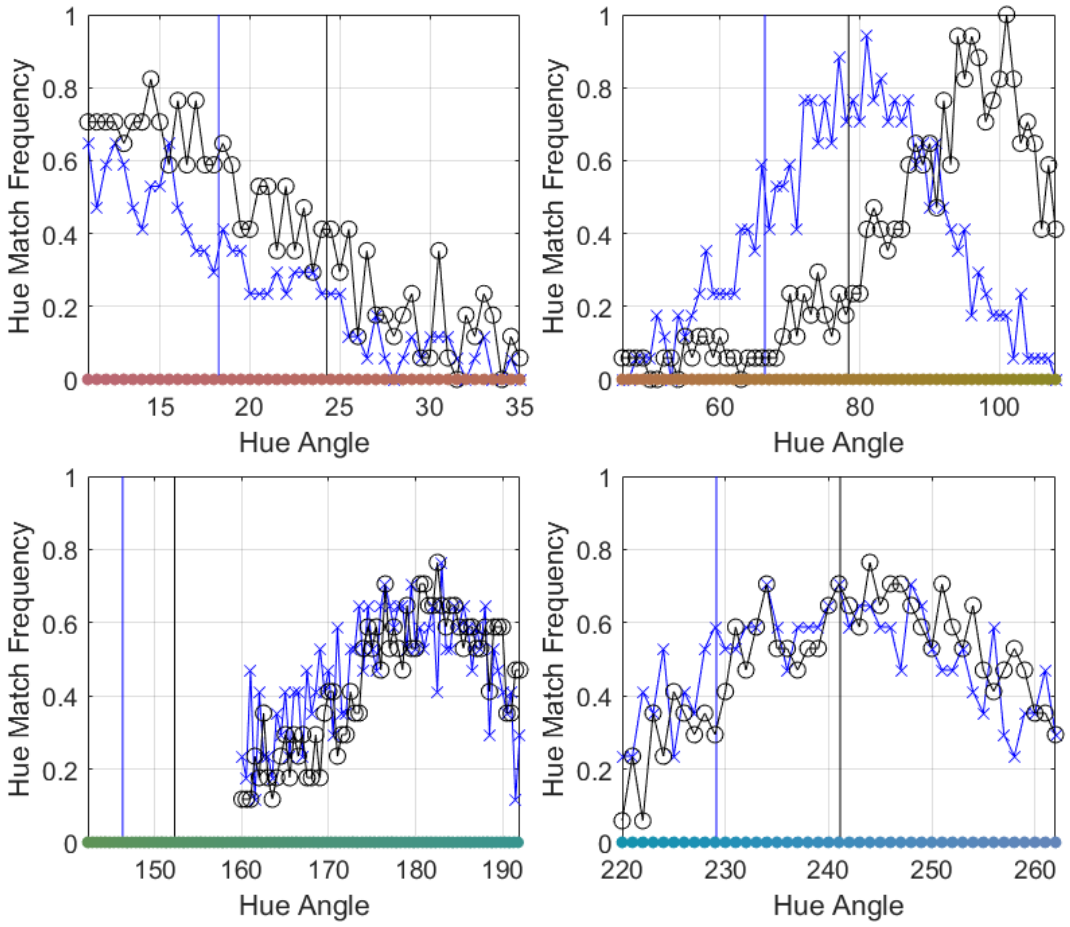


Figure 4.7: The percent of responses that indicated a match in hue between boxes as a function of the CIECAM16 hue angle for each of the Illuminant A boxes divided into four color centers. Vertical lines indicate the CIECAM16 hue angle of the corresponding reference boxes under D65, assuming full adaptation to D65.

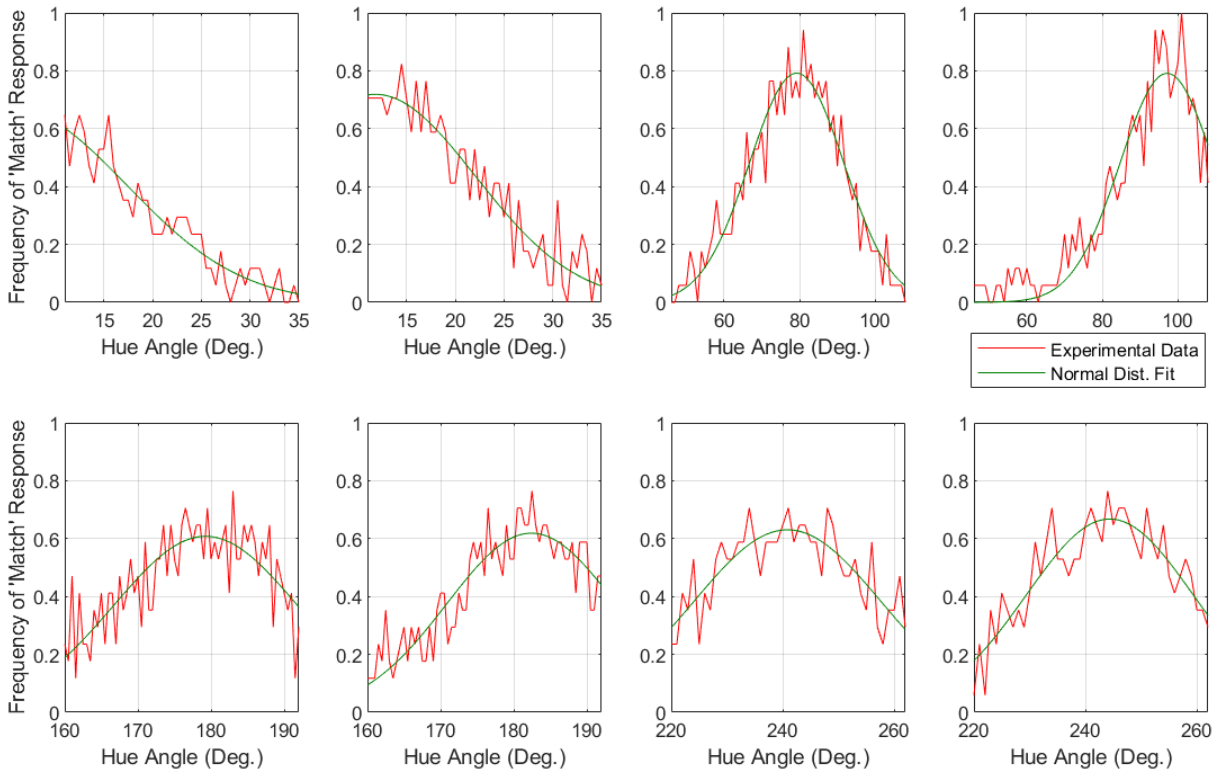


Figure 4.8: Normal probability distributions fit to each curve from Figure 4.7, representing the percent of responses indicating a hue match between boxes.

matches the corresponding reference box hue under D65.

For the other color centers, the hue match frequency never reached one; all observers never agreed that a single pair of cubes matched in hue. This result can be further explained by examining the individual results for each observer (Figure 4.9) For the yellow color center (top right of each set), the observers had similar tolerances for hue matching and their individual ranges of acceptability had similar centers, leading to the normal shape of the yellow results. For the green color center (bottom left of each set), there were discrepancies amongst observers both in relation to the range of tolerance and which hue was the best match. Some observers only saw a very narrow range of bluish cubes under Illuminant A as matching the D65 reference cube hue, whereas others had a very broad range of hues that they viewed as an acceptable match, and these hues tended towards the greener part of the displayed hues compared to the first set of observers.

A similar distinction between the tolerance range of observers can be seen in the results for the red color center (top left of both sets of Figure 4.9). However, these results are further complicated by the fact that the D65 reference cube hues were too purple (low hue angle) compared to the box hues that could be generated by the Illuminant A metamers. Thus, the

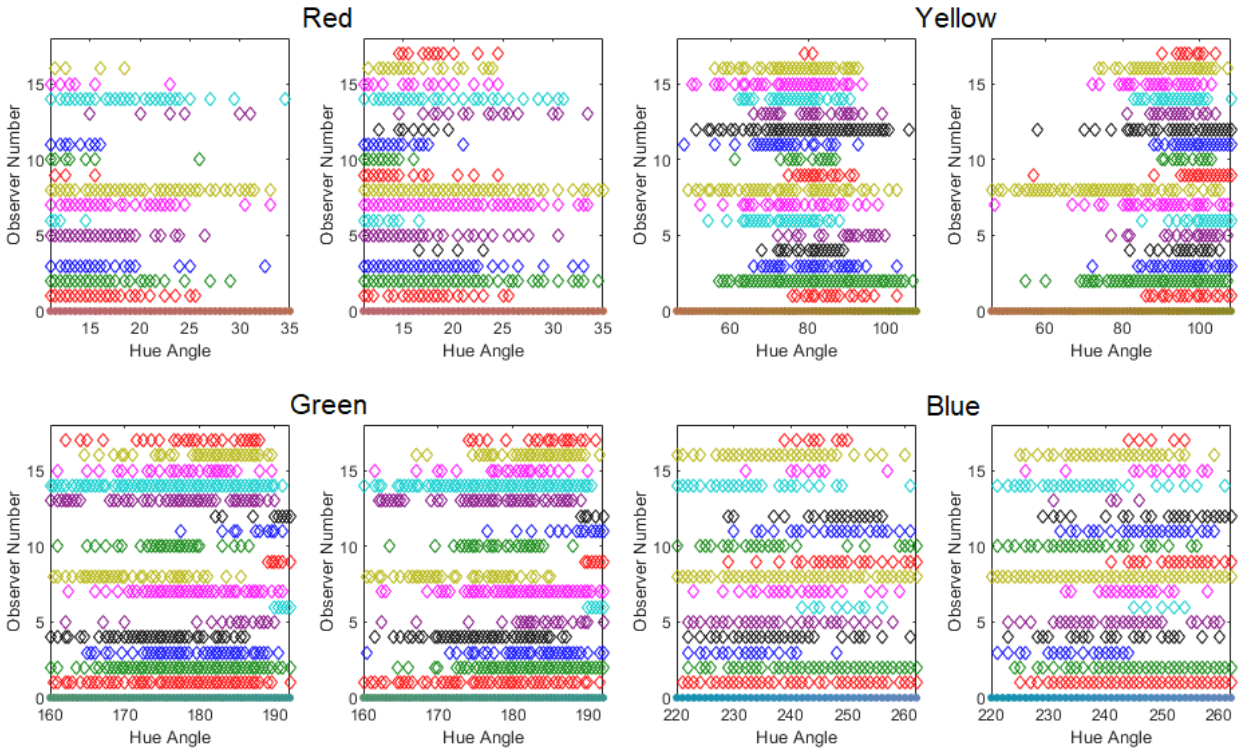


Figure 4.9: Each diamond represents a "hue match" response for a specified stimulus for each of the 17 observers that participated in the experiment, whereas an empty space indicates the observer did not observe a hue match.

probability distributions for the red color center may never have reached their peak (Figures 4.7 and 4.8). Thus, we may not have been able to determine the correct hue match for that color center.

