ABSTRACT

HASLUP, JENNIFER REID CLONTS. Perception of Blackness and the Foundations of a Blackness Index. (Under the direction of Drs. Renzo Shamey and David Hinks.)

Many researchers have investigated various aspects of blackness, but a detailed examination of the perception of blackness has not been reported to date. This study aims to generate a thorough review of blackness from theoretical and perceptual perspectives. Since the quantification of blackness, due to relatively large signal to noise ratio, is specifically subject to errors and has not been sufficiently elucidated, this work also aims to examine the measurement of blackness and determine inter- and intra-instrument agreement. Finally, perceptual assessments are to be analyzed to propose a blackness index that sufficiently correlates with measured colorimetric values and visual perceptions.

Experimentally, this research covers three stages. In the first, twenty samples of uniform chroma and lightness with varying hues were ranked in order of perceived blackness. In the second stage, thirty samples were produced that varied in hue and chroma, but remained of approximately uniform lightness. Observers first categorized these samples as black or not black, and then rated them in relation to an “ideal” black, which was essentially a black velvet light trap. The third stage assessed the measurement of near-black objects and determined the repeatability and reproducibility of such measurements.

From the observation trials, it was determined that samples with hue angles between 200° and 270° are considered blacker than samples with hue angles in other regions of color space. Samples with hue angles above 315° or below 45° are considered less black than other samples. The effect of chroma is much less significant than that of hue angle, but samples with lower chromas are considered to be blacker than samples with higher chromas. The measurement of near-black objects is subject to more variation than that of lighter samples.

The Perception of Blackness and the Foundations of a Blackness Index

by
Jennifer Reid Clonts Haslup

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Fiber and Polymer Science

Raleigh, North Carolina

2011

APPROVED BY:

Renzo Shamey Committee Chair	David Hinks Committee Co-Chair
Peter Bloomfield	Brent Smith

DEDICATION

To my mother, for without her encouragement and help, graduate school would have been surely impossible.

To my husband, who has seen me through to the end of this journey.

BIOGRAPHY

J. Reid Clonts Haslup grew up in Lexington, South Carolina. She graduated from the South Carolina Governor's School for Science and Math in 2000, and received a BS in chemistry and a BS in textile chemistry in 2004 from North Carolina State University. Her main research interest is in color science, and she enjoys teaching chemistry.

ACKNOWLEDGEMENTS

Without the help of the following people and organizations, this research could not have been completed:

- All of my observers
- Dr Shamey, Dr Hinks, Dr Smith, and Dr Bloomfield
- The Department of Textile Engineering, Chemistry, and Science
- Judy Elson and Jeff Krauss
- Janet Clonts
- Christopher Haslup
- Lina Cardenas, Rebecca Klossner, David Padgett

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. Introduction

In 1980, W.D. Wright highlighted the lack of research into blackness in an article published in *Die Farbe* (Wright, 1980). At that time, some work was beginning on the induction of blackness (Shinomori, *et al.*, 1994; Volbrecht, *et al.*, 1989a, 1989b; Volbrecht, *et al.*, 1990; Werner, *et al.*,1984), while a group of German scientists were formulating the beginnings of a blackness index for use with carbon black pigments (Bode, *et al.*, 1979; Ferch, *et al.*, 1981; Gläser, 1978; Lippok-Lohmer, 1986). The induction studies have proven to be useful in indicating potential neurophysiological mechanisms for the perception of blackness. The early German indices presented a foundation for the creation of a general blackness index, but are not applicable to systems other than soot-based pigments in inks and print pastes. All of these works were fairly limited in scope, while it was our intention to take a broader view of blackness.

There is a noticeable lack of published research into the perception of blackness. For this reason, and since black materials are major component of many commercial products, it was decided to undertake a systematic investigation into the subjective assessment of blackness and measurement variability of black materials. The ultimate goal of this work was to formulate the foundations of a blackness index to quantify the perception of blackness. Initial experiments focused on identifying hue preference in the perception of blackness, while later experiments provided data for the generation and testing of an index. Furthermore, a comprehensive measurement study was conducted to determine the inherent variability in the measurements of near-black objects.

Currently a significant amount of effort is being extended in the color science community to approximate Ewald Hering's suggested generic unique hues—that is, a red and a green that perceptually contain no blue or yellow, and a blue and yellow that contain no green or red. These four hues are considered to be elementary, and are necessary to describe other hues. Hering, amongst many others, also considered white and black as elementary colors.

Thus it might be expected that, when asked, white and black would be described as neutral and contain no hints of the four chromatic elementary hues. However, research on whiteness has shown that a slight blue tinge causes a white sample to be perceived as whiter than a sample of identical lightness that is measured to be perfectly neutral. Part of this work endeavored to identify whether slight blueness or indeed the presence of other hues also caused an increased perception of blackness.

There have been multiple models proposed for the neurological basis of blackness. Early theories assumed that the achromatic perceptions arise from the neurological summation of the signals from three cone types, and that blackness was the perceptual opposite of whiteness. In contemporary models, it is assumed that summing the cone responses only provides lightness information, and that a comparison of logarithmic transformations on the responses is required to obtain a blackness perception. Such models require contrast for the perception of blackness. The fact that, in the absence of contrast, a black field can appear grey as shown in the Eigengrau phenomenon—the perception of a grey color when one's eyes are closed in a dark room—supports this theory. These models are presented in the article in Section 2 in greater detail.

The article in Section 2 comprises the bulk of the literature review for this work. Submitted to *Vision Research* for publication, *Perception and Quantification of Blackness: A Review* covers all of the work on blackness that can currently be found. Reviews of published work on whiteness and current practices in coloration are found in Appendices A and B, respectively. While these topics are certainly relevant to the current work as a whole, they were deemed beyond the scope of the published article. Whiteness would have deserved a review article all its own. The discussion of current practices in coloration covers practical textile problems too specific for publication in that journal. However, it is most helpful to review the formulation of the whiteness indices when considering the formulation of a similar blackness index. The current practices in coloration give a starting point for sample generation which was a major portion of this work.

The article in Section 3, *Effect of Colorimetric Attributes on Perceived Blackness of Materials*, provides a brief overview of the observation trials completed for this work. It was presented as a poster at the 5th European Conference on Colour in Graphics, Imaging and Vision, CGIV2010/MCS10, in Joensuu, Finland. This article covers data from both the first and second observation trials.

The article in Section 4, *Effect of Hue on Perceived Blackness of Materials*, discusses the first observation trial in greater detail. It will be submitted to *Color Research and*

Application. The first observation trials isolated the potential role of hue in the perception of blackness. Samples with uniform lightness and chroma were presented to observers. The data collected was used to determine if black samples containing small amounts of one hue (or a range of hues) are perceived as blacker than other hues (or ranges).

The article in Section 5, *Effects of Hue and Chroma on Perceived Blackness of* Materials, details the second observation trial. It will be submitted to the *Journal of the Optical Society of America*. The second observation trials included the role of change in chroma on perceived blackness. By controlling lightness, the contributions of hue and chroma to the perception of blackness were elucidated.

Together, the three articles in Sections 3, 4, and 5 provide the analysis for experiments covering the contributions of hue and chroma to the perception of blackness. To complete the observation trials, appropriate sets of samples had to be generated. The samples for the trials were prepared by coating paper (for the first trial) or dyeing uniformly textured fabrics (for the second trial). The coated paper samples were obtained from Munsell, and the knit acrylic samples were dyed in-house Sample preparation (either generation or collection), which is not a trivial task, was a significant component of this work. A brief overview of failed sample preparation attempts may be found in Appendix C. Many of the samples from the failed generation or collection attempts may be useful in future works.

The final article in the series, *Assessment of inter- and intra-instrument variability in the measurement of very dark surfaces*, will be submitted to *Coloration Technology*, and pertains to the measurement of near-black objects. Despite several decades of work and systematic study on color measurement, the measurement of very dark objects is subject to significant error. With any analytical instrument, the signal-to-noise ratio (S/N) is used to compare the desired signal to the background noise inherent in a measurement. When dealing with spectrophotometers, the signal is the light reflected from or transmitted through the object being measured, while the noise arises from stray light or absorption due to defects in the sphere coating. As any signal gets weaker, the S/N ratio becomes more significant, and if the signal is weak enough, the S/N ratio approaches unity, and measurements become highly unreliable. In spectrophotometers, the signal weakens as the object becomes less reflective; that is, darker. For this reason, it is important to determine the inter- and intra- instrument agreements in the assessment of dark objects, typified by blacks, and to examine repeatability of measurements. The usefulness of any blackness index would thus be highly dependent on reliability of measured values. Any formula will only be as trustworthy as its inputs. Due to increasing significance of global sourcing in manufacturing colored goods and the need to accurately communicate color, and since black is one of the most widely produced colors, accuracy in determining the degree of blackness of colorants, inks and finished colored goods is of primary importance.

Section 7 provides a summary of the conclusions and achievements from this work, and Section 8 gives some suggestions for future work that would be appropriate based on those conclusions.

2. Perception and Quantification of Blackness

R. Shamey* and J. R. Clonts Haslup
Polymer and Color Chemistry Program, North Carolina State University, Raleigh, NC, 27695-8301, USA
Corresponding author: rshamey@ncsu.edu, T +1 919 515 6546, F +1 919 515 6532

Abstract

Blackness is a fundamental concept in vision and color science along with whiteness; yet very little has been published on the topic. Blackness has long been considered to be the opposite of whiteness and is more difficult to study than whiteness. This review examines the historical discussions surrounding the subject matter and highlights an examination of studies aimed at elucidating the physiological as well as psychological aspects of blackness. Studies based on temporal and spatial induction of blackness are briefly described and methods employed to quantify blackness for use in paint industry are explored. Recent works aimed at development of methodology to quantify blackness from a perceptual perspective are also briefly examined.

Keywords: Blackness, induced blackness, blackness index, physiological basis of blackness.

1. Introduction

Blackness has surprisingly been largely ignored as a research topic in color science. Blackness in output devices such as monitors and projectors, and several studies on the induction of blackness, based on certain surround conditions, (Cicerone, *et al.*, 1986; Fuld, *et al.*, 1986; Shinomori, *et al.*, 1994, 1997; Volbrecht, *et al.*, 1989a, 1989b, 1990; Werner, *et al.*, 1984), constitute the main body of the reported literature. However, very few published works pertaining to determination of blackness in objects can be found. Specific examples of practical examinations of blackness include the determination of blackness for pencil lead (Kenya Bureau of Standards, 2005) and black pigments (Bode, *et el.*, 1979; Gläser, 1978; Lippok-Lohmer, 1986; Schumacher, 1976). This review examines blackness from an historical perspective and provides the reader with a summary of efforts aimed at development of an index to specify the magnitude of blackness.

2. Early Discussions of Blackness

At the turn of the twentieth century, the study of color vision and perception was still in its infancy. Theories about the perception of color abounded; two of the well-known theories came from Ewald Hering and Helmholtz. Black was not studied in isolation, however, and was considered within the framework of the models.

Hering first published his *Outlines of a Theory of the Light Sense* in 1878, arguing that black was a sensation similar to other colors and that surfaces reflecting a significant amount of light could still appear black (Hering, 1964). He believed that black could only be perceived if part of the visual field appeared white or grey and that without strong contrast only dark greys could be seen. Hering also argued that blackness and whiteness were two distinct properties of a color, and that a perfect medium grey was as similar to black as it was to white, although it was its own separate color. He advocated defining achromatic colors using the proportion of their similarities to white and black (W/B or B/W) while maintaining that this scale was only theoretical, as a perfect white and perfect black are unobtainable. Hering espoused a black–white opponency while maintaining that intermediate black–white colors (greys) also existed.

2.1. Works of Ladd-Franklin, Titchener and Ward

Some of the earliest published work on blackness came from a somewhat philosophical standpoint. Ladd-Franklin published her theory in 1892, and recognized that the sensation of black could be due to the cortical visual processes reaching a resting state (Ladd-Franklin, 1932). She, however, argued that the sensation of black was a constant background sensation, and not a distinct sensation caused by light. This theory treated black separately from other visual sensations, based on the argument that black arises when no light is present. It is impossible to vary the intensity of black, it was stated, and thus black should be treated differently from white, red, blue, yellow or green. Ladd-Franklin disagreed with Hering's theory that blackness increases as whiteness decreases and preferred to consider black in isolation.

In his 1905 paper "Is 'Black' a Sensation?", James Ward discussed positive and negative sensations and perceptions. Eleven years later, E. B. Titchener (1916a) responded with "A Note on the Sensory Character of Black", prompting a response from Ward (1916) which was followed by a final note by Titchener (1916b). A strict separation between psychology and physiology was maintained throughout the exchange.

Ward stated that “to the unbiased mind, black is positive, *i.e.* for perception, and darkness negative, *i.e.* for sensation; but the unbiased mind would prefer to say not that black is a lower degree of the sensation of white, but that dark is a lower degree of the sensation of light” (Ward, 1905). He also stated that “the so–called sensation of the deepest black [Hering] has shown to be possible only under the influence of light” (Ward, 1905). Ward thus believed darkness to be the sensation arising when little or no light reached the retina; whereas blackness could only occur when light was striking the retina.

In 1916, Titchener (1916a) focused on the psychological aspects of blackness, ignoring the visual theories discussed in Ward’s 1905 paper. Titchener disagreed with the assumption that darkness and silence are essentially the same, and argued that the opposite of a blinding light is an impenetrable dark. He pointed out that while one can only hear a deep silence in the context of a noise; deep darkness can be seen without an adjacent bright light. He also maintained that black and white may be completely independent, as in the case of a black or white figure on a grey background.

Ward (1916) stated that “Experience cannot begin with negation and vision cannot begin with black,” for black is only sensed under the indirect action of light. Therefore he considered black to be a negative sensation that must be preceded by a positive sensation of light. Ward differentiated between black and darkness—the latter being an unlit region, while the former necessarily existed only in a lit area. He argued that black and white are as distinct as “land and water, island and lake,” (Ward, 1916) and that black is the opposite of white, not merely the absence of white. Ward agreed that there must be a visual stimulus for black, but he was unsatisfied with the two available theories—Helmholtz and Ladd-Franklin's idea that black corresponded to the resting state of the eye, and Hering's idea that black was seen during a relaxation of the eye. The analogy between silence and blackness was made as an argument against Helmholtz, Ward stated, and Titchener had interpreted it incorrectly to be an argument against Hering. Ward further discussed his statement (and Titchener's response to it) that "in a light–field many colours may be distinguished, but in a dark field none" (Ward, 1916). Ward claimed that there was only one true black and myriad darks (that is, the darkest dark is the only black).