The observer variation in tolerances was even more severe for the blue color center (bottom right of both sets of Figure 4.9), with some observers reporting that all of the box hues under Illuminant A matched the D65 reference box hue and other observers reporting very few matches, if any. This lack of discernment for some observers flattens the resulting probability distributions (Figure 4.8). These results suggest that some observers may have reverted to judging color categories as opposed to judging the precise hue composition of each box, despite instruction provided to the observers to specifically exclude any pairs of boxes where one box appeared slightly bluer, or slightly greener, or slightly yellow, or slightly redder than the other box. In future experiments, it may be beneficial to train observers to discern hue more precisely using monochromatic illumination and perhaps even exclude observers that fail to sufficiently discern hue differences. These limitations of the data described above are important to keep in mind as we analyze the results of the normal fitting.

The color of the best hue match of a box under an Illuminant A metamer against each reference D65-illuminated box is shown in Figures 4.10 and 4.11. Each diagram was generated using the proposed revisions to CIECAM16 (Section 3.1). In the case of Figure 4.10, the individual illuminant for each section (either D65 or Illuminant A) was used as the CIECAM16 white point. In the case of Figure 4.11, the mean neutral point from the relevant conditions from Experiment 1 (Table 4.1) was used as the white point for calculations. For both figures, opponent color axes α and β were calculated from CIECAM16 J, C, and h using a polar to rectangular Cartesian conversion, because CIECAM16 does not explicitly specify such axes in the output of the model.

If observers underwent instant and complete chromatic adaptation to each of the illuminants in the heterochromatic viewing environment as they were judging the hue of each box, the matching CIECAM16 hue angle under Illuminant A would be the same or close to the CIECAM16 hue angle of the D65 reference. Instead, the Illuminant-A-condition hue angles are offset from the D65-condition hue angles (Figure 4.10). The direction of the offset in the data is consistent with the observers partially compensating for the color of the illumination. For instance, for the green color center, a yellower (after complete chromatic adaptation) cube under D65 was matched to a bluer cube under Illuminant A. This suggests that the color of the illumination was partially mixed with the perceived color of the boxes—the tristimulus values of the box under the D65 metamers are bluer than its tristimulus values under the Illuminant A metamers. The same dynamic can readily be seen in the red color center, where the Illuminant-A-adapted hue angles of the preferred Illuminant A boxes are

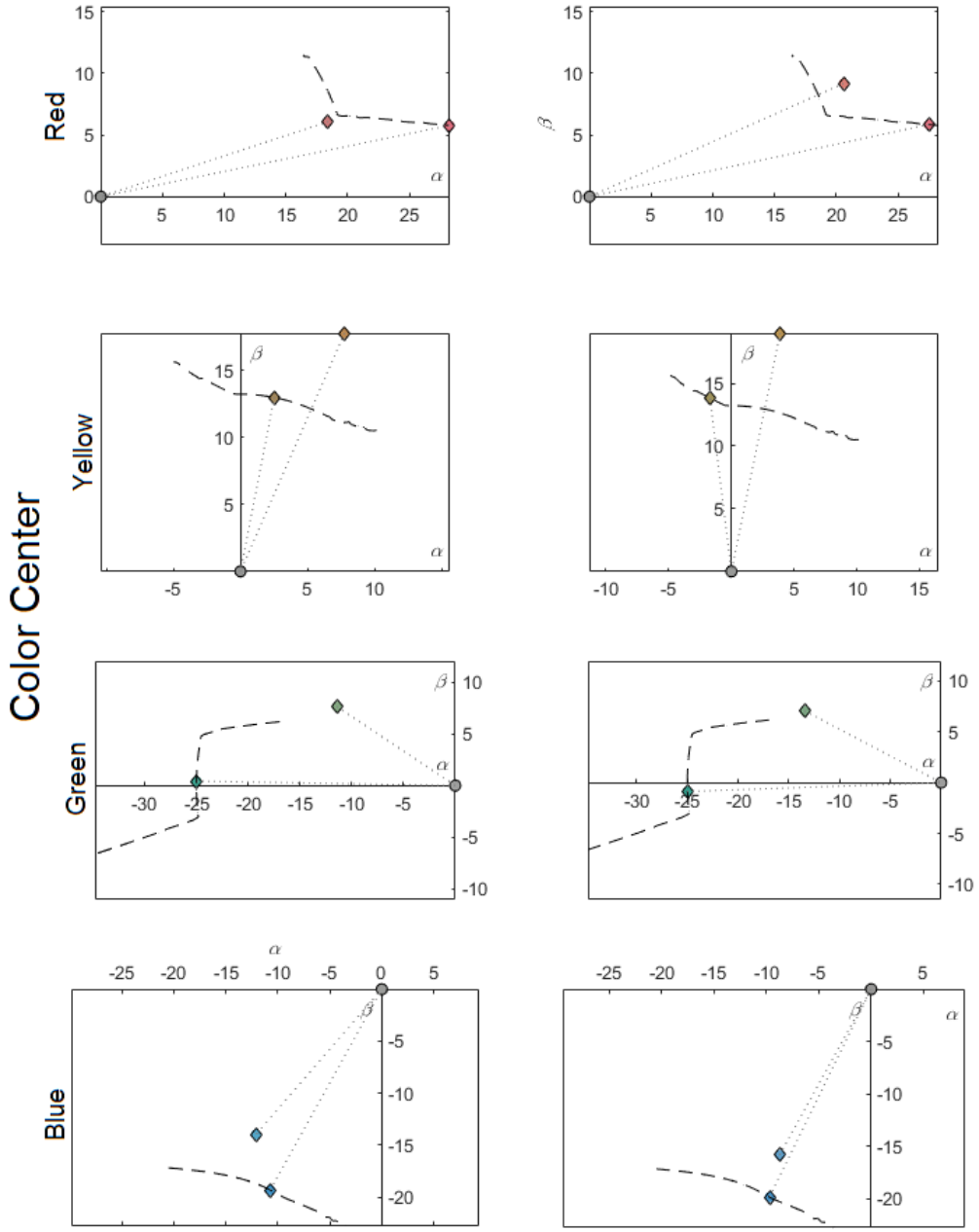


Figure 4.10: The color of hue-matched boxes are plotted as diamonds on $\alpha - \beta$ opponent color axes derived from predicted chroma and hue angle scales following the proposed revisions to CIECAM16 (Section 3.1). The color appearance of each box was calculated assuming complete adaptation to the color of each box's specific illumination. The dashed lines indicate the range of possible box colors that were tested under Illuminant A metamers and are thus also useful in identifying which of each pair of diamonds refers to the Illuminant A box and which refers to the D65 box (which is not connected to the dashed line).

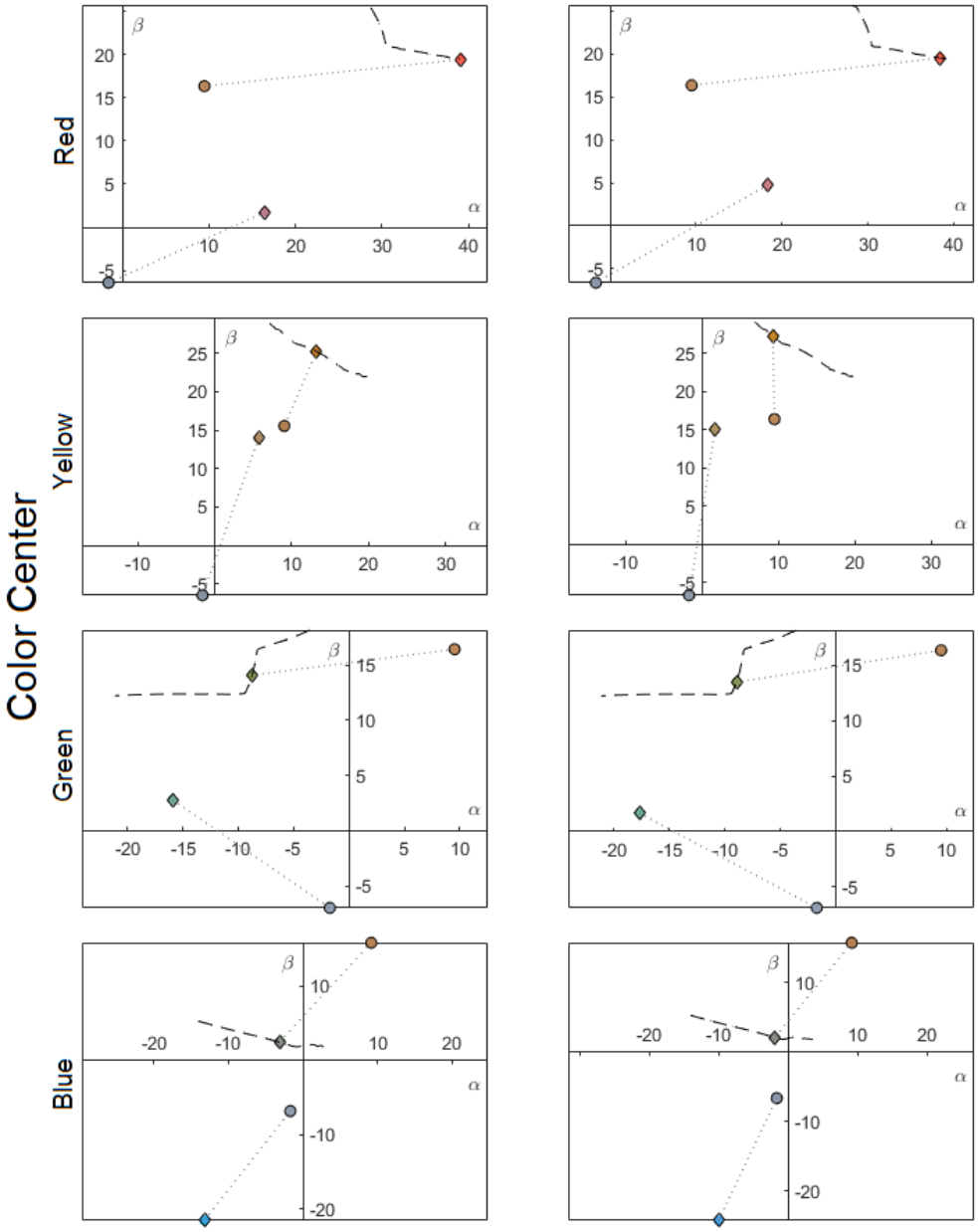


Figure 4.11: The color of hue matched boxes are plotted as diamonds on $\alpha - \beta$ opponent color axes derived from predicted chroma and hue angle scales following the proposed revisions to CIECAM16 3.1. The color appearance of each box was calculated assuming adaptation to the observers' predicted sensory neutral point, 5180 K daylight. The colors of the illumination are therefore not neutral and are also plotted as circles on the diagrams.

bluer than the hue angles of the matched boxes under D65.