Titchener (1916b) described a daylight scene where the eye first notices a patch of burnt grass or the entrance to a cave. In such a scene both black and white are present, but the visual experience starts with black rather arbitrarily, as it is the first color noticed. Titchener elaborated on the differences between "darks" and "blacks" and argued that the distinction is subtle where darks give a three–dimensional feeling, while blacks are superficial. He pointed out that the psychophysical processes corresponding to blackness were not understood. Titchener further disagreed with Ward on his analogy of blackness to silence. Silence, he argued, is not solely auditory, but may be felt or even seen while blackness may only be seen. In reference to 'blinding white' and 'impenetrable black' descriptions he stated that while a brown may be heavy, a yellow is similarly light, and while a high pitch is piercing, a low one may be rumbling. He maintained that the empirical facts did not back up Ward's view, however, he agreed with Ward that seeing black and seeing nothing are different.

2.2. Works of Neifeld, Müller and Michaels

In 1924, Morris Neifeld discussed "The Ladd-Franklin Theory of the Black Sensation". In the years between Titchener's last note and Neifeld's article, black was regarded as a sensation that corresponded to the resting state of the visual system in the brain. Greys, however, were still not well understood. Various theories concerning grays are schematically illustrated in Figure 2.1. Hering's original theory claimed that each grey was a mixture of variable amounts of white and black. This is illustrated by the diagonal line between black and white in Figure 2.1. Müller (Müller, 1896, and 1897) added a "constant neutral grey" (denoted by the square with equal amounts of black and white) to the variable amounts of white and black to account for processes occurring in the cortex (Hering only considered retinal processes). The Ladd-Franklin theory ignored the variable amount of black in each grey and considered that black was a background sensation to which white could be added in variable amounts to make greys. Neifeld defined black as "the psychic correlate of the absorption, by objects, of all the visible light–ray frequencies" (Neifeld, 1924). It was stated that, because of the Eigengrau, absolute black could only be seen in the context of spatial or temporal contrast with a bright light. This indicated a shift from the attitudes of Ward (1905, 1916) and Titchener (1916a, 1916b) where black was regarded as a sensation that could be empirically studied.

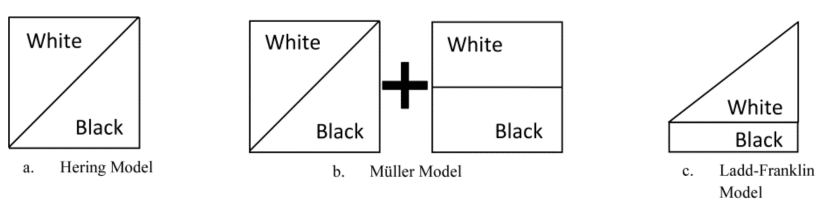

Figure 2.1. Schematic representation of various theories on the mixing of black and white (Neifeld, 1924)

Neifeld (1924) favored the Ladd-Franklin theory not only because it elegantly explained the retinal and cortical processes, but also because it explained why the black sensation occurs. He argued that hearing silence and hearing nothing are practically the same sensation, because we are not able to accurately locate the source of a sound. Visually, however, our spatial awareness is very good. The difference between seeing black and seeing nothing, then, has more to do with the retinal area activated than the actual light reflected from an object.

In 1925, George Michaels (1925) explained that shining a bright light on a black object causes the object to appear dark grey—empirical evidence that greys are a mixture of a single intensity of black and a variable intensity of white. He supported Neifeld's statement that spatial or temporal contrast is required to obtain a perfect black. He again stated that only one intensity of black was possible; any change in the intensity of black arising from a change in the white component of a given color, and a black that is not as intense as the true black is simply grey.

2.3. Works of Dimmick, Rich, and Judd

Forrest Dimmick studied the nature of the achromatic scale (1920, 1929, 1933, 1941). In 1920, he proposed that the black–white scale was analogous to the red–green and yellow–blue scales (Dimmick, 1920). That is, each of these scales consist of two separate scales, each ending with a medium grey, joined 180° apart. He argued that just as it was impossible to go from red to green without passing through a neutral grey point, it was impossible to go from black to white without passing through the same neutral point. To validate this theory, he had observers add small amounts of red, yellow, blue, green, black or white to a constant medium grey and note whether the mixtures showed any trace of the added color. The observers were all naïve, and the results for each of the six colors were very consistent, seemingly verifying that the black–white series is not unique.

Gilbert Rich took up the question of blackness in 1926, responding predominantly to Neifeld's arguments (Rich, 1926). He agreed with Ladd-Franklin and Neifeld that black was the result of retinal inactivity (that is, no light striking the retina), in contrast to Hering and Müller who thought a neutral grey would result. He recognized the difficulty in making black the neutral point of the black–white unidirectional series, however—black and white cannot add together and cancel each other out similar to other opponent colors. He also pointed out that if black and white were truly opposite colors, adaptation to one over a long period would result in seeing the other. However, adaptation to either eventually results in seeing a neutral grey. As such the Hering theory seemed to provide more answers than the Ladd-Franklin theory. Rich also objected to Dimmick's experiment and provided some experimental data to support his hypothesis (Rich, 1926). Rich felt that grey was no more the midpoint of a continuous black–white scale than it was the midpoint of a continuous blue–yellow scale. He stated, "grey, not black, is the center from which visual experience radiates, so to speak." Rich seemingly assumed that grey was an elemental color. He suggested that rods, being primitive, only provide information about the brightness of greys, and that the cones are responsible for not only color vision, but for the differentiation of black and white.

Dimmick undertook an experiment to test Rich's hypothesis regarding gray as an elemental color (Dimmick, 1929). Observers were shown mixtures of test lights and monochromatic colors (standards) simultaneously. They were asked to determine if the two samples had any resemblance to each other. In this first trial, observers designated colors as red, blue, yellow, orange, purple, and green, without using black, white, or grey as descriptors. In the second trial, observers were shown two full color standards and an intermediate variable color (between the two shown standards) simultaneously. The samples were of varying degrees of lightness. A third trial was performed where observers were shown a variable color that was not between the two full colors being tested. Dimmick found that while orange, purple and green were used to describe colors early in his experiments, by the third trial, observers were content to use red, yellow and blue for hues, and they separated grey from black and white. Because more well–trained observers used "grey" to describe colors, Dimmick argued that grey was an elemental color. In 1933, he reported an experiment wherein observers were asked to determine if colors were similar to black, white, or neither. Observers were shown neutral samples of varying lightness, and a "neutral grey" was found (Dimmick, 1933).

Deane Judd argued that surface colors and aperture colors should be distinguished from each other, and that black and white should apply to surfaces while dim and bright should apply to apertures (Judd, 1940). In 1941, Dimmick responded by quoting Titchener's 1916 response to Ward, arguing that there were no psychological reasons to separate black and dark (Dimmick, 1941). He pointed out that there was likewise no reason to separate white from bright, and thus the "achromatic scale" was truly comprised of two scales both anchored in grey. Dimmick also argued that all colors could be reduced to "areal colors", regardless of the appearance mode of the original color. He emphasized that while a black velvet substrate gives the best black surface, the color seen through an opening into a completely dark chamber would be even blacker.

In a subsequent work Judd defined black and white to be extreme achromatic surface colors characterized by being completely opaque and matte, differing only in lightness (Judd, 1941). This concept was expressed using Equation 1:

$$L = \frac{A(A_t+1)}{(A_t+A)}, A_t > 0 \tag{1}$$

where A is the luminous apparent reflectance of the surface in question, A_t is the luminous apparent reflectance of the scene, and L is the defined lightness term. Equation 1 regarded only surfaces with no apparent luminous reflectance to be black, surfaces with perfect reflectance to be white, and everything in between to be a matte surface. This equation could be interpreted to support either Hering's or Ladd-Franklin's theories on blackness, depending on whether the reflectance of the surface in question is set equal to the reflectance of the scene or zero, but Judd was quick to state that since the equation only applied to photopic vision, and because of the relationship between the reflectance of the surface and the scene, the model reinforced neither of the previous theories. He recognized that surfaces of intermediate lightness values (that is, a surface with imperfect reflectance) might be called black or white in the absence of surfaces that perfectly reflect or absorb light. In Equation 2, this concept was taken into account:

$$L = \frac{(A - A_n)(A_t + A_x)}{(A_x - A_n)(A_t + A)}, A_t > 0, 0 < A_x \leq 1, A_n < kA_x \qquad (2)$$

where A_n and A_x are the minimum and maximum apparent reflectance in the scene, k is a constant for each observer, and A_x is sufficiently larger than A_n. Judd argued that while observers tend to perceive scenes as being uniformly illuminated this equation also held in many cases of unequal illumination. He noted that in some cases objects present in the scene can change the observer's perception, but that equation 2 was not broad enough to include such cases.

Judd examined the appearance of non-surface achromatic colors and concluded that volumic blacks and whites were rare, but existed; aperture blacks and whites could be seen by trained observers; and luminous whites were possible while luminous blacks were self-contradictory. He then suggested that the appearance of blacks was based on the surface mode, and that this was in agreement with the works of Helmholtz, Troland (1921), and Neifeld. Thus he restated his original definition of black and white as surface colors, at least until further experimentation proved otherwise (Judd, 1941).

2.4. More Recent Discussions

In 1980, W. D. Wright wrote a short article in *Die Farbe* for the occasion of Richter's 75th birthday, in which he discussed the perception of blackness (Wright, 1980). He pointed out that the separation of television color signals into a luminance channel and a chrominance channel, and their subsequent combination, would result in the regeneration of the original signal. R.W.G. Hunt separated several photographs in this manner, preparing a black-and-white slide and a chromatic slide of uniform luminance of the image. Hunt and Wright noticed that combining the slides gave a reproduction of the original image, while the colors in the chromatic slide appeared fantastical. Wright discovered that the chromaticity was unchanged in the chromatic slide, and that the contrast provided by the luminance slide was critical to seeing the original image. Projecting the chromatic slide onto a uniform dark grey background did not restore the colors seen in the original image. From this, Wright drew a few conclusions. His first pertained to neural coding: current thinking at the time assumed that the black-white opponent channel was coded from the sum of the three types of cones in the retina. This idea led to the conclusion that blackness and luminance would be coded simultaneously, but his experiment showed that luminance must be separated from blackness at some point. His second conclusion was that the luminance alone was insufficient for color reproduction, and contrast was crucial. Finally, he stated: "This raises the question of how we should measure or specify blackness. We might, perhaps, use the black-content scale of the Ostwald charts, or the black-white scale of the Natural Colour System, or the darkness-degree scale of the DIN System. What we do realize is that the blackness we are interested in is a subjective perception and not something that can be measured on any simple photometric scale. Moreover, the depth of the blackness that we perceive can be affected by quite small areas of contrasting lightness, for example, by the small highlights on the black grapes of [one of] Dr. Hunt's [separated images]" (Wright, 1980). Using luminance and chrominance slides of "The Arnolfini Marriage", supplied by Hunt and Ward, he reaffirmed that the blackness element in a color occurs through contrast with nearby lighter colors and that blackness is a positive perception and not just the absence of any luminous sensation (Wright, 1981). He also stated that many formulae for the specification of whiteness were reported, while no comparable formula for blackness existed.

Quinn, Wooten, and Ludman revived the issue of achromatic colors in 1985 (Quinn, *et al.*, 1985). They asked whether black and white alone were sufficient to describe all achromatic colors, or if grey was elemental (that is, necessary for describing the entire achromatic spectrum). In an experiment observers were shown a series of stimuli and asked to assign percentages of colors to them. The observers were restricted to certain color names for each session, and were told that the percentages did not have to add up to 100%. A response function for the colors that were not allowed to be mentioned was then calculated. In the first session, observers were asked to use only "white" and "black", and the calculated response curves for "grey" indicated that the achromatic series could be fully described using only those two words. In the second and third sessions, observers were only allowed to give the percentage of black and white in each sample, respectively. It was determined that the responses from the third session (percent white), the percent white responses from the first session, and the calculated white curves from the second session matched fairly well.

Similarly, the percent black responses from sessions one and two matched the calculated responses from session three well. In the fourth session, observers were allowed to use “grey”, and a final fifth session was held where observers could again use only “white” and “black”. The results from the fifth session matched the results from the first session, indicating that even after observers were asked to use grey as an elemental color, it was not needed. Thus they determined grey as a mixture of black and white, which should not be treated as a separate elemental color. Their results also indicated black and white as the two ends of one continuous scale, which are not opponents, unlike, for instance, blue and yellow. This was a departure from previous thinking, which held black and white as endpoints of two separate scales.

3. Induced Blackness

World War II seems to have been a turning point in the research of blackness. Up to this point, much of the discussion focused on the nature of the perception of blackness, and its relationship with whiteness. These topics, however, largely fell out of favor with researchers. Meanwhile, technological advances in the following years made the study of color on devices such as monitors and projection screens more popular. Significant effort was devoted to inducing the perception of blackness through simultaneous spatial or temporal contrast.

In 1983, Werner, Cicerone, Kliegl, and DellaRosa endeavored to determine the spectral efficiency of blackness induction (Werner, *et al.*, 1983). They assumed that blackness would be the inverse of the spectral response function of whiteness (the prevailing assumption at the time), but this approach did not provide a complete answer. Several spectral sensitivity functions had been found, depending on the methodologies employed (e.g., flicker, brightness, contrast, etc.). It was hypothesized that either a function based on flicker photometry or one based on brightness would most closely match the actual achromatic spectral response function (Werner, *et al.*, 1983). In a study, observers were asked to adjust the brightness of a monochromatic ring of light until a broadband white spot in the center of the ring just appeared to turn black. In a second trial, heterochromatic flicker photometry was used with and without the central spot. The results from both trials were found to match. It was determined that the blackness induction function was similar to the CIE spectral-luminosity curve at medium to long wavelengths. The variation at short wavelengths was attributed to scattered light. They determined that the blackness function is not dependent on the hue of the inducing light, and is essentially the inverse of the luminance of a stimulus.

In 1986, Cicerone, *et al.*, restated that black can only be seen in a context of spatial or temporal contrast, and that although black and white are opponent colors (similar to the opponency of red and green, or blue and yellow), the nature of their opponency is different from that of the chromatic opponent pairs (Cicerone, *et al.*, 1986). This was in agreement with the findings of Quinn, Wooten, and Ludman (Quinn, *et al.*, 1985). Their experiment involved three color normal female observers who were asked to adjust the luminance of an annulus of varying wavelength until a central spot of varying wavelength appeared black (Werner, *et al.*, 1983). These results also indicated that the wavelength of the inducing annulus (or the test spot) is insignificant, and that blackness arises as a function of luminance. Thus they stated that the achromatic neural pathway likely functions parallel to the chromatic neural pathways.

Fuld, Otto, and Slade took a slightly different approach to the question of the spectral response functions for white and black in 1986 (Fuld, *et al.*, 1986). Using an annulus–central spot technique, they looked for the intensity of the annulus required to make the wavelength variable central spot appear to contain equal amounts of black and white. Observers were asked to adjust the luminance of the central spot until they found the limits between "predominantly black" and "equally white and black," as well as between "equally white and black" and "predominantly white." Observers were found to be unable to adjust the spot until the color contained no white or black, indicating that the achromatic neural channels are not entirely analogous to the chromatic pathways. In a second experiment, observers were asked to adjust the luminance of the central spot until it was equal in brightness to the annulus, and results were found to be similar to those from the first experiment. A third experiment was essentially a repetition of the first, but the annulus and spot were matched in brightness, not luminance. A fourth experiment involved an annulus of fixed luminance. The results of all these experiments agreed with each other, and suggested a link between the perception of equal whiteness/blackness and brightness.

In 1989, Kulp and Fuld reported a study of black spectral responsivity, in an attempt to reconcile his findings with those of Werner and Cicerone (Cicerone, *et al*., 1986; Kulp and Fuld, 1989; Werner *et al*., 1984). This experiment involved six subjects, and compared their blackness induction curves to their heterochromatic flicker photometry and heterochromatic brightness matching functions. Subjects were instructed to use either a contour-disappearance or an absolute blackness criterion, and the effects of the instructions were also analyzed. They found that the instruction criterion did not predict the fit of the blackness induction curve, but the results were otherwise inconclusive. They did not find strong evidence for either flicker photometry or brightness matching functions to represent the blackness spectral responsivity.

In 1989, Volbrecht, Werner and Wooten investigated the temporal induction of blackness (Volbrecht, *et al.*, 1989a). They reasoned that their results and those from Fuld were not contradictory, although they found a correlation between blackness and flicker photometry functions, while Fuld correlated whiteness/blackness to brightness (Fuld, 1986). If one assumes that whiteness and blackness are not strict opponents, but are based on multiple processes (opponency and a spectrum based scale), then these two results are reconcilable. For this experiment, observers were shown an inducing stimulus (their unique blue, green, yellow, as well as white) followed by a reference stimulus (another of their unique hues). They were then asked to describe the colors seen during the reference flash using percentages of black, white, red, blue, and yellow. At low inducing field luminances, observers reported chromatic induction. At higher inducing field luminances, however, they described the colors using only black and white. The luminance at which this first occurred was independent of the hue of the inducing field. These findings indicated that the spectral response curve for temporally-induced blackness is related to the flicker photometry response curve.

Volbrecht and Werner followed up the previous publication with a look at the spectral response function of temporally induced blackness, as well as its additivity (Volbrecht, *et al.*, 1989b). For the temporal induction, observers adjusted the intensity of a monochromatic inducing stimulus until the reference stimulus appeared black. They then adjusted the intensity of a monochromatic stimulus to match the intensity of a broadband white standard using flicker photometry. For their final task, they adjusted the intensity of a monochromatic stimulus to match that of a broadband stimulus in a bipartite field. Each of the results from these three tasks was normalized, and the shapes of each curve were compared. Two of the three observers gave temporally induced spectral response functions that matched their flicker functions only, while the third observer's curves matched for all three tasks. The two observers whose brightness functions differed from their other functions were further tested by the presentation of a conditioning field between the inducing and test stimuli. One observer showed no effect in this test, while the other observer's blackness induction function changed as a function of the conditioning wavelength.

Previous works indicated that flicker photometry results are additive, while brightness results are not (Volbrecht, *et al.*, 1989b). Thus, if temporally-induced blackness shows additivity, it could be assumed that the neural pathways being used are similar to those used for flicker tests, while nonadditivity could point to the pathways being more similar to those used for brightness determinations. The three tests were thus repeated, using either monochromatic or dichromatic light. They were then repeated a third time, using varying ratios of two monochromatic lights as stimuli. After these tests, it was seen that the blackness induction tests and the flicker tests showed additivity, while the brightness tests did not. Blackness, then, seems to arise from different neural pathways compared to brightness. It was also found that the spectral sensitivity curves for temporally induced blackness matched the previously found curves for spatially induced blackness. In their conclusion, the authors stated that blackness was a function of the luminances of the reference and inducing stimuli, regardless of the chromatic content of either. It is critical to notice that this applies to blackness in self-luminous displays only.

Once it was determined that temporally induced blackness was additive, Volbrecht, Werner and Cicerone examined the additivity of spatially induced blackness (Volbrecht, *et al.*, 1990). They again found that induced blackness was additive, and was therefore more closely related to flicker photometry results than brightness results. In 1994, Shinomori, Nakano, and Uchikawa revisited spatially induced blackness, studying the effects of illuminance and hue of the annulus on the perception of blackness in a central spot (Shinomori, *et al.*, 1994). Their results agreed with the previous results: hue was insignificant. The Japanese researchers also concluded that induced blackness is determined by luminance, not brightness, and is thus more closely related to flicker functions. In 1997, Shinomori, Schefrin and Werner presented a quantitative model for spatially induced blackness (Shinomori, *et al.*, 1997). This model accounted for the appearance of blackness in the center spot, regardless of ring or annulus hue. They determined that the hue of the center may affect blackness, but the hue of the surround is irrelevant. Interestingly, they found that the influence of the hue of the center is variable with wavelength. At medium to long wavelengths, the hue is less significant than it is at short or very long wavelengths. Plotting contours of chromatic influence on a CIE u’v’ plot showed that the smoothest regions for each observer are in the cyan region, while the steepest regions are in the red/purple areas.

4. Mechanisms of Perceived Blackness

In 1998, Volbrecht and Kliegl presented an historical and contemporary review on the perception of blackness (Volbrect and Kliegl, 1998). A comprehensive overview of the Helmholtz and Hering theories of blackness and a focused examination of recent studies on the physiological and neurological basis of blackness perception was given.

According to the Helmholtz’s theory (1962) the sensation of blackness is due to the total absence of light, a view that was held by Ladd-Franklin (1932). Helmholtz’s theory holds that greyness arises from the red, blue, and green sensations coinciding and does not consider greyness as a unique sensation. Hering (1964) felt that blackness could only be seen under the influence of some external light stimuli and that contrast, either spatial or temporal, was required to truly sense blackness. Hering considered the phenomenon of Eigengrau (the gray color seen when one closes his eyes) as supporting evidence. Helmholtz’s trichromatic theory is upheld in the presence of three broad types of cones in the retina, while Hering’s opponent color theory matches the post-receptor coding that occurs during signal transmission leading to color vision. A combined model was described by Hurvich and Jameson (1957) in the 1950’s.

Two schematics were presented to illustrate the physiological mechanisms of color vision. The first, reproduced in Figure 2.2, shows the link between the cones in the retina and the opponent processing in the brain according to Hurvich and Jameson (1957). The second, reproduced in Figure 2.3, shows Shinomori's proposed model (1997) for the perception of blackness under center/surround spatial contrast. The model shown in Figure 2.2 omits the cone inputs for blackness because of complexity, while Figure 2.3 only shows the inputs for black. Using these models, Volbrecht and Kliegl presented a brief discussion of the physiological mechanisms responsible for the perception of blackness (Volbrecht and Kliegl, 1998).

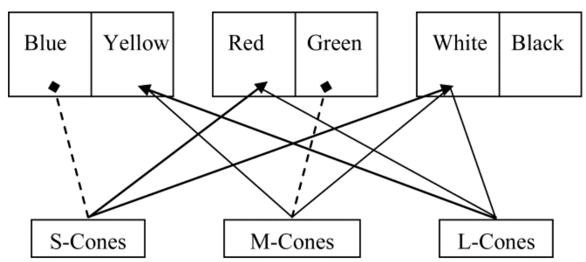

Figure 2.2. Hurvich and Jameson's opponent-process model of color vision. Solid lines indicate excitatory processes, while dashed lines indicate inhibitory processes.

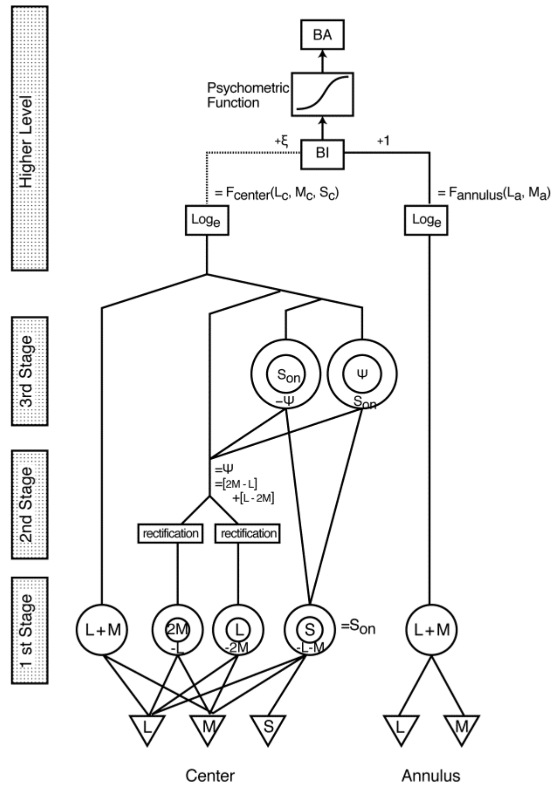

Figure 2.3. Shinomori's model of blackness (Shinomori, *et al.*, 1997). Reprinted with Permission.

Bipolar cells, either directly or through horizontal cells, receive input signals from only rods or cones, never both, although cone bipolar cells may be connected to more than one type of cone. Cones connect to either ON or OFF bipolar cells, beginning the ON/OFF pathways to the brain (Wu, 2010). Only one type of rod bipolar cell exists, and the coding for data coming from rods begins in the amacrine cells which form a feedback loop with bipolar cells, amplifying the signals sent along the neurons. Both of these cell types feed into the ganglion cells, which transmit signals directly, via the axons forming the optic nerve, to the brain. Five types of ganglion cells have been established: midget (connected to a single cone cell, also called X-type), parasol (connected to many rods and/or cones, also called Y-type), bistratified (connected to several rods and/or cones), photosensitive (responsible for circadian rhythms), and other (responsible for miniature eye movements). Midget and parasol ganglion cells send signals to the parvocellular and magnocellular pathways, respectively, and are characterized by responsive fields having surround/center geometry. Bistratified ganglion cells send signals to the koniocellular pathway, and have uniform responsive fields that are always ON to S-cones and OFF to M- and L-cones, however, their role in color vision has not been fully elucidated (Dacey and Lee, 1994). Axons from the ganglion cells form the optic nerves. From the optic nerves, signals may pass into either the geniculostriate or tectopulvinar visual system. The receptive fields of the bipolar or ganglion cells form the basis for post-receptor coding of colors. As Figure 2.3 shows the S+M+L cells handle the whiteness or luminance, while signals corresponding to blackness are more complex. This is due to the requirement of contrast for blackness to be perceived.

In 2005, Stockman, Plummer and Montag used flicker photometry to attempt to further elucidate the achromatic pathways (Stockman, *et al.*, 2005). Their research indicated that seven mechanisms are involved in the perception of luminance. These mechanisms may be relatively fast (f) or slow (s), and include inputs from S, M, and L cones. The seven identified mechanisms are +fM, +fL, +sM, +sL, -sM, -sL, and -sS. In their postreceptoral model, the two fast mechanisms are added together and set opposite the summation of three slow signals. The three slow signals include M and L signals of opposing polarities. The reason for the difference in speed between the mechanisms is not understood, but it is supposed that it may be due to differences in transmission rates among ganglion cells (Stockman, *et al.*, 2005). A simple schematic of this model is shown in Figure 2.4.

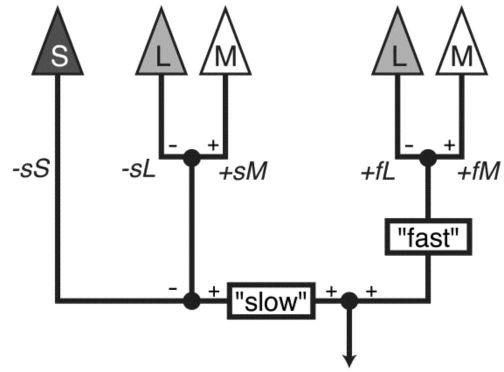

Figure 2.4. Stockman's simplified model (Stockman, *et al*., 2005). Reprinted with Permission.

In 2009, Bimler, Paramei, and Izmailov published the results of a color naming experiment. They found hue and saturation shifts in spatially induced blackness (Bimler, *et al.*, 2009). The experiment consisted of a center of varying wavelength surrounded by a white annulus of varying luminance. Observers were asked to use one, two, or three basic color terms to describe the color of the center of the stimulus. Point values were assigned to the responses based on their salience and color-naming functions were derived from these point values. Using multidimensional scaling, they separated induced blackness from desaturation and found nonlinearities in visual processing. Their analysis led to a four-dimensional solution; green/red and blue/yellow were two dimensions, while the remaining dimensions represented blackness/lightness and desaturation. From these results, a more comprehensive model of induced blackness was presented. This model differs from the Shinomori model in that it included saturation effects as well as brightness, and was neither quantified nor tested and none of the channels were associated with any physiological system or the striate cortex layer. The model is schematically shown in Figure 2.5.

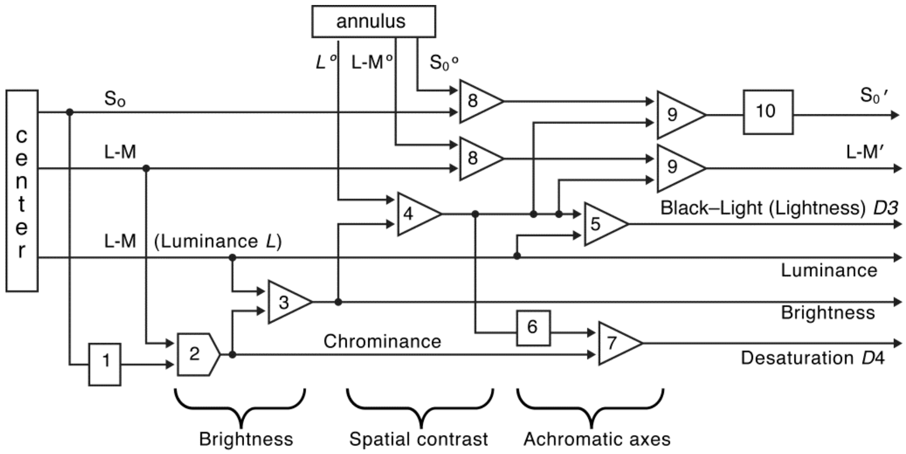

Figure 2.5. Bimler and Paramei's model (Bimler, *et al.*, 2009). Reprinted with permission.