Directly interpreting these results from the superficial trends is simplest for the red and green color centers, since they are perceptually orthogonal to the blue-yellow axis of variation of the color of the illumination. The correct relationship between the hue angles is more difficult to read from the yellow and blue hue angles, where it's not apparent without further calculation which direction the illumination would be pushing the perception of the boxes. Such calculation involves using partial chromatic adaptation to derive an adapted white point for the observer between the white point of the illumination in question and the observer's internal resting reference white point.

Figure 4.11 suggests an alternative paradigm for analyzing whether the illuminant was discounted by observers. The calculation of hue angle represented by complete adaptation (as in Figure 4.10) implies that cognitive discounting of the illuminant follows the same mathematical von Kries form as sensory adaptation, but there is no explicit evidence for such an assumption. By using our best estimate of the state of sensory adaptation of the average observer, Figure 4.11 shows the predicted perceived colors of both the boxes and the illuminations in the entire heterochromatic scene. Thus, this represents the perceptual information that the observer is using to judge the colors. However, the angular relationship between each box and its respective illumination is not significantly changed from the full von Kries state and thus does not provide additional insight into the mechanism of discounting the illuminant.

A third potential path for evaluating these hue match pairs is suggested by Berns and Choh [96]. This study evaluated appearance matches across hard copy and soft copy media within different white points and found that the appearance match preferred by observers when they assumed that the observer would be 75% adapted to whichever medium they were viewing at that moment and 25% adapted to the other medium. Unique, partially adapted states would also be plausible given our experimental method and may be the result of cognitive discounting of the illuminant.

In the case of our heterochromatic lighting condition, we can use the results of Experiment 1 to estimate a resting state of sensory adaptation for the observers in Experiment 2. This adaptation state will serve as the baseline white point from which observers' white points are changed as they use cognitive mechanisms to partially discount the color of each illumination. (This approach also allows for fast chromatic adaptation that would occur as observers look back and forth between the light booth sections.) The mean correlated color temperature (CCT) of the white point from the observers in Experiment 1 in the D65 versus Illuminant A condition was 5180 K (daylight). The accuracy of this value has surprisingly little effect on the results of the following calculations, thus we need not overly concern ourselves with

the interobserver differences observed in Experiment 1 nor the fact that Experiment 1 and Experiment 2 consisted of separate observer pools.

We can then find the degree of partial chromatic adaptation from this baseline white point towards each of the adapting illuminants at which the stimuli would have the same CIECAM16 hue angle. Instead of using the default CIECAM16 method for implementing partial chromatic adaptation, it is more appropriate in this context to use the weighted-geometric-mean method proposed by Shen & Fairchild[6], which improves upon several of the flaws in the CIECAM16 formula. Specifically relevant for this case, predicted white points follow the Planckian and daylight loci more closely in the weighted-geometric-mean method, as did our stimuli in Experiment 1.

The implementation of the model is as follows. Given the CIECAM16 pseudo-cone values for the (reference) baseline white point, $R_rG_rB_r$, and the pseudo-cone values for the adapting illuminant, $R_wG_wB_w$, the effective white point, $R_eG_eB_e$, after partial chromatic adaptation to degree D is calculated by:

$$R_e = R_w^D R_r^{1-D} \tag{4.1}$$

$$G_e = G_w^D G_r^{1-D} \tag{4.2}$$

$$B_e = B_w^D B_r^{1-D} \tag{4.3}$$

This effective white point can then be used to model typical von Kries chromatic adaptation as implemented in CIECAM16.

The hue angle of each of the matched stimuli as a function of the degree of adaptation is shown in Figure 4.12, with the dashed lines representing the reference stimuli under D65 and the solid lines representing the median matched stimuli under the Illuminant A metamers. If observers were able to completely discount the color of the illumination when making hue matches, each pair of dashed and solid lines would intersect where $D = 1$. (The ordinate of each line at $D = 1$ matches the hue angles from Figures 4.7 and 4.8.) If observers did not discount the illuminant at all nor experienced any fast chromatic adaptation as they looked back and forth, then the solid lines would intersect at $D = 1$. Instead, all intersections occur between these endpoints, indicating that observers partially, but not completely, discounted the illuminant in their judgment of the boxes' hues.

As discussed earlier, the red boxes were too blue under D65 to match any of the box hues under Illuminant A. Thus, solid lines (representing Illuminant A) should be lower than they are in Figure 4.12 and the degree of adaptation value should be less. Similarly, the inter-observer variability discussed above leads to a lack of confidence in the blue curves at the top of the figure. Thus, the two central pairs of lines represent the highest quality

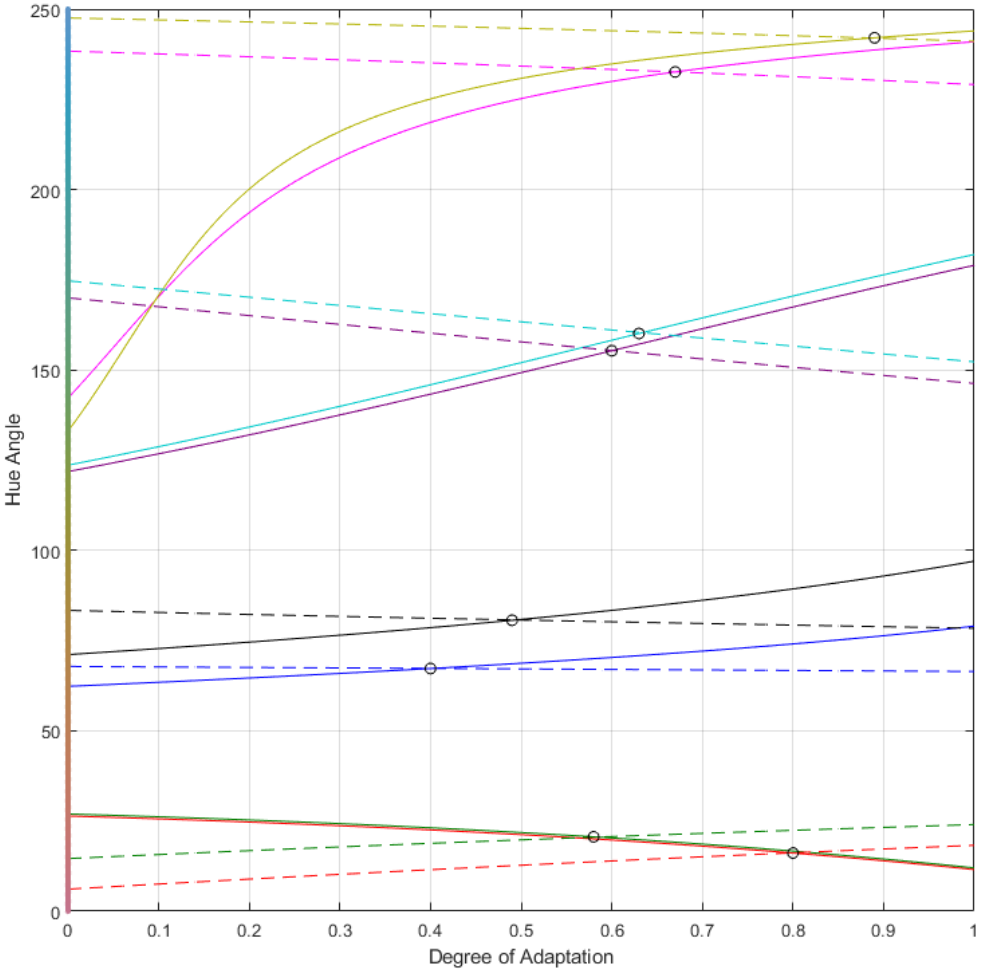


Figure 4.12: Predicted hue angle of each box as a function of the degree of partial adaptation to the test illumination from 5180 K daylight used to calculate a CIECAM16 adaptation point. Solid lines indicate the Illuminant A boxes and dashed lines indicate D65 boxes, with the color of each line used to identify its pair. The circles indicate the degree of adaptation at which the boxes are predicted to match in hue.

data and predict effective degrees of adaptation ranging from 0.40 to 0.63. These predicted white points are shown relative to the resting sensory adaptation point (5180 K daylight) and with the colors of the chosen boxes as well in Figure 4.13. While the Berns and Choh paper only tested one degree of partial adaptation, 0.75, their value agrees with the values that we report here [96].

4.4 Conclusion

Heterochromatic lighting environments and the associated psychophysical method of simultaneous asymmetric color matching are important experimental tools for researchers of chromatic adaptation. Our experimental development and approach rooted in color appearance modeling represent significant improvements on the current standard for these experiments. Taking advantage of the metamerism of multi-primary, spectrally tunable illumination alongside the tunability of surface reflectance via Kubelka-Munk simulations allowed our experiment to supersede the limiting assumptions made by past "color constancy" researchers to study chromatic adaptation with objects without the chosen spectral reflectances biasing the results or analysis of the experiment. Our more rigorous approach also gave us the opportunity to quantify how observers cognitively discounted the illuminant, opening the path for future studies that complement existing experimental paradigms for measuring chromatic adaptation.