5. Efforts Aimed at Development of a Blackness Index

Several one-dimensional scales of blackness have been used in various industries, although they are narrow in scope. The most familiar of these is the NCS Blackness scale. The NCS scale is inappropriate for use with near black objects, because of the small numerical range afforded to near black samples. This scale runs from a value of zero, representing pure white, to 100, representing pure black. Additionally, the NCS scale has no hue or chroma component and is based purely on lightness, much like the Munsell value scale. The Kenya Bureau of Standards recommends a method for specifying the blackness of pencil leads by drawing thirty lines 1.5 mm apart on a white paper with a reflectance of 79±2% (Kenya Bureau of Standards, 2005). The reflectance is measured via a photocell connected to a sensitive galvanometer in a specially fitted box and used to quantify the blackness of each pencil. This method gives values generally between 60 and 80 percent reflectance. At least one patent (Nakajima, *et al.*, 2007) recommends rating blackness as a function of the lightness of a sample.

In the 1970's and 1980's, several blackness indices were suggested for use in the German printing industry (Schumacher, 1976; Bode, *et el.*, 1979; Gläser, 1978; Lippok-Lohmer, 1986). As these were created for specific applications they mostly dealt with blackness in terms of lightness, although a few also incorporated chromaticity terms. These formulae included an assumption that bluish blacks appear deeper than brownish blacks of similar lightness values.

In 1976, Schumacher presented a measure of blackness at the Fédération d'Associations de Techniciens des Industries des Peintures, vernis, émaux et encres d'imprimerie de l'Europe Continentale (Federation of Associations of Technicians of Paint, Varnishes, Enamels and Printing Inks in Continental Europe) (Schumacher, 1976). This blackness metric was based on the ratio of absorption at 430 to that at 750 nm, which gave higher values for browner soot pigments (which were more refined than bluer ones). Earlier, Hardy at MIT had designed a "nigrometer", an apparatus created that allowed for visual comparisons of black samples to a standard whose lightness could be varied (Hardy, 1930). The nigrometer was quickly found to give no better results than simple visual assessments, and thus its use was never widespread (Schumacher, 1976). Schumacher's work was based on readings taken with a Zeiss RFC 3 spectrophotometer. By using a grey standard in place of white for calibration, and implementing a new black standard based on a large trap lined with black velvet, compared to traditional black standards, more accurate readings on very dark samples were obtained. Working with 22 samples, 19 trained observers ranked the samples in order of blackness and then assigned a blackness value to each sample, with 0 being the brightest (*die hellste Probe)* sample and 100 being the darkest sample (*die dunkelste Probe)*. Observers were finally asked to arrange the samples from blue to brown, but this was found to be especially difficult due to lightness differences amongst the samples.

In accordance with the German standards, the hue and chromaticity of the samples were calculated using the Adams coordinates, a_1, a_2, a_L, F, and φ (Schumacher, 1976). Schumacher conceded that it was simple to describe the colors of the samples using these variables, but recognized that a single value would be more useful in daily practice. The scale was arranged to give higher values for deeper colors, as this would be more intuitive. This blackness scale was denoted *M*, after the ancient Greek *μελας* , meaning black, and it was calculated using Equation 3 which was subsequently modified to that shown in Equation 4. Here a_L is the Adams' coordinate for lightness, while a_1 and a_2 are the Adams' chromaticity coordinates.

$$M = 100(2 - \log a_L) \tag{3}$$

$$M_F = 100[2 - \log(a_L + a_2 - a_1)] \tag{4}$$

The correction implemented in Equation 4 adjusted the formula for determining the blackness differences of equi-lightness bluish and brownish samples since Equation 3 rated them identically. Notes on the derivation of either formula cannot be found and it is not clear why the variables were weighted equally.

In 1978, Gläser published a model for the colorimetric behavior of black prints based on the practical motivation that adding a small amount of blue pigment to black printing inks caused them to appear deeper (that is, blacker) (Gläser, 1978). Due to the low reflectance, high gloss, and nonuniformity of print samples, measuring the effects of the added blue pigments was difficult. To better determine the effects of the added pigments, a thorough model of a black print on a substrate was presented. This model dealt with aspects such as thickness of the print layer and print translucency. Three measurement techniques were considered: 0/d specular included, 0/d specular excluded, and 45/0 specular included. None of these were found to be completely satisfactory, but it was noted that 0/d specular included gave the worst measurement geometry for matching observer results, as humans are able to ignore specular reflection when assessing samples. It was recommended to standardize the size and location of gloss traps in spectrophotometers so that their results would be more easily comparable. Schumacher's blackness indices, converted from the Adams' coordinates to the CIELAB coordinates, shown in Equations 5 and 6, were also discussed.

$$M^* = 100(2 - \log L^*) \tag{5}$$

$$M_F^* = 100[2 - \log(L^* + b^* - a^*)], \tag{6}$$

where L^*, a^*, and b^* are the colorimetric attributes of the sample. Equation 6 was found to be particularly useful as reddish blacks would give a reduced calculated blackness. The use of the colorimetric attributes of a sample could allow for low brightness or color to be weighted differently, as different applications required.

In 1979, Schumacher,along with Bode, Ferch and Koth, published a redefinition of the *M* scale using CIELAB coordinates (Bode, *et al.*, 1979). It was stated that prior to 1976, no reproducible procedure for the assessment of blackness was known that would be nearly as accurate as the human eye. The darkness scale, *D*, which was analogous to lightness and defined by DIN 6164, as shown in Equation 7, was also discussed.

$$D = 10 - 6.1723 \log\left(40.7 \frac{A}{A_0} + 1\right) \quad (7)$$

Here A is the lightness value determined by DIN 5033, and A_0 is the maximum lightness of a color in the same category. The value of A is equal to the Y tristimulus value calculated for Illuminant C and 1931 Standard Colorimetric Observer. For achromatic colors (and for most near-blacks), A_0=100, and the Equation 7 simplifies to that shown in Equation 8.

$$D = 10 - 6.1723 \log(0.407A + 1) \quad (8)$$

Using this scale, an ideal white sample (A=100) gives D=0, while an ideal black sample (A=0) gives D=10. It was noted that most near-blacks have lightnesses between 0.04 and 0.8, which led to D values between 9.96 and 9.24. It was recognized that this range was too small to practically describe the differences in blackness that the human eye can recognize, and that a larger range was needed. The formula for calculating D was used as the basis for a larger scale, giving rise to the logarithmic form in the M scale. In 1977, DIN 6174 was revised such that the expression for L^* included a cube root term, analogous to the CIE W*. After this revision, the M scale was revised to the form shown in Equation 9.

$$M = 104.4 - 100\log Y = 100(2 - \log L^*), \text{ for } Y \leq 0.8856 \qquad (9)$$

Bode, et al also introduced the Bode scale, B, based on a visual derivation of the blackness scale (Bode, *et al.*, 1979). The Bode scale had a linear relationship with the Hunter L values, and thus $\sqrt{Y}$, for the samples tested. No further details on the Bode scale were reported.

In 1986, Lippok-Lohmer proposed new versions of the blackness scales, M_Y and M_C (Lippok-Lohmer, 1986). M_Y, given in Equation 10, was analogous to the M scale, but used the Y tristimulus in place of L^*.

$$M_Y = 100(2 - \log Y) \qquad (10)$$

Samples of constant brightness were visually ranked by hue, with bluish samples being considered blacker than brownish samples. This was consistent with results based on the M_F^*

scale shown in Equation 6. However, it was noted that the M_F* formula tended to fail for violet-to-purple samples, and a new formula, M_C, given in Equation 11, was proposed.

$$M_C = 100\left(\log\frac{X_n}{X} - \log\frac{Z_n}{Z} + \log\frac{Y_n}{Y}\right), \tag{11}$$

where X, Y, and Z are the tristimulus values of the sample, and X_n, Y_n, and Z_n are the tristimulus values of the reference white sample. This formula weights blueness less heavily than the former M_F* formula, and was found to agree better with visual observations. As with the previous formulae, the difference found by M_C-M_Y is a measure of the hue and chroma of the sample (previously found by the difference M-M_F).

6.1. Recent Works on Development of Blackness Index

In 2006, Westland reported work on the density of the ink as a metric for blackness (Westland, *et al.*, 2006). One hundred samples were printed and mounted on medium grey cards. The samples were divided into five sets of twenty samples each and 33 observers were asked to rank each set in order of perceived blackness. Four forms for the blackness index were proposed as shown in Equations 12-15.

$$B1 = a_1 + a_2Y + a_3(x - x_n) + a_4(y - y_n), \tag{12}$$

$$B2 = a_1 + a_2Y + a_3(x - x_n)^2 + a_4(y - y_n)^2, \tag{13}$$

$$B3 = a_1 + a_2L^* + a_3a^{*2} + a_4b^{*2}, \tag{14}$$

$$B4 = a_1 + a_2X + a_3Y + a_4Z. \tag{15}$$

Observers disagreed with the mean rankings 22% to 58% of the time. The proposed equations disagreed with the mean rankings 28% to 47% of the time. It was reported that a formula of form B3 would be the most effective and a proposed blackness index, shown in Equation 16, was given.

$$B = 9.6523 - 0.3118L^{*} - 0.0054a^{*2} - 0.0003b^{*2}. \qquad (16)$$

The blackness index values calculated for samples (except for the samples that were never called black) were plotted against the percentage of observers that called them black and only a weak trend was obtained. It must be noted that the lightness (L^*) of samples ranged from 5 to 40 and some highly chromatic samples (a^* = –90to 20, b* = –45to 65) were also included.

More recently, Clonts, Shamey and Hinks published results from work on perception and quantification of blackness (Clonts, *et al.*, 2010). They completed two experiments using Munsell colored sheets (X-Rite) as well as acrylic fabrics dyed to different shades of black. For the experiment with the Munsell papers, observers were asked to force rank two sets of ten samples from least like black (assigned a value of 10) to most like black (assigned a value of 1). Results are given in Figure 2.7 which shows samples with hue angles, h, in the CIELAB system, between 180° and 270° are perceived blacker than those of other hues. The selected set, shown in Figure 2.6, consisted of the six blackest samples for each observer that were selected during the initial part of the experiment from two sets of ten samples each, with assigned values of 1 (blackest) to 6 (least black), where unselected samples were assigned a value of 13.5.

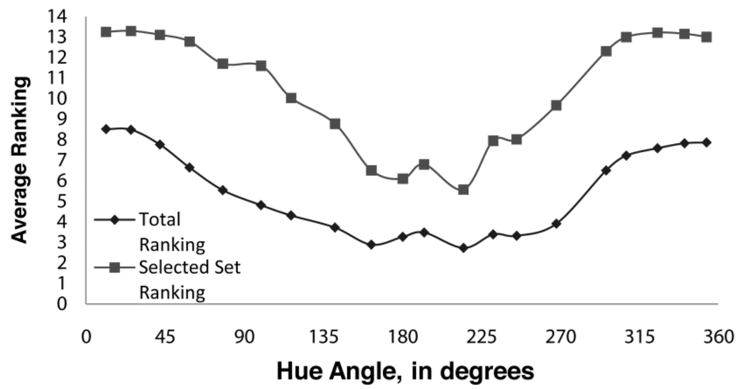

Figure 2.6. The effect of hue on perceived blackness of Munsell glossy samples for the total set and the blackest samples.

Similar results were obtained in tests using dyed acrylic fabric samples. The samples were divided into five groups based on C^* values: A (C^* = 0.12-0.20), B (C^* = 0.42-0.57), C (C^* = 0.89-0.97), D (C^* = 1.58-1.86), and E (C^* = 3.34-3.46). Observers were asked to first indicate whether the samples were perceived as black or not. Results are shown in Figure 2.7 and are represented by Equation 17. Low chroma samples were considered black by more observers compared to relatively higher chroma samples, regardless of hue angle, and that samples with hue angles between 225° and 270° were considered black more often than samples with other hue angles, regardless of chroma.

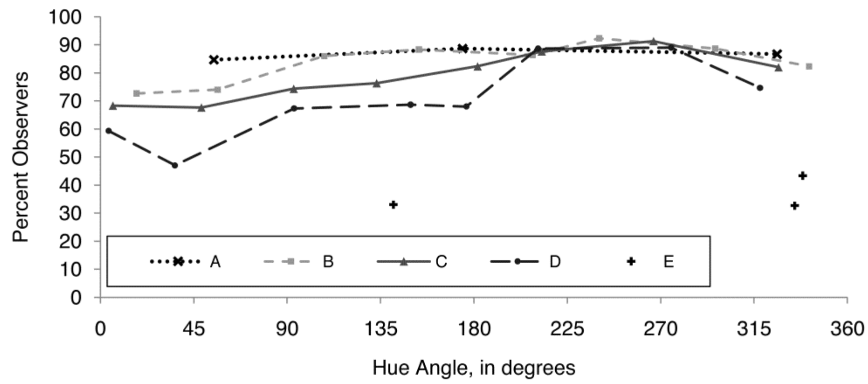

Figure 2.7. Percent of observers considering each sample “black”.

The second task for observers of the dyed acrylic samples was to rate each sample individually. A “black box”, consisting of a wooden box whose interior was lined with black velvet and whose outside was painted medium grey to match the interior of a Spectralight III (X-Rite) viewing booth, was presented to observers. A 2” × 2” aperture was cut in the lid of the box and minimal, if any, light was allowed to escape. Observers were told to consider the color seen through the cavity as perfect black, with a rating of zero. All blacks were assigned a value between 0 and 10 and non-blacks were assigned a value higher than 10. Figure 2.8 shows the average rating of each set.

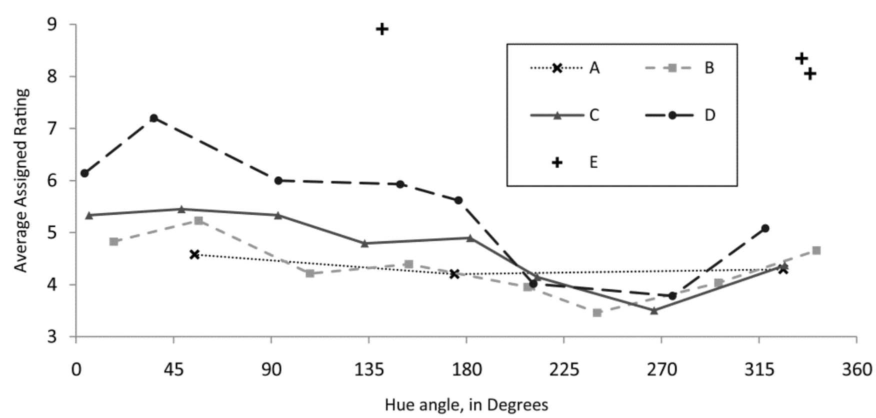

Figure 2.8. Mean assigned ratings for all acrylic samples.

Again, samples with hue angles between 225° and 270° were considered blacker than samples with other hues. A preliminary blackness index, shown in Equation 18, was also proposed based on the experimental results (Clonts, *et al.*, 2010).

$$\%NCSU_{BK} = 92.89 - 14.96C^{*} + 10.75\cos(165.22 - \mathrm{h}) \qquad (17)$$

$$NCSU_{BK} = 3.6 + 1.27C^{*} + 0.90\cos(113.80 - \mathrm{h}) \qquad (18)$$

6. Final Remarks

While early works on blackness focused on a philosophical debate they may have provided the impetus for the limited experimental findings that followed. Although it may be argued that whiteness and blackness are the opposite ends of the same scale or two ends of adjoining scales, with gray in the middle, undeniably both are of fundamental importance in colour and vision science. Surprisingly, however, the perception and quantification of blackness has not attracted the same degree of attention as whiteness. Indeed over one hundred indices have been developed to determine the degree of whiteness of samples yet our extended effort has amounted to unearthing only a small number of German reported models that discuss blackness. Reported literature on the induction of blackness, from both temporal and spatial perspectives, has contributed to our current understanding of blackness. As W.D. Wright noted, luminance and blackness are likely separated during neural coding of color and contrast seems to be as important as luminance in the determination of blackness. A detailed examination of blackness would not only enhance our understanding of human perception, but also aid the technology sector in the development of a universally accepted blackness index and appropriate production protocols and standards leading to reduced discrepancies amongst all sectors when communicating color. While recent models have attempted to elucidate the post-receptoral coding of the achromatic visual channel, blackness is still a topic that is not fully understood and additional concerted effort, perhaps through the CIE, is needed to provide a platform for examination of the topic.

References

Bimler, D. L.; Paramei, G. V.; Izmailov, C. A., 2009. Hue and saturation shifts from spatially induced blackness. J. Opt. Soc. Am. A 26, 163-172.

Bode, R.; Ferch, H.; Koth, D.; Schumacher, W. 1979. Schwarzgradskala f ür Pigmentrußе. Farbe + Lack 85, 7-13.

Cicerone, C. M.; Volbrecht, V. J.; Donnelly, S. K.; Werner, J. S., 1986. Perception of Blackness. J. Opt. Soc. Am. A 3, 432-437.

Clonts, J. R.; Shamey, R.; Hinks, D., 2010. Effect of Colorimetric Attributes on Perceived Blackness of Materials. CGIV 2010 -- Fourth European Conference on Colour in Graphics, Imaging, and MCS/10 Vision 12th International Symposium on Multispectral Colour Science, 83-87.

Dacey, D. M.; Lee, B.B., 1994. The 'blue-on' opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. Nature 367, 731-735.

Dimmick, F. L., 1920. A Note on the Series of Blacks, Greys, and Whites. American Journal of Psychology 31, 301-302.

Dimmick, F. L., 1929. Gray and the Color Pyramid. Amer. J. Psychol. 41, 284-290.

Dimmick, F. L., 1933. The Psychophysical Determination of the Limits of Pure Gray. Amer. J. Psychol. 45, 313-314.

Dimmick, F. L., 1941. Black and White. Amer. J. Psychol. 54, 286-289.

Evans, R., 1974. The Perception of Color. Wiley: New York.

Fuld, K.; Otto, T. A.; Slade, C. W., 1986. Spectral responsivity of the white-black channel. J. Opt. Soc. Am. A 3, 1182-1188.

Gläser, F., 1978. Probleme bei der Farbmessung geschönter Schwarzdrucke. Farbe + Lack 84, 579-584.

Hardy, A. C., 1930. US Patent 1 780 231.

Helmholtz, H. v., 1962. Handbook of physiological optics. J. P. C. Southall, (Ed. and trans.), Dover: New York.

Hering, E., 1964. Outlines of a Theory of the Light Sense; Harvard University Press: Cambridge, MA, 1964.

Hurvich, L. M. and Jameson, D., 1957. An opponent-process theory of color vision. Psychological Review 64, 384-404.

Judd, D. B., 1940. Hue, saturation, and lightness of surface colors with chromatic illumination, J. Opt. Soc. Amer. 30, 2-32.

Judd, D. B., 1941. The Definition of Black and White. J. Psychol. 54, 289-294.

Kenya Bureau of Standards, 2005. Black lead pencils - specification. KS 90:2005.

Kulp, T. D.; Fuld, K., 1989. Black spectral responsivity. J. Opt. Soc. Am. A 6, 1233-1238.

Ladd-Franklin, C., 1932. Colour and Colour Theories; Harcourt, Brace and Company: New York.

Lippok-Lohmer, K., 1986. Praxisnahe Schwarzmessungen. Farbe + Lack 92, 1024-1029.

Michaels, G. M., 1925. Black: A Non-light Sensation. Psychol. Rev. 32, 248-250.

Muller, G. E. (1896). Zur Psychophysik der Gesichtsempfindungen. Kapitel 2. Zeitschrift fur Psychologie und Physiologie der Sirulesorgane 10, 321-413.

Muller, G. E. (1897). Zur Psychophysik der Gesichtsempfindungen. Kapitel 4. Zeitschrift fur

Psychologie und Physiologie der Sinnesorgane 14, 1-76.

Nakajima, Y.; Yukawa, Y.; Kamimori, I.; Yamanouchi, A.; Hoshida, Y., 2007. European Patent 1 270 624 B1.

Neifeld, M. R., 1924. The Ladd-Franklin Theory of the Black Sensation. Psychol. Rev. 31, 498-502.

Quinn, P. C.; Wooten, B. R.; Ludman, E. J., 1985. Achromatic Color Categories. Perception and Psychophysics 37, 198-204.

Rich, G. J., 1926. Black and Grey in Visual Theory. Amer. J. Psychol. 37, 123-128.

Schumacher, W., 1976. Photometrische Schwarzmessung. pp 582-587.

Shinomori, K.; Nakano, Y.; Uchikawa, K., 1994. Influence of the illuminance and spectral composition of surround fields on spatially induced blackness. J. Opt. Soc. Am. A 11, 2383-2388.

Shinomori, K.; Schefrin, B. E.; Werner, J. S., 1997. Spectral mechanisms of sptially induced blackness: data and quantitative model. J. Opt. Soc. Am. A 14, 372-387.

Stockman, A.; Plummer, D. J.; Montag, E. D., 2005. Spectrally opponent inputs to the human luminance pathway: slow +M and −L cone inputs revealed by intense long-wavelength adaptation. The Journal of Physiology 566, 61-76.

Titchener, E. B., 1916a. A Note on the Sensory Character of Black. Journal of Philsophy: Psychology and Scientific Methods 13, 113-121.

Titchener, E. B., 1916b. A Further Word on Black. Journal of Philsophy: Psychology and Scientific Methods. 13, 649-655.

Troland, L. T. (1921). The enigma of color vision. American Journal of Physiological Optics 2, 317-337.

Volbrecht, V. J.; Werner, J. S.; Wooten, B. R., 1989a. Temporal Induction of Blackness--I. Color Appearance. Vision Res. 29, 1425-1436.

Volbrecht, V. J.; Werner, J. S.; Wooten, B. R., 1989b. Temporal Induction of Blackness--II. Spectral Efficiency and Tests of Additivity. Vision Res. 29, 1437-1455.

Volbrecht, V. J.; Werner, J. S.; Cicerone, C. M., 1990. Additivity of spatially induced blackness. J. Opt. Soc. Am. A 7, 106-112.

Volbrecht, V. J.; Kliegl, R., 1998. Color Vision: Perspectives from Different Disciplines, in: Backhaus, W. G. K., Kliegl, R. and Werner, J. S., (Eds.), The Perception of Blackness: An Historical and Contemporary Review. Walter de Gruyter & Co.: New York, pp 187-206.

Ward, J., 1905. Is 'Black' a Sensation? Brit. J. Psychol. 1, 407-427.

Ward, J., 1916. A Further Note on the Sensory Character of Black. Brit. J. Psychol. 8, 212-221.

Westland, S.; Cheung, T. L. V.; Lozman, O. R., 2006. A metric for predicting perceptual blackness. 14th Color Imaging Conference Final Program and Proceedings, 14-17.

Werner, J. S.; Cicerone, C. M.; Kliegl, R.; DellaRosa, D., 1984. Spectral efficiency of blackness induction. J. Opt. Soc. Am. A 1, 981-986.

Wright, W. D., 1980. Perception of Blackness. Die Farbe 28, 161-166.

Wright, W.D., 1981. The nature of blackness in art and visual perception. Leonardo 14, 236- 237.

Wu, S. M., 2010. Synaptic Organization of the Vertebrate Retina: General Principles and Species-Specific Variations: The Friedenwald Lecture. Invest. Ophthalmol. Vis. Sci. 51 1264-1274.

Xin, J. H.; Shen, H.-L.; Lam, C. C., 2005. Investigation of Texture Effect on Visual Colour Difference Evaluation. Color Res. Appl. 30, 341-347.

3. Effect of Colorimetric Attributes on Perceived Blackness of Materials

Reid Clonts, Renzo Shamey, David Hinks, Polymer & Color Chemistry Program, TECS Department, North Carolina State University, Raleigh, NC 27695-8301 USA

Abstract

While black is one of the most prevalent industrial colors in the world, the colorimetric attributes of what is considered black vary significantly and the range of subtle hue undertones can be numerous. However, no systematic study can be found in the literature pertaining to the potential role of colorimetric attributes in the perceptual assessment of blackness. We have experimentally determined that the perception of blackness is influenced by hue and chroma using psychophysical assessments of a range of black materials.

In the initial part of this study a series of 2 × 2" precision cut glossy Munsell color samples comprising a hue circle with a lightness (L) of approximately 20.5 and chroma (C*) between 4 and 6 were assessed using thirty color normal observers and a filtered tungsten daylight simulator (D65). Observers were asked to arrange samples in order from most like black to least like black with no time limits in three separate sittings. In the second part of the study 27 over-dyed woolen samples were arranged in 2"× 3" dimensions. Samples in this set had a lightness range of 14-16 and C* of 0.5-3.5, and were assessed by 25 observers in two sittings in the same manner. The third set of samples comprised 24 precision cut 2" × 2" dyed acrylic samples with a L* range of 10.5-12 arranged around the hue circle. Samples were selected such that they comprised three concentric hue circles of eight evenly spaced samples each. The samples were divided into five sets according to chroma: A (C* = 0.12-0.20), B (C* = 0.42-0.57), C (C* = 0.89-0.97), D (C* = 1.58-1.86), and E (C* = 3.34-3.46).*

For the assessment of samples in the third set 100 color normal observers were employed that repeated the assessments in three separate sittings with at least 24 hours gap between each sitting. Analysis of the data indicates that, irrespective of chroma, on average samples with hue angles between approximately 200° and 270° were perceived to be the most black, i.e., cyan to bluish-blacks. Blacks with hue angles above 315° or below 45° (reddish-blacks) were considered to be the least black. Chroma and lightness also influenced the perceived blackness but for the majority of samples the effect was less pronounced.

Introduction

The search of literature yields a very small number of manuscripts pertaining to the assessment of blackness. In 1980, W. D. Wright wrote a short article in *Die Farbe* on the perception of blackness in which he discussed the separation of television signals into a luminance channel and a chrominance channel [1]. To test whether this separation was valid, R.W.G. Hunt prepared a black-and-white slide and separated a chromatic slide of uniform luminance. Hunt and Wright found that combining the slides gave a reproduction of the original image, while the colors in the chromatic slide alone appeared garishly bright and unnatural. Wright discovered that the chromaticity was unchanged in the chromatic slide, and that the contrast provided by the luminance slide was critical to seeing the original image. Projecting the chromatic slide onto a uniform dark grey background did not return the colors to their original shades. From this, Wright drew a few conclusions. His first pertained to neural coding: current thinking at the time assumed that the black-white opponent channel was coded from the sum of the three types of cones in the retina. This idea led to the conclusion that blackness and luminance would be coded simultaneously, but his experiment showed that luminance must be separated from blackness at some point. His second conclusion was that the luminance alone was insufficient for color reproduction, as contrast is also important. Finally, he stated:

"*This raises the question of how we should measure or specify blackness. We might, perhaps, use the black-content scale of the Ostwald charts, or the black-white scale of the Natural Colour System, or the darkness-degree scale of the DIN System. What we do realize is that the blackness we are interested in is a subjective perception and not something that can be measured on any simple photometric scale. Moreover, the depth of the blackness that we perceive can be affected by quite small areas of contrasting lightness, for example, by the small highlights on the black grapes of Dr. Hunt's first demonstration slide.*"

To further the understanding of perceptual blackness a formal study was initiated at North Carolina State University in 2004 to examine the role of colorimetric attributes on the degree of blackness perceived. Concurrently, the study aimed to assess color vision various models pertaining to the perception of blackness. Preliminary results of the psychophysical assessments were presented in a special meeting of the ISCC in Portland, Oregon, USA [2]. In a different study in 2006 the preliminary results of work towards the development of a blackness index were reported [3]. This paper reports some of the additional results of the on-going endeavor at North Carolina State University.

Method

The location of all samples used in the study on the CIEa*b* plane is shown in Figure 3.1.

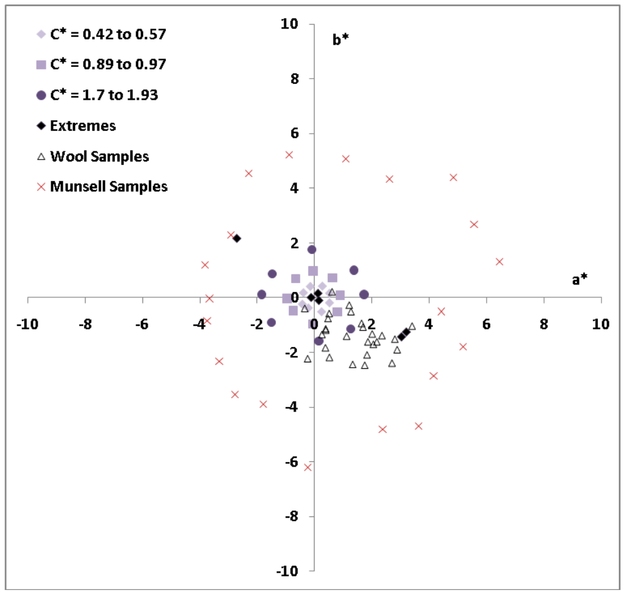

Figure 3.1. Location of glossy Munsell, over-dyed wool and over-dyed acrylic samples used on the CIE

All observers who participated in this study were tested for having normal color vision using the Neitz test [4]. All samples were illuminated using a filtered tungsten daylight simulator (Macbeth SpectraLight III, X-Rite) calibrated to 6500K. All observers were adapted to the viewing conditions for at least two minutes.