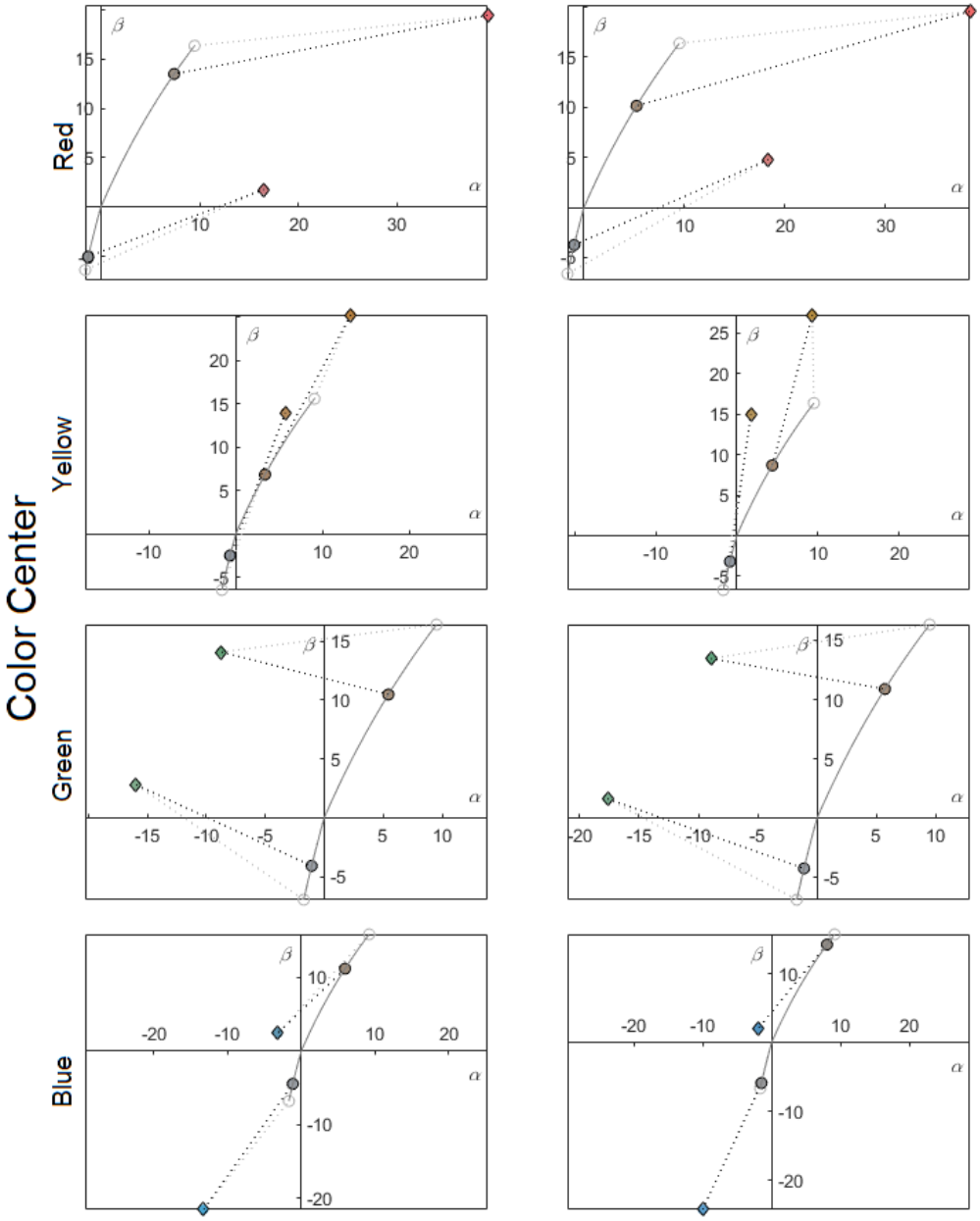


Figure 4.13: The color of hue matched boxes are plotted as diamonds on $\alpha - \beta$ opponent color axes derived from predicted chroma and hue angle scales following the proposed revisions to CIECAM16 3.1. The color appearance of each box was calculated assuming adaptation to the observers' predicted sensory neutral point, 5180 K daylight. The solid lines indicate the range of color coordinates of neutral points assuming partial adaptation to each individual illumination, ranging from no adaptation (the origin) to full adaptation (empty circles). Filled circles indicate the neutral point for the degree of adaptation at which the pair of boxes are predicted to have equal perceived hue.

Chapter 5

Conclusion

5.1 Summary

This thesis reports the following improvements to modeling brightness within the framework of color appearance modeling:

- Improved CIECAM16 by linearizing the relationship between lightness and brightness.
- Simplified and improved the CIECAM16 measures of chromatic intensity.
- Developed an initial model of the brightness of chromatic color that accounts for the Helmholtz-Kohlrausch effect.
- Conducted a series of three psychophysical experiments to generate new data on the Helmholtz-Kohlrauch effect.
- Fit a new model of the brightness of chromatic colors to our data.
- Testing our model of brightness on high-dynamic-range images.

Additionally, this thesis reports the following achievements in our studies of cognitive phenomena related to chromatic adaptation:

- Developed new psychophysical procedures to measure the sensory adaptation of observers to heterochromatic viewing environments.
- Conducted an asymmetric hue matching experiment with the heterochromatic lighting environment that overcame the shortcomings of previous studies.
- Measured the degree to which observers cognitively discount the color of the illumination in their perception of hue.

5.2 Discussion

This thesis demonstrates the power of color appearance models in designing and analyzing psychophysical experiments. Our finding that the strength of the Helmholtz-Kohlrausch (H-K) effect is invariant to chroma and hue could only have been seen with a chroma metric that was perceptually linear and uniform to hue. Thus, the greater computational investment on the front end of color appearance modeling paid off with a final model that was simpler than a model that only used CIELAB. Using CAMs in this way has a benefit for the models, as well, as this type of implementation can reveal flaws or areas for improvement in a model. This was demonstrated by our refinement of the relationship between achromatic brightness and achromatic lightness in CIECAM16 following our discovery of a perceptual paradox in the model.

CAMs were also key to designing our experiments on chromatic adaptation. The 7-channel LED illumination, along with our ability to spectrally tune reflectances with Kubelka-Munk theory, provided an exceptional number of degrees of freedom that could only be properly constrained and properly deployed via color appearance modeling. The use of multiple white points in the experiment also called for the reliable chromatic adaptation transforms implemented by CAMs.

This integration of color appearance modeling into our experimental modalities also reveals surprising instances where observer behavior seems to depart from the definitions used by color scientists. In our HDR image brightness experiment, this occurred when some observers appeared to conflate saturation and brightness and rated less luminant (but more saturated) images as brighter. In the heterochromatic viewing condition experiment, we saw a wide range in observer tolerance for hue matches. Future experiments that incorporate explicit color appearance modeling in their design and analysis have the potential to shed more light on how observers interpret these color appearance terms when completing experimental tasks.

Both sets of experiments have led to important new insights into cognitive mechanisms of color appearance. However, challenges remain for broader application of color appearance models to the applications we have introduced. A stated goal of CAMs is to be able to properly account for the effects of viewing conditions, such as luminance level, background luminance, and surrounds conditions, on visual perception. The current generation of CAMs, such as CIECAM16, are not able to reliably and rigorously account for these effects. Thus, while the scales work very well within one set of viewing conditions, we cannot use the CAM to maintain constant color appearance for images across viewing conditions. Further research should be done to improve the rigor of these parameters within CIECAM16 and other models.

Additionally, even within one set of viewing conditions, our model had trouble making predictions of brightness for high-dynamic-range images. This failure is at least partially due to spatial effects, which are not accounted for in CIECAM16. Modular additions to the CAM that would allow for spatial or temporal effects would greatly increase the utility of the model.

5.3 Contributions

5.3.1 Journal Articles

Hellwig, L., Stolitzka, D., & Fairchild, M. D. (2023). The Brightness of Chromatic Colors. *Color Research & Application*, [Accepted].

Hellwig, L., & Fairchild, M. D. (2022). Brightness, lightness, colorfulness, and chroma in CIECAM02 and CAM16. *Color Research & Application*, *47*(5), 1083–1095.

Hellwig, L., Stolitzka, D., & Fairchild, M. D. (2022). Extending CIECAM02 and CAM16 for the Helmholtz–Kohlrausch effect. *Color Research & Application*, *47*(5), 1096–1104.

Hellwig, L., & Fairchild, M. D. (2020). Using Gaussian Spectra to Derive a Hue-linear Color Space. *Journal of Perceptual Imaging*, *3*(2), 020401-1–020401-8

5.3.2 Conference Proceedings, Presentations, and Posters

Hellwig, L., Stolitzka, D., & Fairchild, M. (2022, November). Novel methods of brightness and saturation testing for high-dynamic-range images. In *Color and Imaging Conference* (Vol. 31, No. 1, pp.122-125). Society for Imaging Science and Technology.

Hellwig, L., Stolitzka, D., & Fairchild, M. (2022, September). Improvements to CIECAM16 and Future Directions. In *Proc. 30th Session*. Commission Internationale de l'Eclairage.

Hellwig, L., Stolitzka, D., & Fairchild, M. D. (2023, June). The Brightness of Chromatic Colors. *Inter-Society Color Council: Color Impact* [Presentation Only].

Hellwig, L., & Fairchild, M. D. (2023). Chromatic Adaptation to Heterochromatic Illumination. *Journal of Vision*, 23(11), 51. [Abstract – poster presented at Optica Fall Vision Meeting 2022]

Hellwig, L., Stolitzka, D., & Fairchild, M. (2022, November). Why Achromatic Response is not a Good Measure of Brightness. In *Color and Imaging Conference* (Vol. 30, No. 1, pp. 1-5). Society for Imaging Science and Technology.

Hellwig, L., Stolitzka, D., & Fairchild, M. D. (2022, November). Revising CAM-16. *Color and Imaging Conference* [Poster Only].

Hellwig, L., Stolitzka, D., Yi, Y., & Fairchild, M. D. (2022, June). 75-1: Student Paper: Brightness and Vividness of High Dynamic Range Displayed Imagery. In *SID Symposium Digest of Technical Papers* (Vol. 53, No. 1, pp. 1009-1012).

Hellwig, L., & Fairchild, M. D. (2020, November). Using Gaussian Spectra to Derive a Hue-linear Color Space. In *Color and Imaging Conference* (Vol. 28, No. 1, pp. 244-251). Society for Imaging Science and Technology.

Bibliography

[1] "CIE 248:2022 the CIE 2016 colour appearance model for colour management systems: CIECAM16," tech. rep., CIE, Vienna, 2022.

[2] A. Stockman and L. T. Sharpe, "Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype," *Vision Research* **40**, pp. 1711–1737, 2000.

[3] A. Stockman, L. T. Sharpe, and C. C. Fach, "The spectral sensitivity of the human short-wavelength cones," *Vision Research* **39**, pp. 2901–2927, 1999.

[4] C. Li, Z. Li, Z. Wang, Y. Xu, M. R. Luo, G. Cui, M. Melgosa, M. Brill, and M. Pointer, "Comprehensive color solutions: CAM16, CAT16, and CAM16UCS," *Color Res. Appl.* **42**, pp. 703–718, 2017.

[5] F. W. Dunn, M. J. Lankheet, and F. Rieke, "Light adaptation in cone vision involves switching between receptor and post-receptor sites," *Nature* **449**, pp. 603–607, 2007.

[6] C. Shen and M. D. Fairchild, "Weighted geometric mean (wgm) method: A new chromatic adaptation model," *Plos one* **18**(8), p. e0290017, 2023.

[7] T. Kunkel and E. Reinhard, "A reassessment of the simultaneous dynamic range of the human visual system," in *APGV '10: Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pp. 17–24, 2010.

[8] F. Jiang and M. D. Fairchild, "Preliminary result on the direct assessment of perceptible simultaneous luminance dynamic range," *Journal of Imaging Science and Technology* **65**(5), p. 50401, 2021.

[9] M. D. Fairchild and G. M. Johnson, "iCAM framework for image appearance, differences, and quality," *Journal of Electronic Imaging* **13**(1), pp. 126–138, 2004.

[10] M. R. Luo, A. A. Clarke, P. A. Rhodes, A. Schappo, S. A. R. Scrivener, and C. J. Tait, "Quantifying colour appearance. Part I. LUTCHI colour appearance data," *Color Res. Appl.* **16**(3), pp. 166–180, 1991.

[11] M. R. Luo, A. A. Clarke, P. A. Rhodes, A. Schappo, S. A. R. Scrivener, and C. J. Tait, "Quantifying colour appearance. Part II. Testing colour models performance using LUTCHI colour appearance data," *Color Res. Appl.* **16**(3), pp. 181–197, 1991.

[12] J. Wu, L. Zhang, S. Isikman, and C. Chen, "Enhanced viewing experience considering chromatic adaptation," *SID Digest* **2019**, pp. 857–860, 2019.

[13] K. A. Smet, Q. Zhai, M. R. Luo, and P. Hanselaer, "Study of chromatic adaptation using memory color matches, Part II: colored illuminants," *Optics express* **25**(7), pp. 8350–8365, 2017.