In the first part of the study a range of low value, low chroma glossy Munsell samples that constituted a full hue circle were precision cut to 2” × 2” dimensions and mounted on PVC backings to facilitate observer handling during assessments. Samples had a L* range of 19.3-

20.75 and C* of 3.66-6.58. For the assessments Munsell samples were divided into two groups as shown below:

•5R, 5YR, 5Y, 5GY, 5G, 5BG, 5B, 5PB, 5P, 5RP and

•10R, 10YR, 10Y, 10GY, 10G, 10BG, 10B, 10PB, 10P, and 10RP.

Due to the glossy nature of samples a chin rest was used during psychophysical assessments and samples were arranged such that the illumination/viewing geometry approximated 0/45 for all observers. This arrangement is shown in Figure 3.2.

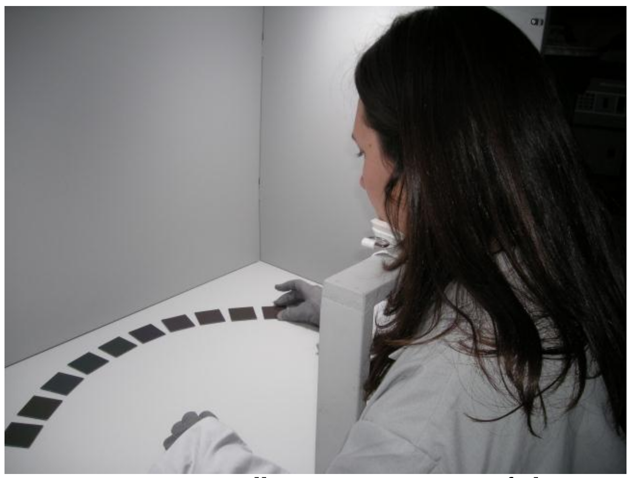

Figure 3.2. Viewing Illumination geometry of glossy Munsell samples inside a SpectraLight III booth

The five most perceptually black samples ranked by each observer in each set were then exhibited to them as a new group to obtain the final ranks. Assessments were repeated twice with a time gap between assessments of at least 24 hours. The data thus collected were statistically analyzed.

In the second part of the study 45 wool samples were dyed to different hues within a color triangle. The colored wool samples were then over-dyed with C.I. Acid Black 194. Using a

Datacolor SX600 spectrophotometer the colorimetric attributes of samples were determined using D65 illuminant and CIE 1964 Supplementary Standard Observer, specular and UV excluded and a large aperture. Samples were measured four separate times on four different locations to ensure uniformity and accuracy. Each sample was visually assessed and 27 out of the 45 dyed samples were chosen and separated into three groups of nine based on visual color variation and arranged based on their hue angle. Each sample was made into a 2” × 3” dimensions for easy handling. Each observer was asked to order four randomly presented sets of nine black samples from least black to most black. The three blackest samples from each set were set aside and used in the fourth set which was also assessed in the same manner. This test was administered to a different group of 25 color normal observers twice, each time on a different day.

Since samples representing the full hue circle could not be produced on over-dyed black wool samples in the third part of this study a large number of black samples were produced on an acrylic knit fabric. Two black cationic dyes at two concentrations were employed to initially dye the acrylic knitted fabric. The black fabric was then cut into smaller pieces and over-dyed with one of three concentrations of a trichromatic cationic dye mixture to produce twelve nominally black color triangles. From this set 30 precision cut 2” × 2” square samples with L* values between 10.5 and 12 were selected and mounted onto medium grey plastic backings. Twenty-four of these samples were selected such that they comprised three concentric hue circles of eight evenly spaced samples each with C* ranges of 0.42-0.57, 0.89-0.97, and 1.58-1.86. Six additional samples, with C* between either 0.12-0.20 or 3.34-

3.46 were also used. The samples were divided into sets according to chroma: A (C* = 0.12-0.20), B (C* = 0.42-0.57), C (C* = 0.89-0.97), D (C* = 1.58-1.86), and E (C* = 3.34-3.46). The samples were mounted in a custom built display easel at a 45° angle and viewing was set normal to the plane of the display. One hundred color normal observers including fifty men and fifty women completed two tasks three times each on separate sittings and with at least 24 hour gap between assessments. In the first task, viewers were randomly presented with each of the thirty samples and each sample was categorized as either "black" or "not black." In the second task, a reference black (an 'ideal black' that was essentially a light trap) was placed in the viewing booth and used to rate each sample on a custom scale as shown in Figure 3.3. The reference black comprised a wooden cube mounted at 45^{o} with a 2" square hole in the center of the plane facing the observer. The interior of the cube was lined with black velvet and the exterior was painted grey to approximately Munsell N7 to resemble the interior of the viewing booth. No light could escape the box.

Figure 3.3. Viewing Illumination geometry of acrylic samples against perfect black (the black light trap box

All extraneous light was excluded during the assessments. Observers were asked to rate each of the thirty samples using a scale where the reference black was assigned a rating of 0. Observers were instructed that a rating of 10 should be given to what they consider as their borderline black sample and, consequently, that all samples that were not perceived as black should be rated 11 or higher. No endpoint was defined for the scale. Observers were allowed to respond with zero if they felt the sample matched the black reference box perfectly.

Results and Discussion

Due to the relatively high L* and C* values of Munsell samples the majority of observers did not consider these samples to be black. Observers were thus asked to rate samples in terms of most-like to least-like black. Results from this study were analyzed in terms of auto-concordance and concordance to determine inter and intra-observer variability in assessments. Results showed relatively high degrees of repeatability amongst observers with 81% calculated concordance. In addition the agreement among observers was also high with

76% calculated concordance. Results from this task were also analyzed in terms of hue angle associated with selections as shown in Figure 3.4.

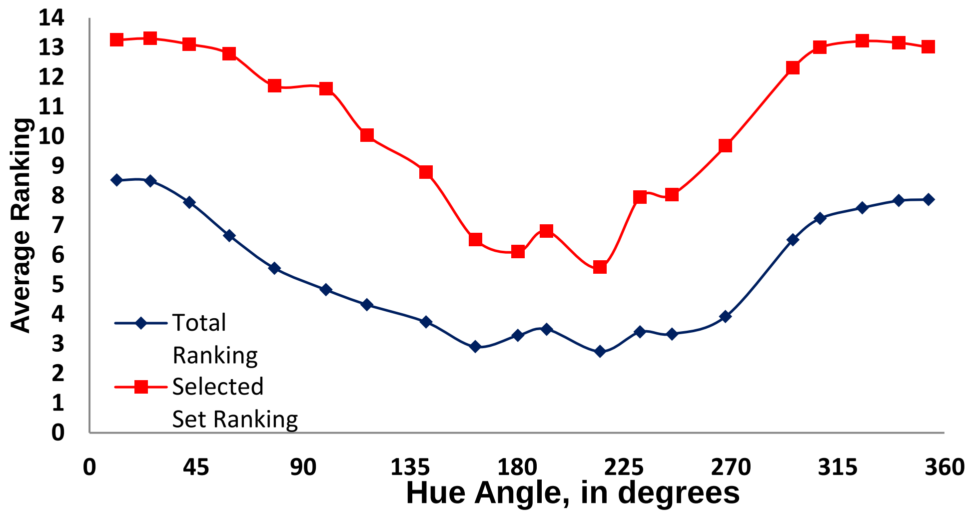

Figure 3.4. The effect of hue on perceived blackness of Munsell glossy samples for the total set and the blackest samples.

In order to address observer objections to Munsell samples' blackness, a range of over-dyed wool samples were produced and assessed by observers. The difficulty with these samples, however, was that a full hue circle could not be obtained via over-dyeing and the majority of over-dyed samples were purplish blacks with only a few in the blue-green region and none in the yellow or yellow-green region. Nevertheless since the chroma and lightness of these samples were significantly lower than those of glossy Munsell samples they were assessed by a group of observers twice to evaluate the potential role of chroma and lightness on perceived blackness. Results in terms of diminished blackness against hue angle are shown in Figure 3.5. Increased mean diminished blackness values indicate the sample is perceived as less black by the observer.

The number of over-dyed woolen samples which varied in hue across the visible range was insufficient and therefore conclusions on the potential role of hue on perceived blackness cannot be generalized. However, it can be seen that samples in the cyan to blue region were selected by observers as most black and those in the purple region were selected as being the least black. This was in agreement with the results of the Munsell sample set. In terms of lightness dyed samples had nearly constant values but their chroma varied between 0-4.

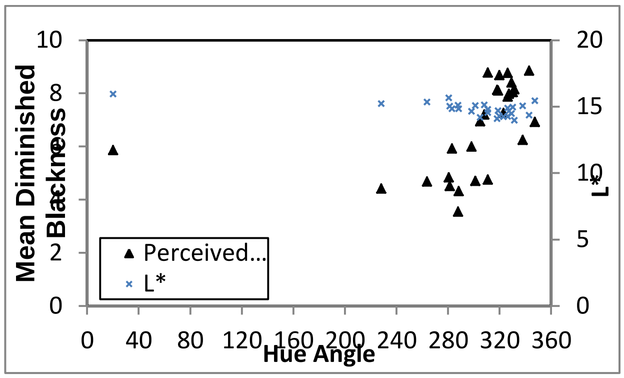

Figure 3.5. Relationship between perceived blackness against and hue angle of dyed samples.

While there were insufficient numbers of samples to assess the role of colorimetric attributes of dyed samples on their perceived blackness the potential role of variations in lightness and chroma on the perceived blackness were plotted as shown in Figure 3.6. The CIELAB color difference of samples against the most neutral black dyed sample were also separately calculated for the samples used in the study which are also shown in the figure. As can be seen there seems to be a direct relationship between increased chroma and diminished blackness which would seem expected. However, the correlation between increased chroma and diminished blackness is relatively weak (R^2 = 0.54). In terms of lightness the figure shows an inverse relationship, however, this trend is insignificant as determined by the weak correlation between parameters (R^2 = 0.38) and moreover the narrow range of lightness among these samples would make such comparisons inconclusive.

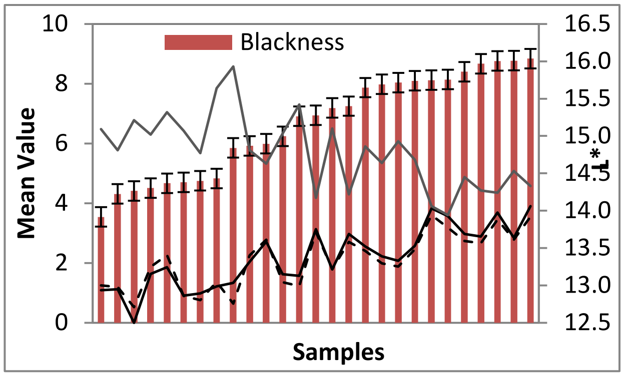

Figure 3.6. Mean blackness rank values including standard error as a function of ΔE^*_{ab} against standard (sample 3), C* and L* of samples.

Results from the third set of samples obtained by over-dyeing acrylic knitted fabric are separated into two tasks. The goal of task one was the elimination of all samples considered to be too distant from what would generally be considered black. This was analyzed using a binomial approach. Regardless of the response given during the first task all thirty samples were presented during the second task which rated the blackness of each sample against the reference black. A multivariate model was used to analyze the responses from the second task. The standard deviations of the mean rating given in the second task were also calculated, as were confidence intervals for the mean ratings. Finally, the results from the two tasks were compared, and the multivariate model for mean rating was tested against the results from the first observation trial. Excel software was used for the calculations of standard deviation and confidence intervals. A traditional t-test was used in this assessment, as the ratings of each sample were independent from each other. JMP software was used for the creation of the binomial and multivariate models.

It was decided to treat the first task as a separate experiment to help determine the boundaries of blackness in a given color space. Figure 3.7 shows the percent of observers identifying each sample black. The samples are divided by chroma group for clarity. Figure 3.7 also shows that higher chroma samples were considered black less often than samples of similar hue but lower chroma. It can also be seen that more observers agreed that samples were black when their hues were between 200° and 270°.

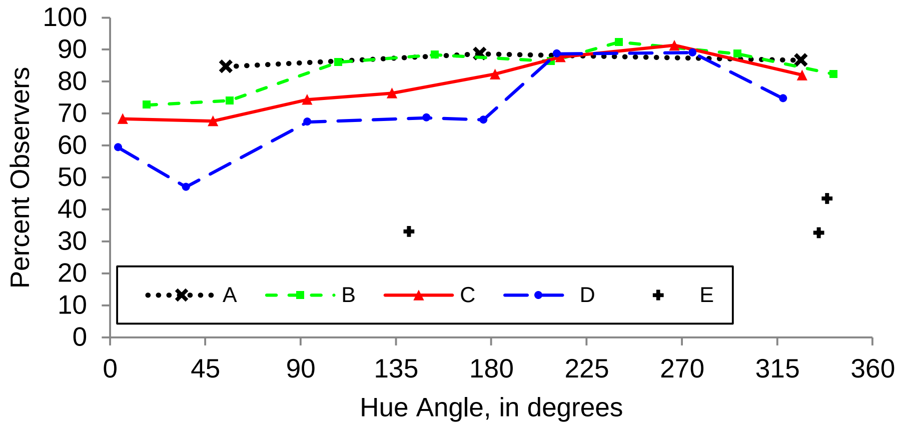

Figure 3.7, Percent of observer considering each sample black as a function of sample's hue angle.

Task two enabled assessment of perceived blackness as a function of hue and chroma. Figure 3.8 shows a graph of blackness rating as a function of hue angle for chroma groups A-E. The lower the average blackness rating, the more black a sample is perceived. Again, the higher chroma samples were perceived to be less black than those with similar hues and lower chromas, and samples with cyanish hues were rated blacker than those with other hues, regardless of chroma.

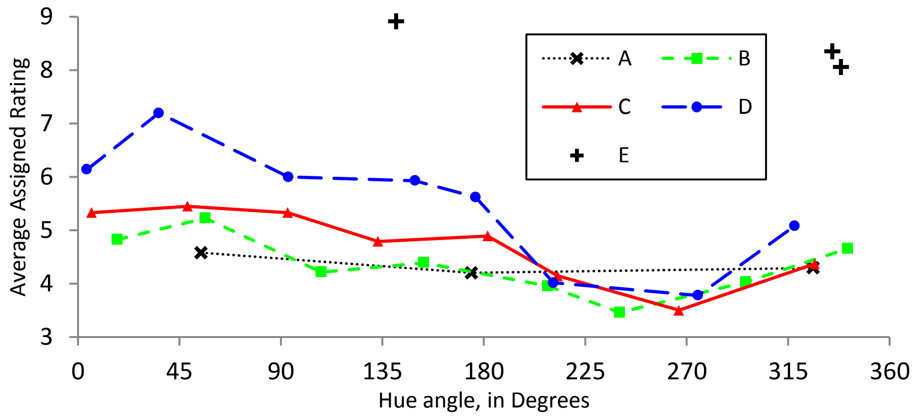

Figure 3.8. Mean perceived blackness ratings as a function of hue angle, h°, for chroma sets A-E.

The standard deviations for the mean ratings were also obtained during the second task. Observers agreed more with each other for the lower chroma samples, and more inter-observer variation was seen in the assessment of samples with increased chroma. Figures 3.9 and 3.10 show the standard deviation and the confidence intervals calculated for each mean rating respectively.

Two statistical models were obtained based on the results of this analysis. The first model gives the percent of observers who consider each sample to be black, while the second models the mean assigned rating of each sample. Both models are functions of the chroma (C*) and hue angle (h°). The correlation coefficients for these models are 0.86 and 0.87, respectively which are relatively high considering the psychophysical nature of the study.

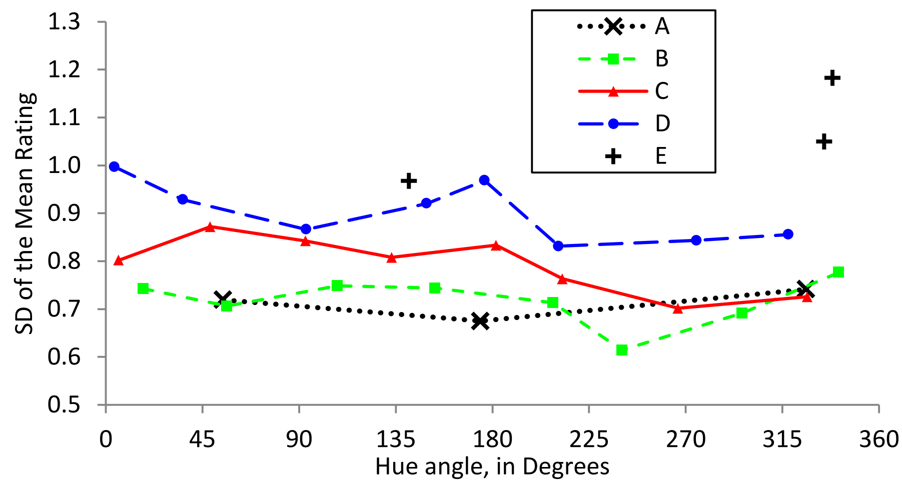

Figure 3.9. Standard deviations of mean assigned ratings as a function of hue angle.

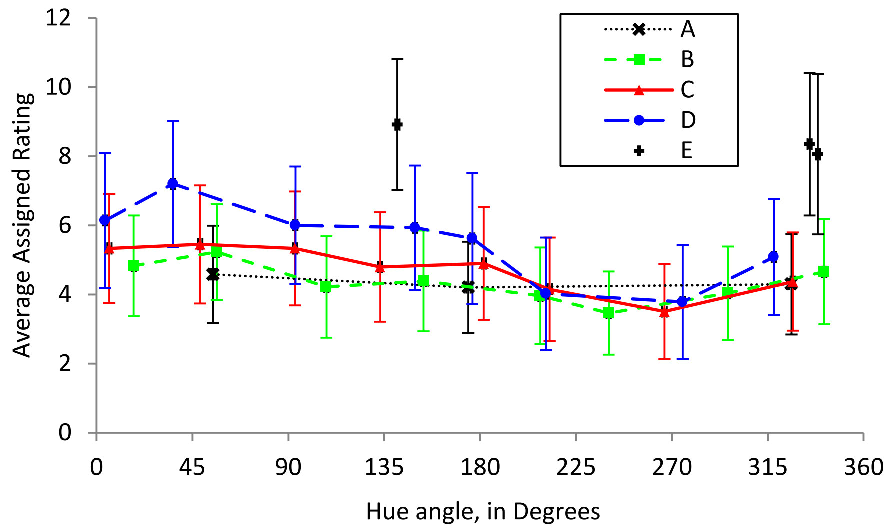

Figure 3.10. Confidence intervals for mean assigned ratings for each sample.

The models are shown in Equations 3.1 and 3.2. It should be noted that no lightness term is included in either model due to the limited range of variability in lightness of samples examined.

$$\%yes = 92.89 - 14.96(C^*) + 10.75\cos(165.22 - h) \qquad (3.1)$$

$$Rating = 3.60 + 1.27(C^*) + 0.90\cos(113.80 - h) \qquad (3.2).$$

Finally, the ratings model was used to predict ratings for the samples used in the first experiment. The model fits the data fairly well, indicating that the chroma effects are applicable to different sample types. Figure 3.11 shows how the predicted ratings agree with the empirical rankings. The predicted ratings have the same rough shape as the reported rankings, although the model seems to be less effective at hue angles in the purple region. It is interesting to note that some Munsell samples were assigned ratings below 10, indicating that they would be identified as black by the predictive model. This is not compatible with visual results. However, it is expected that the modification of the model to account for the increased variations in lightness and chroma values would resolve this inconsistency.

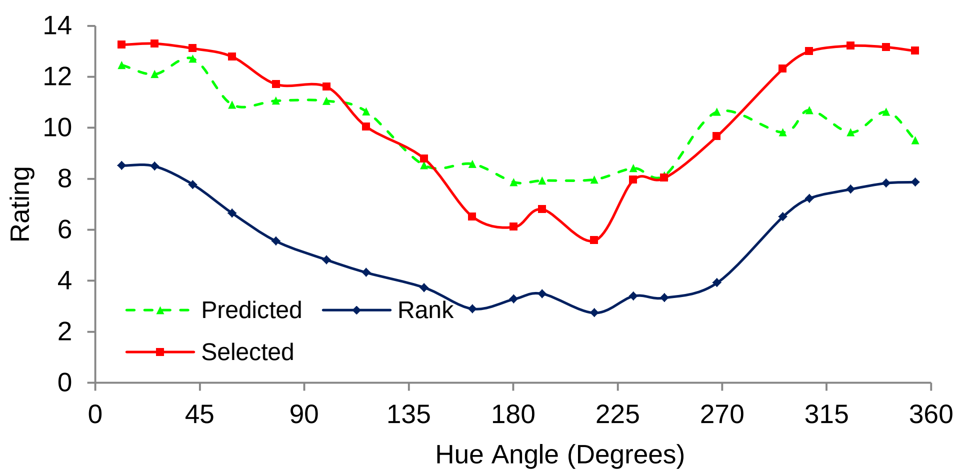

Figure 3.11. Modeled ratings for Munsell paper samples as a function of hue angle.

Conclusions

Analysis of the data indicates that, irrespective of chroma, on average samples with hue angles between approximately 200° and 270° were perceived to be the most black, i.e., cyan to bluish-blacks. Blacks with hue angles above 315° or below 45° (reddish-blacks) were considered to be the least black and the ratings trended between the most and least blacks as a function of hue angle. In general, the blackness rating was inversely proportional to C* for the samples that were not greenish- to bluish-black. Hence, for the observers studied, increasing C* has a deleterious effect on perception of blackness for all samples except greenish blacks and bluish blacks. This was in agreement with results obtained from glossy Munsell and over-dyed wool samples. Independent verification of the findings of this work should help color vision scientists in their modeling of achromatic channel as it seems that both perceived whiteness and blackness are increased with the introduction of bluish undertones.

Acknowledgments

The authors thank Prof. Bloomfield for his useful input in statistical analysis of results and all observers who participated in the study.

References

[1]. Wright, W. D., Die Farbe 1980, 28, 161-166.

[2]. Clonts R., Shamey, R. “Perceptual Assessment of Blackness”, ISCC Special Topics Meeting: Black & White, Inter-Society Color Council, 20-22 Nov, (2008) Portland, OR, USA.

[3]. Westland, S.; Cheung, T.L.V.; Lozman, O.R. In 14th Color Imaging Conference Final Program and Proceedings; 2006; pp 14-17.

[4]. Neitz J., Manual: Neitz Test of Color Vision (Western Psychological Services, 2001).

4. The Effect of Hue Angle on the Perception of Blackness Using Munsell Samples

Jennifer Reid Clonts Haslup, David Hinks, Renzo Shamey*

* Corresponding Author rshamey@ncsu.edu

Abstract

Black is one of the most commercially and artistically important colors in the world, and yet over the last several decades the perception of black objects has not been examined. The purpose of this work was to determine the influence of hue angle on the perception of preferred blackness. Once hue preference is determined the work aimed to develop a chroma and a lightness metric to clarify the contributions of these parameters to the perception of black objects. By combining these three variables, blackness may be quantified into a single blackness index scale for general use..

A set of 20 glossy low chroma Munsell Color papers were purchased from X-Rite comprising a complete hue circle with a value and chroma of two and one, respectively (L* = 19.3 – 20.75, and C* = 3.66 – 6.58.) The Munsell samples were then divided into two interleaved groups: (5R, 5YR, 5Y, 5GY, 5G, 5BG, 5B, 5PB, 5P, and 5RP) and (10R, 10YR, 10Y, 10GY, 10G, 10BG, 10B, 10PB, 10P, and 10RP). Fifty color-normal observers force-ranked the two sets of ten samples from “most like black” to “least like black”. After completion of the forced rank experiments, observers assessed a set of six samples that

represented the three samples from each set of ten that the observer chose to be "most like black".

The 50 observers were found to have fairly good autoconcordance and concordance values. In repeat experiments observers agreed with themselves in 81% of the pair-wise decisions, and they agreed with the grand mean rank 76% of the time. The blue-green samples (with Munsell hue notations 10.G, 5.BG, and 10.BG) were most selected (and were considered blackest), followed by green, blue, and purple-blue. The samples selected the fewest times by any observer as being most black were the red samples. The grand mean rankings demonstrate that greenish to bluish blacks are perceived by the observers as "blacker" than yellowish and reddish blacks.

Key words: Blackness, blackness preference, psychophysical assessment of blackness, blackness in Munsell system.

Introduction

Only relatively recently has the perception of black objects been studied, although black is one of the most commercially and artistically important colors worldwide. In the late 1970's, Bode, *et al*., reported preliminary work on the quantification of blackness using black print pastes, but this work started with the assumption that bluish blacks were the "best" without testing that assumption..[1,2,3] Commercial printers have also often found that adding a small amount of blue ink improves the appearance of dark black patches.

It is only in the past decade that a thorough, systematic examination of blackness has been attempted.[4,5] The purpose of the work reported here was to determine the influence of hue angle on the perception of preferred blackness of objects and extend the preliminary results reported previously.

Westland proposed a blackness index shown in equation (1), based on forced rankings of one hundred samples, in which a higher B_3 value indicates a darker and thus blacker sample.

$$B_3 = 8.6542 - 0.2583L^* - 0.0052a^{*2} + 0.0045b^{*2} \qquad (1)$$

Some of the samples used in the Westland study were highly chromatic, including samples with C* values near 50 that were apparently considered black by at least 50% of the 23 observers tested. This preliminary examination of the perception of very dark objects of low chroma in Westland's study and the number of observers employed to determine a potential blackness index was not sufficient.

Over, the same period a parallel study was conducted at North Carolina State University to systematically determine the effect of hue on blackness of objects. To that end the Munsell color order system was initially selected. In the Munsell color order system, there are five primary colors, red, yellow, green, blue and purple. The main motivation for inclusion of purple was to express the system in an approximately perceptually uniform manner since the distance between blue and red was considered to be much larger than that between other four colors. The other two attributes representing color in the Munsell system are lightness (called value) and “chroma” a term coined to designate the radial distance from the neutral gray in the center[6]. The value scale in the Munsell system, *V*, is related to the luminance term and places black and perfect white at opposite ends of a scale ranging from zero to ten. Therefore, the (potential) contribution of blackness on perception of color is neglected in this color order system. However, the system provides a nearly uniform distribution of hue across low chroma samples that were ideally suited for the examination of the potential role of hue on perception of dark objects.

Once hue preference is determined the work aimed to develop a chroma and a lightness metric to elucidate the contributions of these parameters to the perception of blackness in surface mode. The follow up work aimed to combine these three variables to quantify blackness using a single blackness index scale, similar to the accepted and widely used whiteness indices. This second study is reported in a separate publication.

Experimental Method

A set of 20 glossy low chroma coated papers were purchased from Munsell Color (X-Rite) comprising a complete hue circle with a value and chroma of two and one, respectively ($L^* =$ 19.3 – 20.75, and $C^* =$ 3.66 – 6.58). This set was based on the lowest value at which samples comprising a full Munsell hue circle could be obtained. Samples were measured four times on a DataColor SF600X spectrophotometer equipped with a 30 mm aperture with specular light excluded and UV included, with illuminant D65, and the CIE 1964 supplemental 10^o

standard observer and averaged to obtain final colorimetric data. The chromatic information (*a** versus *b**) for the twenty samples are given in Figure 4.1 and a summary of colorimetric data is shown in Table 4.1. Specular excluded measurement conditions were used to more closely match the experience of the observers, who were required to utilize a chin rest to view samples at a predetermined angle and avoid the specular reflection from the glossy samples.

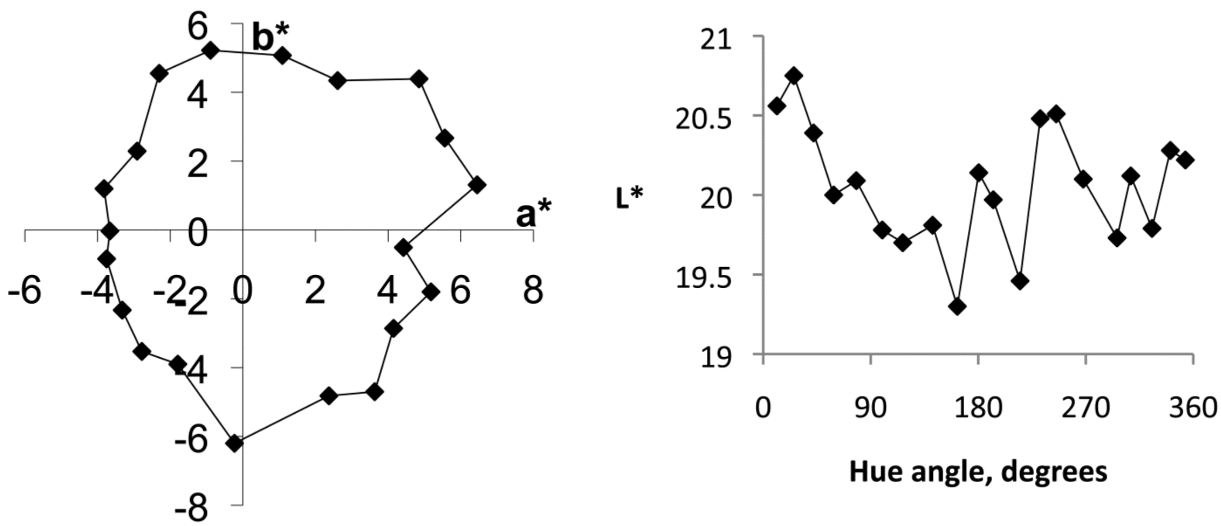

Figure 4.2. Position of twenty samples in the a^*b^* plane, and lightnesses vs. hue angles for samples.