[14] R. Peng, M. Cao, Q. Zhai, and M. R. Luo, "White appearance and chromatic adaptation on a display under different ambient lighting conditions," *Color Res. Appl.* **46**(5), pp. 1034–1045, 2021.

[15] Y. Zhu, M. Wei, and M. R. Luo, "Investigation on effects of adapting chromaticities and luminance on color appearance on computer displays using memory colors," *Color Res. Appl.* **45**(4), pp. 612–621, 2020.

[16] M. Wei and S. Chen, "Effects of adapting luminance and CCT on appearance of white and degree of chromatic adaptation," *Optics express* **27**(6), pp. 9276–9286, 2019.

[17] Z. Huang and M. Wei, "Effects of adapting luminance and CCT on appearance of white and degree of chromatic adaptation, part II: extremely high adapting luminance," *Optics Express* **29**(25), pp. 42319–42330, 2021.

[18] H. E. Smithson, "Sensory, computational and cognitive components of human colour constancy," *Phil. Trans. R. Soc. B* **360**, pp. 1329–1346, 2005.

[19] M. D. Fairchild, *Color Appearance Models*, John Wiley & Sons, Ltd, West Sussex, third ed., 2013.

[20] A. D. Logvinenko, B. Funt, H. Mirzaei, and R. Tokunaga, "Rethinking colour constancy," *PLOS One* **10**(9), p. e0135029, 2015.

[21] D. H. Foster, "Color constancy," *Vision Research* **51**, pp. 674–700, 2011.

[22] D. H. Brainard, N. P. Cotaris, and A. Radonjic, "The perception of color and material in naturalistic tasks," *Interface Focus* **8**, p. 20180012, 2018.

[23] L. Arend and A. Reeves, "Simultaneous color constancy," *J. Opt. Soc. Am. A* **3**(10), pp. 1743–1751, 1986.

[24] L. E. Arend, A. Reeves, J. Schirillo, and R. Goldstein, "Simultaneous color constancy: papers with diverse Munsell values," *J. Opt. Soc. Am. A* **8**(4), pp. 661–672, 1991.

[25] R. W. G. Hunt, "Colour terminology," *Color Res. Appl.* **3**(2), pp. 79–87, 1978.

[26] R. S. Berns, *Billmeyer and Saltzman's Principles of Color Technology*, John Wiley & Sons, 2019.

[27] "CIE 159:2004 a colour appearance model for colour management systems: CIECAM02," tech. rep., CIE, Vienna, 2004.

[28] N. Moroney, M. D. Fairchild, R. W. G. Hunt, C. Li, M. R. Luo, and T. Newman, "The CIECAM02 color appearance model," tech. rep., Rochester Institute of Technology, 2002.

[29] R. W. G. Hunt, "A model of colour vision for predicting colour appearance," *Color Res. Appl.* **7**(2), pp. 95–112, 1982.

[30] R. W. G. Hunt and M. R. Pointer, "A colour-appearance transform for the CIE 1931 standard colorimetric observer," *Color Res. Appl.* **10**(3), pp. 165–179, 1985.

[31] R. W. G. Hunt, "A model of color vision for predicting colour appearance in various viewing conditions," *Color Res. Appl.* **12**(6), pp. 297–314, 1987.

[32] R. W. G. Hunt, "Revised colour-appearance model for related and unrelated colours," *Color Res. Appl.* **16**(3), pp. 146–165, 1991.

[33] R. W. G. Hunt, *The Reproduction of Colour, 5th Ed.*, ch. A model of color vision for practical applications. Fountain Press, 1995.

[34] C. J. Bartleson, "Measures of brightness and lightness," *Die Farbe* **28**, pp. 132–148, 1980.

[35] R. Hunt, C. Li, and M. Luo, "Dynamic cone response functions for models of colour appearance," *Color Res. Appl.* **28**(2), pp. 82–88, 2003.

[36] C. Li, M. R. Luo, R. W. G. Hunt, N. Moroney, M. D. Fairchild, and T. Newmann, "The Performance of CIECAM02," in *28th Color and Imaging Conference*, pp. 28–32, 2002.

[37] R. W. G. Hunt, "An improved predictor of colourfulness in a model of colour vision," *Color Res. Appl.* **19**(1), pp. 23–26, 1994.

[38] M. R. Luo and B. Rigg, "Chromaticity-discrimination ellipses for surface colours," *Color Res. Appl.* **11**(1), pp. 25–42, 1986.

[39] M. Melgosa, R. Huertas, and R. S. Berns, "Performance of recent advanced color-difference formulas using the standardized residual sum of squares index," *J. Opt. Soc. Am. A* **25**(7), pp. 1828–1834, 2008.

[40] Y. Nayatani, "Why two kinds of color order systems are necessary?," *Col. Res. Appl.* **30**(4), pp. 295–303, 2005.

[41] S. M. Newhall, D. Nickerson, and D. B. Judd, "Final report of the O.S.A. subcomittee on the spacing of the Munsell colors," *J. Opt. Soc. Am.* **33**(7), pp. 385–418, 1943.

[42] M. H. Kim, T. Weyrich, and J. Kautz, "Modeling human color perception under extended luminance levels," in *ACM SIGGRAPH 2009 papers*, pp. 1–9, 2009.

[43] "Comission internationale de l'eclairage proceedings, 1924," tech. rep., CIE, Vienna, 1924.

[44] L. T. Sharpe, A. Stockman, W. Jagla, and H. Jägle, "A luminous efficiency function, $V^*(\lambda)$, for daylight adaptation," *Journal of Vision* **5**(11), pp. 948–968, 2005.

[45] R. A. Bone and J. T. Landrum, "Heterochromatic flicker photometry," *Archives of biochemistry and biophysics* **430**(2), pp. 137–142, 2004.

[46] B. Lee, P. Martin, and A. Valberg, "The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina.," *The Journal of Physiology* **404**(1), pp. 323–347, 1988.

[47] G. Brindley, J. Du Croz, and W. Rushton, "The flicker fusion frequency of the blue-sensitive mechanism of colour vision," *The Journal of Physiology* **183**(2), pp. 497–500, 1966.

[48] J. P. Comerford and P. K. Kaiser, "Luminous-efficiency functions determined by heterochromatic brightness matching," *J. Opt. Soc. Am.* **65**(4), pp. 466–468, 1975.

[49] C. L. Sanders and G. Wyszecki, "Correlate for brightness in terms of CIE color matching data," in *Proc. 15th Session*, pp. 221–230, CIE, (Vienna), 1963.

[50] G. Wyszecki, "Correlate for lightness in terms of CIE chromaticity coordinates and luminous reflectance," *J. Opt. Soc. Am* **57**, pp. 254–257, 1967.

[51] G. Wagner and R. M. Boynton, "Comparison of four methods of heterochromatic photometry," *J. Opt. Soc. Am.* **62**(12), pp. 1508–1515, 1972.

[52] M. Ikeda and Y. Nakano, "Spectral luminous-efficiency functions obtained by direct heterochromatic brightness matching for point sources and for 2 and 10 fields," *J. Opt. Soc. Am. A* **3**(12), pp. 2105–2108, 1986.

[53] P. K. Kaiser and G. Wyszecki, "Additivity failures in heterochromatic brightness matching," *Color Res. Appl.* **3**(4), pp. 177–182, 1978.

[54] Y. Nayatani, Y. Umemura, H. Sobagaki, K. Takahama, and K. Hashimoto, "Lightness perception of chromatic object colors," *Color. Res. Appl.* **16**(1), pp. 16–25, 1991.

[55] Y. Nayatani, Y. Gomi, M. Kamei, H. Sobagaki, and K. Hashimoto, "Perceived lightness of chromatic objects including highly saturated colors," *Color Res. Appl.* **17**(2), pp. 127–141, 1992.

[56] M. D. Fairchild and E. Pirrotta, "Predicting the lightness of chromatic object colors using cielab," *Color Res. Appl.* **16**(6), pp. 385–393, 1991.

[57] S. L. Guth and H. R. Lodge, "Heterochromatic additivity, foveal sensitivity, and a new color model," *J. Opt. Soc. Am.* **63**(4), pp. 450–462, 1973.

[58] B. B. Lee, "Visual pathways and psychophysical channels in the primate," *The Journal of Physiology* **589**(1), pp. 41–47, 2011.

[59] P. K. Kaiser and J. P. Comerford, "Flicker photometry of equally bright lights," *Vision Research* **15**(12), pp. 1399–1402, 1975.

[60] P. K. Kaiser, "Minimally distinct border as a preferred psychophysical criterion in visual heterochromatic photometry," *J. Opt. Soc. Am.* **61**(7), pp. 966–971, 1971.

[61] D. Kelly and D. Van Norren, "Two-band model of heterochromatic flicker," *J. Opt. Soc. Am.* **67**(8), pp. 1081–1091, 1977.

[62] "ISO/CIE 11664:2019(E) Colorimetry," tech. rep., CIE, Vienna, 2019.

[63] M. Kim, J. H. Jo, Y. K. Park, and S. W. Lee, "Amendment of CIECAM02 with a technical extension to compensate Helmholtz-Kohlrausch effect for chromatic characterization of display devices," *Displays* **56**, pp. 1–10, 2019.

[64] S. Hermans, K. A. G. Smet, and P. Hanselaer, "Color appearance model for self-luminous stimuli," *J. Opt. Soc. Am. A* **35**(12), pp. 2000–2009, 2018.

[65] M. Kleiner, D. Brainard, and D. Pelli, "What's new in Pscyhtoolbox-3?," in *ECVP*, 2007.

[66] ITU-R, "BT.2100-2, image parameter values for high dynamic range television for use in production and international programme exchage," *Int. Telecommunication Union* , 2018.

[67] Y. Nayatani, "Simple estimation methods for the Helmholtz-Kohlrausch effect," *Color Res. Appl.* **22**(6), pp. 385–401, 1997.

[68] R. G. Kuehni, "The early development of the Munsell system," *Color Res. Appl.* **27**(1), pp. 20–27, 2002.

[69] S. L. Guth, "Model for color vision and light adaptation," *J. Opt. Soc. Am. A* **8**(6), pp. 976–993, 1991.

[70] R. S. Berns, "Extending CIELAB: Vividness, depth, and clarity," *Color Res. Appl.* **39**(4), pp. 322–330, 2014.

[71] H. Xie and M. D. Fairchild, "G_0 revisited as equally bright reference boundary," in *29th Color and Imaging Conference*, pp. 247–252, 2021.

[72] M. D. Fairchild, "A digital test chart for visual assessment of color appearance scales," in *Color and Imaging Conference*, **29**, pp. 160–165, Society for Imaging Science and Technology, 2021.

[73] R. M. Evans, "Fluorescence and gray content of surface colors," *J. Opt. Soc. Am.* **49**(11), pp. 1049–1059, 1959.

[74] R. M. Evans and B. K. Swenholt, "Chromatic strength of colors: dominant wavelength and purity," *J. Opt. Soc. Am.* **57**(11), pp. 1319–1324, 1967.

[75] M. D. Fairchild and R. L. Heckaman, "Deriving appearance scales.," in *Color Imaging Conference*, pp. 281–286, 2012.

[76] R. L. Donofrio, "Review Paper: The Helmholtz-Kohlrausch effect," *Journal of the SID* **19**(10), pp. 658–664, 2011.

[77] Y. Nayatani, H. Sobagaki, and K. Hashimoto, "Existence of two kinds of representations of the Helmholtz-Kohlrausch Effect. I. The experimental confirmation," *Color Res. Appl.* **19**(4), pp. 246–261, 1994.

[78] Y. Nayatani, H. Sobagaki, and K. Hashimoto, "Existence of two kinds of the Helmholtz-Kohlrausch Effect. II. The models," *Color Res. Appl.* **19**(4), pp. 262–272, 1994.

[79] S. F. Liao, H. Y. Chou, T. H. Yang, C. C. Lee, and K. Chang, "20.1: Perceived brightness of LED projector," *SID Symposium Digest of Technical Papers* **40**, pp. 262–264, 2009.

[80] M. R. Luo and R. W. G. Hunt, "The Structure of the CIE 1997 Colour Appearance Model (CIECAM97s)," *Color Res. Appl.* **23**(3), pp. 138–146, 1998.

[81] M. Withouck, K. A. G. Smet, W. R. Ryckaert, and P. Hanselaer, "Experimental driven modelling of the color appearance of unrelated self-luminous stimuli: CAM15u," *Optics Express* **23**(9), pp. 12045–12064, 2015.

[82] Y. Yang, M. R. Luo, and W. J. Huang, "Assessing glare, part 3: Glare sources having different colours," *Lighting Res Technol* **50**, pp. 596–615, 2018.

[83] L. Hellwig and M. D. Fairchild, "Brightness, lightness, colorfulness, and chroma in CIECAM02 and CAM16," *Color Res. Appl.* **47**(5), pp. 1083–1095, 2022.

[84] M. D. Fairchild, "Von kries 2020: Evolution of degree of chromatic adaptation," in *28th Color and Imaging Conference*, pp. 252–257, 2020.

[85] L. Hellwig, D. Stolitzka, Y. Yi, and M. D. Fairchild, "75-1: Brightness and vividness of high dynamic range displayed imagery," in *SID 2022 Digest*, pp. 1009–1012, 2022.

[86] G. High, P. Green, and P. Nussbaum, "The Helmholtz-Kohlrausch effect on display-based light colours and simulated substrate colours," *Color Res. Appl.* **48**, 2023.

[87] L. Hellwig, D. Stolitzka, and M. D. Fairchild, "Why achromatic response is not a good measure of brightness," in *30th Color and Imaging Conference*, pp. 1–5, 2022.

[88] G. Seong, Y. Kwak, and H. Kim, "Heterochromatic brightness matching experiments to evaluate brightness prediction model including Helmholtz-Kohlrausch effect," *Electronic Imaging* **35**, pp. 186–1, 2023.

[89] S. J. Daly, "Visible differences predictor: an algorithm for the assessment of image fidelity," in *Human Vision, Visual Processing, and Digital Display III*, B. E. Rogowitz, ed., **1666**, pp. 2 – 15, International Society for Optics and Photonics, SPIE, 1992.

[90] J. Lubin, *Vision Models for Target Detection and Recognition*, ch. A visual discrimination model for imaging system design and evaluation, pp. 245–283.

[91] X. M. Zhang and W. B. A., "A spatial extension to CIELAB for digital color image reproduction," in *SID 1996 Digest*, pp. 731–734, 1996.

[92] J. J. McCann, "Retinex at 50: color theory and spatial algorithms, a review," *Journal of Electronic Imaging* **26**(3), p. 031204, 2017.

[93] W. Jung, Y. Kwak, K. J. Kang, and M. Y. Lee, "Image-quality change by white-boost function of WRGB OLED display," *SID Digest* **54**, pp. 1117–1120, 2023.

[94] E. J. Breneman, "Corresponding chromaticities for different states of adaptation to complex visual fields," *J. Opt. Soc. Am. A* **4**(6), pp. 1115–1129, 1987.

[95] K. M. Braun, M. D. Fairchild, and P. J. Alessi, "Viewing techniques for cross-media image comparisons," *Color Res. Appl.* **21**(1), pp. 6–17, 1996.

[96] R. S. Berns and H.-K. Choh, "Cathode-ray-tube to reflection-print matching under mixed chromatic adaptation using rlab," *Journal of Electronic Imaging* **4**(4), pp. 347–359, 1995.

[97] D. H. Brainard, W. A. Brunt, and J. M. Speigle, "Color constancy in the nearly natural image. 1. asymmetric matches," *J. Opt. Soc. Am. A* **14**(9), pp. 2091–2110, 1997.

[98] D. H. Brainard and W. T. Freeman, "Bayesian color constancy," *J. Opt. Soc. Am. A* **14**(7), pp. 1393–1411, 1997.

[99] M. Safdar, J. Y. Hardeberg, and M. R. Luo, "ZCAM, a colour appearance model based on a high dynamic range uniform colour space," *Optics Express* **29**(4), pp. 6036–6052, 2021.

[100] M. D. Fairchild, "Chromatic adaptation in hard copy/soft copy comparisons," in *Color Hard Copy and Graphic Arts II*, **1912**, pp. 47–61, SPIE, 1993.

[101] R. S. Berns, "Artist acrylic paint spectral, colorimetric, and image dataset," in *Archiving Conference*, **19**, pp. 45–49, Society for Imaging Science and Technology, 2022.

[102] R. S. Berns and M. Mohammadi, "Evaluating single-and two-constant Kubelka-Munk turbid media theory for instrumental-based inpainting," *Studies in conservation* **52**(4), pp. 299–314, 2007.

[103] C. Shen and M. D. Fairchild, "The threshold of color inconstancy," in *29th Color and Imaging Conference Final Program and Proceedings*, pp. 374–380, 2021.

[104] A. David, T. Esposito, K. Houser, M. Royer, K. A. Smet, and L. Whitehead, "A vector field color rendition model for characterizing color shifts and metameric mismatch," *Leukos* **16**(2), pp. 99–114, 2020.

[105] N. Schlömer, "Algorithmic improvements for the ciecam02 and cam16 color appearance models," *arXiv preprint arXiv:1802.06067* , 2018.

Appendices

Appendix A

Proposed Color Appearance Model: HellesCAM23

In Section 3.1, this thesis proposed key revisions and corrections to CIECAM16. Additionally, in section 3.4, we proposed an extension to this revised CAM, introducing two new output attributes, J_{HK} and Q_{HK}, to account for the Helmholtz-Kohlrausch effect. Furthermore, the MATLAB implementation of CIECAM16 used for these calculations contains improvements suggested by Nico Schlömer [105]. The full model resulting from these suggested changes is recorded below and can be referred to as HellesCAM23, after the German word for "bright."

A.1 Forward Model

A.1.1 Model Inputs:

- XYZ, tristimulus values of stimulus.
- $X_wY_wZ_w$, tristimulus values of diffuse white within the scene.
- L_a, luminance of adapting field in cd/m^2, which is defined as 20% of the luminance of the diffuse white within the scene, regardless of scene composition.
- Y_b, luminance factor of background.
- D, degree of adaptation.
- Surround parameters, F, c, and N_c, determined via Table A.1.

	F	c	N_c
Dark	0.8	0.525	0.8
Dim	0.9	0.59	0.9
Average	1.0	0.69	1.0

Table A.1: Surround parameters

A.1.2 Values which are independent of stimulus

$$k = \frac{1}{5L_A - 1} \tag{A.1}$$

$$F_L = 0.2k^4(5L_A) + 0.1(1 - k^4)^2(5L_A)^{\frac{1}{3}} \tag{A.2}$$

$$z = 1.48 + \sqrt{\frac{Y_b}{Y_w}} \tag{A.3}$$

$$\mathbf{M}_{16} = \begin{pmatrix} 0.401288 & 0.650173 & -0.051461 \\ -0.250268 & 1.204414 & 0.045854 \\ -0.002079 & 0.048952 & 0.953127 \end{pmatrix} \tag{A.4}$$

$$\begin{pmatrix} R_w \\ G_w \\ B_w \end{pmatrix} = \mathbf{M_{16}} \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} \tag{A.5}$$

$$R_{aw} = G_{aw} = B_{aw} = 400 \left(\frac{F_L^{0.42}}{F_L^{0.42} + 27.13} \right) \tag{A.6}$$

$$A_w = 2R_{aw} + G_{aw} + \frac{B_{aw}}{20} \tag{A.7}$$

A.1.3 Cone responses

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \mathbf{M_{16}} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{A.8}$$

A.1.4 Chromatic Adaptation

$$\begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix} = \begin{pmatrix} \frac{Y_w}{R_w} & 0 & 0 \\ 0 & \frac{Y_w}{G_w} & 0 \\ 0 & 0 & \frac{Y_w}{B_w} \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{A.9}$$

A.1.5 Tone compression

$$R_a = 400 \left(\frac{\left(\frac{F_L R_c}{100}\right)^{0.42}}{\left(\frac{F_L R_c}{100}\right)^{0.42} + 27.13} \right) \tag{A.10}$$

Calculate G_a and B_a similarly.

A.1.6 Opponent dimensions and hue

$$a = R_a - \frac{12G_a}{11} + \frac{B_a}{11} \tag{A.11}$$

$$b = \frac{R_a + G_a - 2B_a}{9} \tag{A.12}$$

$$h = \arctan\left(\frac{b}{a}\right) \tag{A.13}$$

A.1.7 Eccentricity

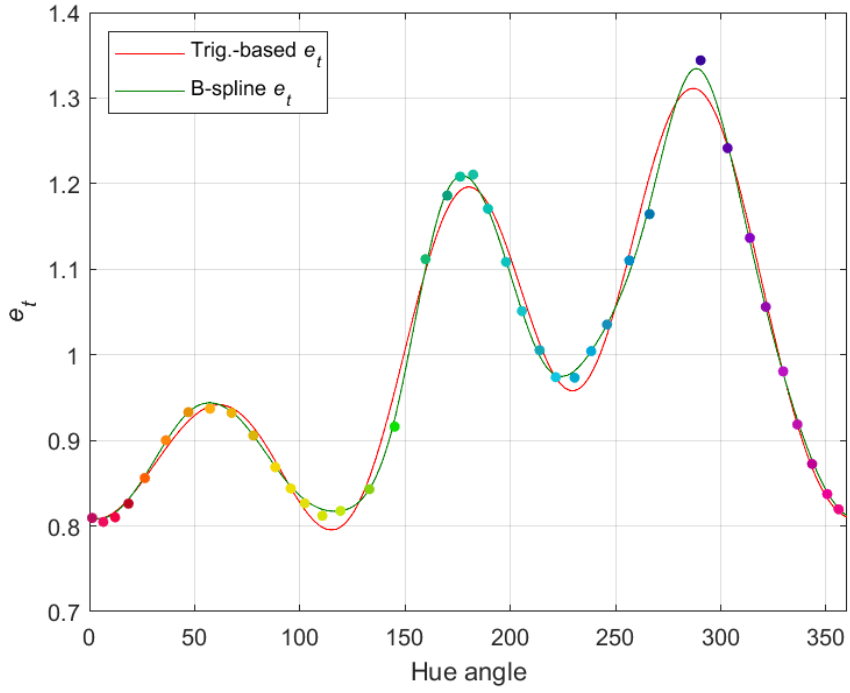


Figure A.1: Current, trigonometric eccentricity function (red; Equation A.14) compared to the B-spline eccentricity function (green; Equations A.15 and A.16).

$$\begin{aligned} e_t = &-0.0582\cos(h) - 0.0258\cos(2h) - 0.1347\cos(3h) + 0.0289\cos(4h) \\ &-0.1475\sin(h) - 0.0308\sin(2h) + 0.0385\sin(3h) + 0.0096\sin(4h) + 1 \end{aligned} \tag{A.14}$$

The above formula was used for all calculations represented in this thesis. Recently, Laslo Hunhold has suggested that a 4th-order B-spline function could achieve the same performance as this trigonometry-based formula with much greater efficiency. The B-spline fit to the same Munsell data is defined by coefficients $\mathbf{c}$ and knots $\boldsymbol{\tau}$:

$$\mathbf{c} = \begin{pmatrix} 0.8121 \\ 0.8003 \\ 0.8397 \\ 0.9582 \\ 0.9371 \\ 0.8326 \\ 0.805 \\ 0.8557 \\ 1.2605 \\ 1.1801 \\ 0.9398 \\ 0.9994 \\ 1.124 \\ 1.4352 \\ 1.1419 \\ 0.9102 \\ 0.8238 \\ 0.8121 \end{pmatrix}^{\mathbf{T}} \tag{A.15}$$

$$\boldsymbol{\tau} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 24 \\ 48 \\ 72 \\ 96 \\ 120 \\ 144 \\ 168 \\ 192 \\ 216 \\ 240 \\ 264 \\ 288 \\ 312 \\ 336 \\ 360 \\ 360 \\ 360 \\ 360 \end{pmatrix}^{\mathbf{T}} \tag{A.16}$$

The trigonometric and B-spline eccentricity functions are compared in Figure A.1.

A.1.8 Achromatic response

$$A = 2R_a + G_a + \frac{B_a}{20} \tag{A.17}$$

A.1.9 Achromatic Lightness

$$J = 100 \left(\frac{A}{A_W} \right)^{c \cdot z} \tag{A.18}$$

A.1.10 Achromatic Brightness

$$Q = \left(\frac{2}{c} \right) \left(\frac{J}{100} \right) (A_W) \tag{A.19}$$

A.1.11 Colorfulness

$$M = 47N_c \cdot e_t \cdot \sqrt{a^2 + b^2} \tag{A.20}$$

A.1.12 Chroma

$$C = 35\frac{M}{A_w} \tag{A.21}$$

A.1.13 Saturation

$$s = 100\frac{M}{A} \tag{A.22}$$

A.1.14 Lightness

$$J_{HK} = \sqrt{J^2 + 66C} \tag{A.23}$$

A.1.15 Brightness

$$Q_{HK} = \left(\frac{2}{c}\right)\left(\frac{J_{HK}}{100}\right)(A_W) \tag{A.24}$$

A.2 Inverse Model

A.2.1 Values which are independent of stimulus

Calculate F_L, z, R_w, G_w, B_w, and A_w as described in the forward model.

A.2.2 Achromatic response

From Q_{HK} and/or M:

$$J_{HK} = 50\left(\frac{c}{Q_{HK} \cdot A_W}\right) \tag{A.25}$$

$$C = 35\frac{M}{A_w} \tag{A.26}$$

Then, from J_{HK}:

$$J = \sqrt{(J_{HK})^2 - 66C} \tag{A.27}$$

From Q:

$$J = 50\left(\frac{c}{Q \cdot A_W}\right) \tag{A.28}$$

Then, from J:

$$A = A_w \left(\frac{J}{100}\right)^{\frac{1}{cz}} \tag{A.29}$$

Note: The model is not analytically invertible if only provided saturation, s, and either Q_{HK} or J_{HK} only. However, the inverse can quickly be computationally solved using Halley's method, which will be the topic of a future publication.

A.2.3 Opponent dimensions

From s:

$$M = \frac{s \cdot A}{100} \tag{A.30}$$

From C:

$$M = \frac{C \cdot A_w}{35} \tag{A.31}$$

Then, from M and h:

$$\begin{aligned} e_t = &-0.0582\cos(h) - 0.0258\cos(2h) - 0.1347\cos(3h) + 0.0289\cos(4h) \\ &-0.1475\sin(h) - 0.0308\sin(2h) + 0.0385\sin(3h) + 0.0096\sin(4h) + 1 \end{aligned} \tag{A.32}$$

$$\gamma = \frac{M}{43 e_t \cdot N_c} \tag{A.33}$$

$$a = \gamma\cos(h) \tag{A.34}$$

$$b = \gamma\sin(h) \tag{A.35}$$

A.2.4 Cone responses

$$\begin{pmatrix} R_a \\ G_a \\ B_a \end{pmatrix} = \frac{1}{1403} \begin{pmatrix} 460 & 451 & 288 \\ 460 & -891 & -261 \\ 460 & -220 & -6300 \end{pmatrix} \begin{pmatrix} A \\ a \\ b \end{pmatrix} \tag{A.36}$$

A.2.5 Inverse cone compression

$$R_c = \frac{100}{F_L} \left(\frac{27.13|R_a|}{400 - |R_a|}\right)^{\frac{1}{0.42}} \tag{A.37}$$

Calculate G_c and B_c

A.2.6 Inverse chromatic adaptation

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} \frac{R_w}{Y_w} & 0 & 0 \\ 0 & \frac{G_w}{Y_w} & 0 \\ 0 & 0 & \frac{B_w}{Y_w} \end{pmatrix} \begin{pmatrix} R_c \\ G_c \\ B_c \end{pmatrix} \tag{A.38}$$

A.2.7 Tristimulus values

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \mathbf{M_{16}}^{-1} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{A.39}$$

Appendix B

Experimental Data

There are two key sets of data collected in this experiment. The first is the brightness matching data from the experiments described in Section 3.4, which are presented in Tables B.1 to B.3. The second is the result of the hue matching experiment from Section 4.3, which is presented in Table. These data may also be downloaded from my profile on Research Gate: www.researchgate.net/profile/Luke-Hellwig-2.

Table B.1: Tristimulus values for the 1931 standard observer for pairs of stimuli matched in brightness following the method described in Section 3.4.1. The white point used to design the background pattern and for CAM calculations has tristimulus values of 474.4, 500, 543.6.

Test Stimulus			Reference Stimulus		
X	Y	Z	X	Y	Z
173.3	182.5	198.4	123.0	87.0	97.8
189.3	199.4	216.8	116.4	88.4	44.8
144.1	151.7	164.9	105.0	90.9	23.9
120.7	127.1	138.2	91.7	93.9	13.7
186.3	196.2	213.3	73.5	97.9	17.4
190.9	201.1	218.6	64.5	99.8	55.4
217.0	228.5	248.5	65.9	99.8	105.4
149.2	157.1	170.8	69.2	99.8	171.7
163.8	172.5	187.5	83.1	97.3	239.2
186.0	195.9	212.9	96.1	93.8	222.5
151.7	159.8	173.7	105.3	91.4	198.5
171.9	181.1	196.9	116.9	88.6	167.9
212.3	223.6	243.1	201.4	210.8	67.0
246.4	261.4	175.4	207.2	210.4	22.8

348.8	367.4	399.4	366.8	386.9	171.2
439.1	465.8	348.1	372.8	386.5	90.1
421.5	445.1	208.9	382.4	386.0	34.1
89.0	93.7	101.9	40.2	25.2	65.2
119.3	107.0	153.0	47.7	24.0	88.5
204.6	189.4	256.7	131.4	85.9	207.9
246.8	204.1	334.5	146.6	83.4	252.9
256.6	188.5	376.7	162.4	80.9	303.1
285.3	188.9	448.3	178.8	78.5	358.8
254.0	267.6	290.9	244.4	202.0	331.5
335.8	318.2	414.5	267.4	197.6	390.9
329.2	279.7	438.3	291.0	193.3	456.0
403.0	314.6	568.3	315.3	189.2	527.2
392.6	275.2	594.4	340.4	185.1	604.5
441.2	464.7	505.2	430.7	374.1	564.9
318.2	455.0	290.7	189.1	368.1	135.6
228.8	381.7	185.5	84.1	239.6	43.4
360.3	447.6	633.5	166.2	269.9	533.3
334.4	352.2	382.9	275.9	348.9	506.7
178.7	178.8	362.3	55.5	39.4	253.2
311.4	323.4	458.3	217.5	210.3	527.6
355.2	335.0	176.7	256.7	200.7	15.3
473.2	476.5	388.8	369.3	326.7	105.5

Table B.2: Tristimulus values for the 1931 standard observer for pairs of stimuli matched in brightness following the method described in Section 3.4.2. The white point used to design the background pattern and for CAM calculations has tristimulus values of 384.2, 400, 640.5.

Test Stimulus			Reference Stimulus		
X	Y	Z	X	Y	Z
89.08	92.26	146.4	39.7	21.27	34.61
130.8	136	215.7	106.8	70.82	114.8
197	152	249.4	134.3	67.55	111.5
218.9	130.4	217.2	150.4	61.82	102.3
234.6	245.6	389.2	212.2	157.9	256.3
289.2	232.6	380.3	259.5	153.1	250

327.8	211.2	347.7	289.5	143.8	239.1
350.4	183.6	300.4	337.8	139.4	239.6
297.6	309.7	494.4	367.3	293.6	470.6
444.2	368	603.6	433.5	276.1	458
419.1	277	466	483.8	269.8	451.3
80.65	84.45	134.4	37.29	21.95	3.888
177.6	186.1	295.4	98.15	72.03	40.52
179.1	148.9	142.9	126.3	71.06	9.118
187.2	195.5	312.4	197.9	161.3	140.4
285.6	247.6	263.4	239.8	156.4	60.08
286.6	205.8	116	264.9	149.6	24.93
331.3	198.2	50.03	306.7	145.1	1.191
350.8	364.8	582.9	343.3	296	294.1
397.7	350	394.5	403.5	292.3	165.6
385.6	289.1	190	439.5	279.7	97.11
112.3	116.5	186.9	30.4	22.34	0.368
101.3	105.5	166.6	82.72	73.76	18.04
174.9	175.4	188.8	97.47	73.58	2.144
215.8	225.7	358.5	174.9	162.9	96.36
219.9	222.2	252.5	186.3	161.7	30.82
239.9	232.3	172.2	199.6	158.7	12.9
248.6	224.4	101.8	209.6	158.9	2.492
285.5	297.2	473.7	310.4	300.7	232.6
330.3	335.8	398.3	330.1	295.3	125.3
341.8	333.3	277.6	340.2	291.8	71.43
355.7	328.4	183.5	356.6	293.6	42.28
383.5	338.5	118.8	389.8	295	4.606
105.7	109.8	173.8	42.7	44.41	1.724
131.5	136.6	216.6	68.79	75.96	10.66
150.2	162	149.8	73.65	73.96	1.405
138.8	144.6	228.4	105.2	116	26.99
136.2	146.8	134.8	113	115.6	3.657
165	170.5	274.7	150.5	165.1	61.9
189.4	207.4	202.3	156.1	167.1	15.04
202.2	220	126	162.2	167.3	3.167
202.4	210.9	340.7	206.9	227.7	114.1

231.8	248.8	255.8	213.3	231	38.82
275	299.6	200	216.7	229	5.72
264.3	276	438.9	279.3	305.8	185.2
315.2	338.7	371.9	283.4	301.1	85.47
305.1	334.8	222.7	284.3	302.2	33.8
339.7	372.5	160.2	293.9	303.7	5.764
90.78	93.92	149.4	14.55	24.03	0.5632
131.2	136.2	216.7	54.53	78.01	19.39
113	134.7	127.1	50.09	79.57	1.784
205.3	214.6	341.4	128.8	171.7	85.72
267.4	310.1	337.4	117.8	175.9	33.49
196.3	253.9	157.2	112.9	175.8	3.882
281.4	293	466.9	246.4	314.5	206
278.7	322.5	351.2	225.8	313.3	110.7
313.3	390	297.3	215	321.2	56.81
285.5	382.2	175.2	211.6	327.4	7.211
101	104.7	168.4	10.89	24.59	12.2
137.1	142.7	225.8	50.48	78.75	67.43
150.9	180.2	231.1	40.4	80.3	49.08
219.3	229.4	363.7	125.7	173.6	180.1
206.3	243.5	321.3	110.1	177.6	146.9
209.5	278	304.1	92.8	179	115.4
186.4	282.1	255.6	73.77	185.8	77.93
303.4	315.8	504	241.5	313.2	357.6
337.2	388.7	532.7	224.7	322.5	314.3
279.9	360.2	414.9	198.5	327.8	261.7
284.1	405.1	401.7	168.9	332.7	199.7
277.8	443.3	368	142.3	337.6	156.9
107.8	111.8	179.6	28.21	47.04	72.18
152.9	158.9	253.8	53.36	78.88	124.5
165.3	174.1	274.3	87.44	120.5	190.2
144	168.3	267.9	72.75	123.5	193.9
209.2	218.6	347.2	131.1	173.1	273.8
224.7	257.4	410.9	114.8	175.3	274.2
262.5	274.3	435.5	182.7	235.8	372.3
264.1	300.7	482.1	167.2	243.1	380.5

256.2	320	509.6	143.3	245.4	382.2
331.8	344.5	553	256	317.4	505.9
352.5	398.2	626.8	228.6	320.1	494.8
315.8	390.4	614.7	205.2	326.5	501.3
103.5	107.3	172.5	58.85	79.36	200.1
146.7	152.2	244.5	94.57	121.1	282.6
204.6	214	339.8	138.9	172.4	390.9
261.3	272.9	433.9	194.9	238.1	515.1
319.7	358.7	651	180.2	240.6	615.4
339.5	352.7	563.8	264.4	317.6	660.8
385.5	426	776.4	248.6	322.6	762.5
126.9	131.6	209.6	46.28	46.23	189.4
151.1	157.7	247.6	72.61	76.88	272
150.5	157	246.9	111	118.2	373.7
237.7	247.6	399.8	159.8	168.2	503
305.1	324.5	662.3	164.1	170	658.8
264.2	275.8	438.4	218.8	232.6	648
331.1	349.9	713.8	223.9	234.6	825.6
328	340.7	546.3	292.4	310.5	810.7
403.3	426.8	858.1	294.8	310.5	1014
86.34	89.11	142	28.17	21.74	113.7
96.02	94.1	217.7	35.34	21.8	171.4
126	130.6	208.1	82.28	72.65	251.6
166.2	165.8	358.5	91.71	73.63	347.1
176.6	166.4	473.6	103.9	72.39	462.3
199.1	208.3	331.7	176.6	166.2	473.3
209.2	209.9	440.3	188.4	165.6	605.2
241.5	231.4	622.7	202.9	163.7	756.7
272.6	246.3	811.2	222.8	163.7	942.6
256.3	214.8	905	240.1	160.2	1140
314.7	326.8	523.7	311.5	300.7	776.3
309.9	311.9	635.3	328.9	302.1	953.8
335.3	324.6	827.4	339.7	296.4	1111
108.8	112.9	180.9	31.28	21.2	97.09
114.4	94.5	282.9	45.58	19.23	194.4
160.6	166.4	267.5	88.99	72.11	230.4

142.2	121.1	340.9	113.1	69.27	383.6
133.9	86.11	435.8	143.1	66.25	579.8
179.1	91.64	690.4	181.8	61.72	846
205.3	214.7	341.1	186.6	162.5	434.1
206.9	180.9	472.5	220.5	156.6	653.3
230.8	167.3	682.6	260.1	152.3	916.6
288.4	300.5	478.5	331.2	299.3	723.5
373.6	338.6	806.9	377.5	291.6	1013
89.4	92.18	148.7	34.37	19.98	75.51
121.8	91.19	235.8	49.85	18.59	132.9
151.2	157	251.6	96.85	70.25	193.3
204.3	162.8	390	122.6	64.62	285.6
156.4	87.65	348.9	157.3	62.01	407.7
206.1	215.4	342.6	201.7	160.3	381.9
248	202.4	472.7	242.9	150.7	519.5
285.9	184.9	608.3	292.2	145.1	691.9
329.3	169.7	768.1	346	139.4	899
337.4	350.3	561.8	348.9	292.5	645.4
405.8	343.7	747.2	410.8	284.1	837.4
413.9	285.8	849.4	473.4	270.5	1056

Table B.3: Tristimulus values for the 1931 standard observer for pairs of stimuli matched in brightness following the method described in Section 3.4.3. The white point used to design the background pattern and for CAM calculations has tristimulus values of 384.2, 400, 640.5.

Test Stimulus			Reference Stimulus		
X	Y	Z	X	Y	Z
153	158.7	254.2	97.47	73.58	2.144
169.8	175.2	282.4	186.3	161.7	30.82
283.9	295.9	470	199.6	158.7	12.9
294.5	306.7	487.6	209.6	158.9	2.492
236.7	247.1	391.5	330.1	295.3	125.3
352.6	366.4	585.4	340.2	291.8	71.43
488.9	507.4	817.3	356.6	293.6	42.28
439.5	456.1	726.3	389.8	295	4.606
129.8	134.5	213.5	73.65	73.96	1.405

139.6	145.3	229.2	113	115.6	3.657
244.5	255.8	405.1	156.1	167.1	15.04
176.3	184.9	292.4	162.2	167.3	3.167
194.6	203.2	322.9	213.3	231	38.82
279.8	291.7	463.8	216.7	229	5.72
248.6	260	411.9	283.4	301.1	85.47
306.3	318.8	508.3	284.3	302.2	33.8
312	324.5	518.8	293.9	303.7	5.764
91.85	94.88	150.7	40.4	80.3	49.08
243.1	254.2	402.9	110.1	177.6	146.9
300.3	312.7	498.3	92.8	179	115.4
253.9	265.3	420.8	73.77	185.8	77.93
370.3	386.4	616	224.7	322.5	314.3
328.4	340.7	546.4	198.5	327.8	261.7
424.7	441.5	703.1	168.9	332.7	199.7
444.6	462.6	740.4	142.3	337.6	156.9
76.3	79.98	126.5	35.34	21.8	171.4
194	202.5	322.2	91.71	73.63	347.1
169	174.6	280.3	103.9	72.39	462.3
313.6	326.2	520.3	188.4	165.6	605.2
244.6	254.6	411.2	202.9	163.7	756.7
337.3	350.5	560.6	222.8	163.7	942.6
334	346.9	555.6	240.1	160.2	1140
308.1	320.7	512.1	328.9	302.1	953.8
382.7	398.1	634.1	339.7	296.4	1111
261.1	262.9	301.6	199.6	158.7	12.9
341.2	346.8	410.5	209.6	158.9	2.492
396.3	402	484.5	340.2	291.8	71.43
318.3	321.4	377.7	356.6	293.6	42.28
511.7	519.4	639.5	389.8	295	4.606
248.2	268.3	275.4	162.2	167.3	3.167
246.8	266.9	274.1	216.7	229	5.72
292.8	317.5	334.3	284.3	302.2	33.8
278.1	299.7	311	293.9	303.7	5.764
261.8	300.7	414.2	92.8	179	115.4
333.7	387.8	517.2	73.77	185.8	77.93

397.9	453.3	628.8	198.5	327.8	261.7
371.8	424.4	587.3	168.9	332.7	199.7
432.7	491	683.5	142.3	337.6	156.9
182.3	182.4	387.3	103.9	72.39	462.3
311.5	313.3	638.1	202.9	163.7	756.7
310.7	312.5	636.5	222.8	163.7	942.6
432.9	436.4	875.9	240.1	160.2	1140
402.6	407	813.9	339.7	296.4	1111
325.2	315.1	250.1	209.6	158.9	2.492
355.4	347.5	275.7	356.6	293.6	42.28
474.3	466.1	421.9	389.8	295	4.606
367.7	405.7	276.6	293.9	303.7	5.764
242.6	314.5	359.6	73.77	185.8	77.93
353.9	444.5	526.8	168.9	332.7	199.7
413.5	509.3	617.5	142.3	337.6	156.9
350.6	338.9	863.8	222.8	163.7	942.6
374	360.5	903.4	240.1	160.2	1140
409.7	380.9	205.9	389.8	295	4.606
296.6	410.7	423	142.3	337.6	156.9
342.9	317.5	978.9	240.1	160.2	1140

Illuminant A			D65		
X	Y	Z	X	Y	Z
47.78	29.57	7.49	35.56	28.61	22.58
47.62	29.74	7.50	39.37	31.23	21.04
31.73	26.23	3.99	31.30	29.56	12.03
34.07	29.66	4.02	32.40	32.23	12.07
25.12	29.92	10.11	27.99	35.13	26.68
24.65	29.53	10.49	26.75	34.30	27.12
25.14	27.38	20.59	27.75	33.62	60.29
25.20	27.00	20.53	27.69	31.57	60.22

Table B.4: Relative tristimulus values for the CIE 2012 10° observer for pairs of stimuli matched in hue following the method described in Section 4.3. All Illuminant-A-colored illumination had relative tristimulus values of 111.77, 100, 31.12. All D65-colored illumination had relative tristimulus values of 94.78, 100, 107.61. The luminance of white was approximately 200 cd/m^2, thus for absolute values, these values should be multiplied by 2.