Table 4.1. Colorimetric Values of Samples.

Sample Number	**Munsell Hue**	L*	a*	b*	C*	h
203	**5 R**	20.56	6.45	1.31	6.58	11.49
403	**10 R**	20.75	5.56	2.67	6.17	25.63
204	**5 YR**	20.39	4.85	4.39	6.54	42.14
410	**10 YR**	20	2.61	4.34	5.06	58.99
205	**5 Y**	20.09	1.09	5.07	5.19	77.92
407	**10 Y**	19.78	-0.89	5.22	5.29	99.67
210	**5 GY**	19.7	-2.3	4.55	5.1	116.8
404	**10 GY**	19.81	-2.91	2.29	3.71	141.78
206	**5 G**	19.3	-3.82	1.2	4.01	162.51
409	**10 G**	20.14	-3.66	-0.03	3.66	180.4
202	**5 BG**	19.97	-3.75	-0.84	3.84	192.58
402	**10 BG**	19.46	-3.32	-2.33	4.06	215.01
207	**5 B**	20.48	-2.78	-3.53	4.49	231.87
405	**10 B**	20.51	-1.79	-3.9	4.29	245.27
201	**5 PB**	20.1	-0.23	-6.2	6.21	267.87
408	**10 PB**	19.73	2.37	-4.82	5.37	296.2
209	**5 P**	20.12	3.63	-4.7	5.94	307.67
401	**10 P**	19.79	4.15	-2.86	5.04	325.46
208	**5 RP**	20.28	5.18	-1.8	5.49	340.84
406	**10 RP**	20.22	4.42	-0.51	4.45	353.37

Fifty color normal observers (24 females and 26 males), ranging in age from 21 to 65 with a mean age of 33, participated in the experiment. Of fifty participating observers only three had prior experience in assessment of color. Observers in this study were originally from North America, South America, Europe and Asia. Munsell sheets were cut to 2 × 2 inch dimensions with sharp edges and mounted on PVC backing to help ensure integrity of the samples. The Munsell samples were then divided into two interleaved groups: (5R, 5YR, 5Y, 5GY, 5G, 5BG, 5B, 5PB, 5P, and 5RP) and (10R, 10YR, 10Y, 10GY, 10G, 10BG, 10B, 10PB, 10P, and 10RP). Observers were given instructions to describe samples as "most like black" and "least like black", as the majority of observers did not consider all the samples to be black. Observers force-ranked the two sets of ten samples in order from "most like black" to "least like black". After completion of the forced rank experiments observers assessed a set of six samples that represented the three samples from each set of ten that the observer chose to be "most like black". Hence, each set of six was specific to each observer. The order of presentation for the two groups of ten samples was randomized in the first trial, and was reversed for each observer in the second trial. At least 24 hours was allowed between assessments. A Macbeth SpectraLight III (SPLIII) standard viewing booth employing a calibrated filtered tungsten daylight simulator (6500K) was used to illuminate samples at normal and all extraneous light was excluded during assessments. A chin rest was used during the experiments to ensure that each observer's eyes were at the same height in relation to the samples. Samples were placed in a well-defined arc on the floor of the SPLIII standard viewing booth such that viewing was at a constant 45° angle, ensuring all specular reflection was eliminated from view. Figure 4.2 shows the setup used for the experiment.

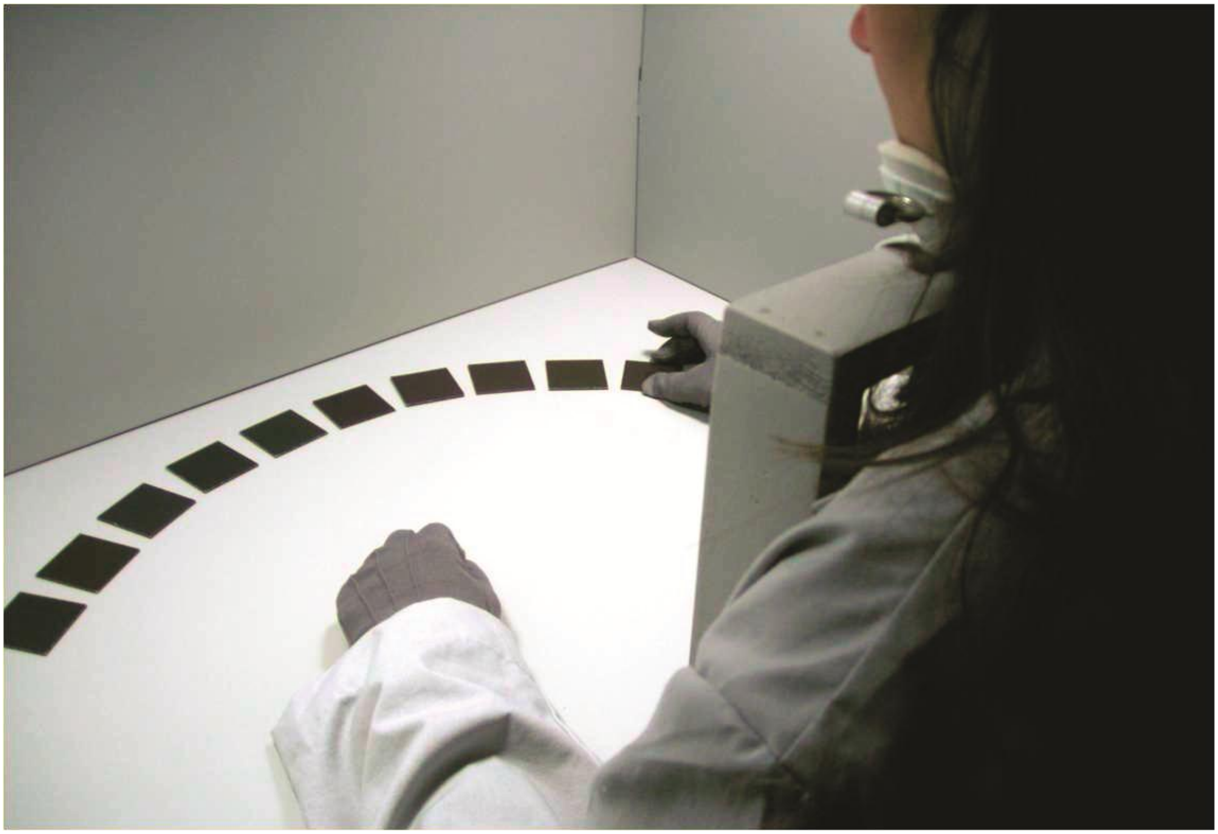

Figure 4.3. Viewing conditions employed inside a Macbeth Spectralight III viewing booth showing head position and arc of samples.

The autoconcordance was computed by comparing results from each pair of samples individually. The number of pairs that the observer ranked in the same order twice was divided by the total number of pairs. The concordance was calculated in the same way; however, each observer's mean rank was compared to the grand mean rank of all observers. Bonferroni analysis was used to determine the significance of the differences between mean rankings for each pair since the rankings for each pair were interconnected. In the Bonferroni analysis all the pairwise comparisons are considered at once. Calculations are similar to those of a t-test, with the α value being replaced by (α/n). This ensures that all of the comparisons are valid at the desired α level. For example, to compare 100 values at the $\alpha = 0.05$ level, 100 t-statistics are calculated with α = 0.0005. Microsoft Excel software was used for the Bonferroni calculations, as well as to obtain the mean rankings. The model of rank as a function of hue angle was calculated using JMP software from SAS.

Results and Discussion

The degree of intra- and inter-observer variability is of interest in studies of human color perception. In the few reported studies results vary based on the number of individuals employed, assessment conditions and age, to name but a few variables[7,8]. Results shown in Table 2, indicate that the 50 observers' autoconcordance and concordance values were in the range of, or better than, values previously reported. On average, observers agreed with themselves in 81% of the pairwise decisions, and they agreed with the grand mean rank 76% of the time. It is important to bear in mind that color perception is subjective, and variations across a pool of color normal observers are common. For this reason, 81% mean autoconcordance and 76% mean concordance are considered to be a good indication that the experiment and results are robust.

Table 4. 2. Average Autoconcordance and Concordance Values from Observer Trials.

Autoconcordance, 200 Set	80.67
Autoconcordance, 400 Set	81.60
Autoconcordance, Selected Set	61.57
Concordance, 200 Set	75.60
Concordance, 400 Set	77.02

Figure 4.3 indicates the number of times that each sample was selected in an observer's top six samples. There are 600 selections represented in this plot, and the distribution is approximately normal.

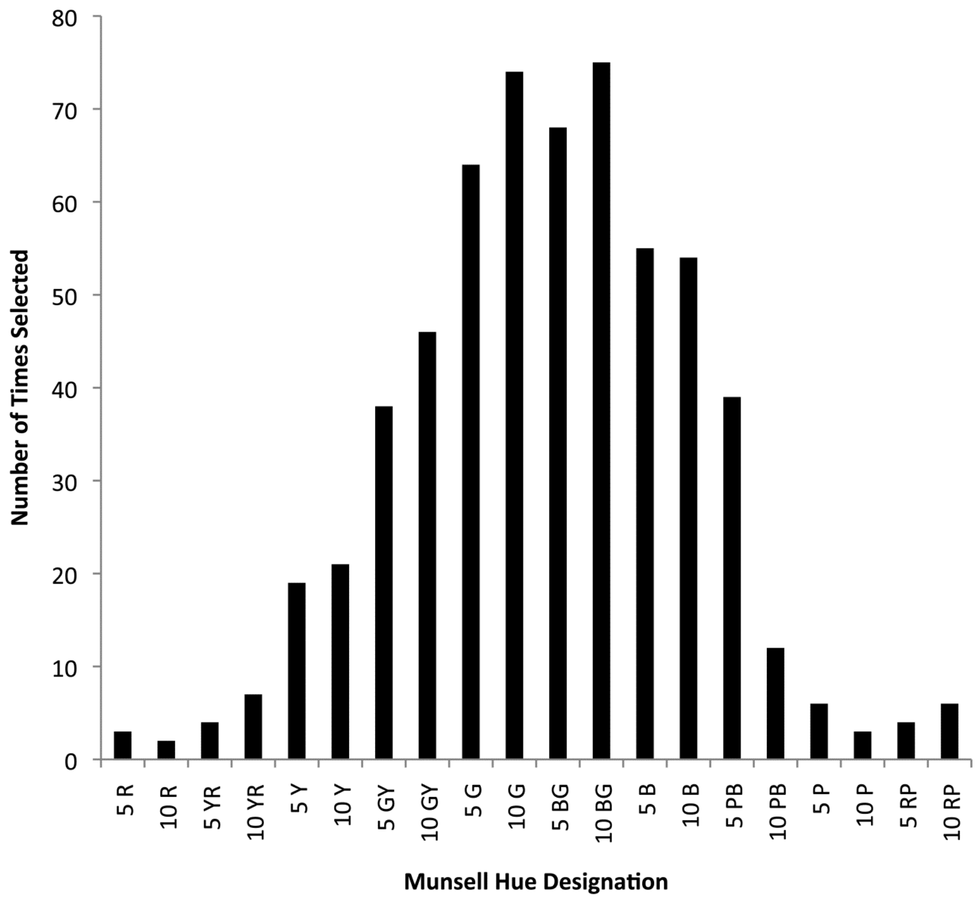

Figure 4.3. Selections in the final set of six samples.

It is interesting to note that the blue-green samples (with Munsell hue notations 10.G, 5.BG, and 10.BG) were most selected (and as Figure 3 shows, considered blackest), followed by green, blue, and purple-blue. This is in agreement with assumptions used in Bode's study and the empirical experience of practical printers. It would be worthwhile to determine any mechanistic reasons for the role of hue on perception of black objects, however, this is outside the scope of this study. The samples selected the fewest times by any observer were the red samples. Hence, reddish blacks are the least preferred samples. One possible reason for this observation may be due to the relatively high C* values of these samples (6.58 and 6.17, compared with an average of 2.83). However, two samples with relatively high C* values, 5YR and 5PB with C* of 6.54 and 6.21 respectively, were selected more often, indicating that the relatively high chroma values is likely not the primary reason for why reds were never selected by any observer.

Figure 4.4 shows that the grand mean rankings demonstrate that greenish to bluish blacks are perceived by the observers as “blacker” than yellowish and reddish blacks. In Figure 4.4, the “Total Ranking” curve refers to the grand mean rankings, while the “Selected Set Ranking” curve refers to the grand mean ranking in the selected subsets. A rank of 1 was assigned to each observer’s “blackest” sample, and higher ranked samples were considered less black. For the selected subsets, samples that were not included were assigned a rank of 13.5, the average of seven and twenty. This was because the unselected samples could not be compared to each other or to the selected samples.

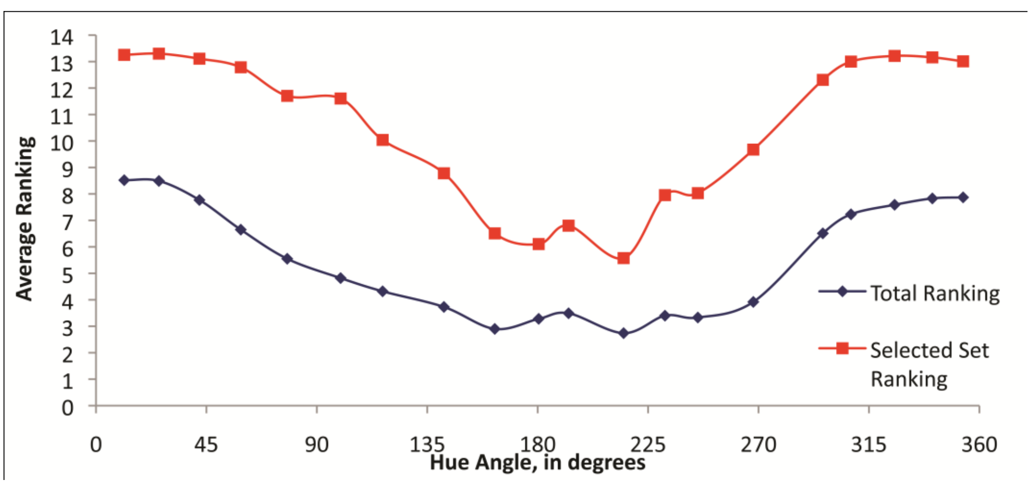

Figure 4.4. Average rank for each sample.

The difference in rank between each pair of samples was calculated for each observer, and the mean difference was calculated for each pair of samples. From these values, 95%

confidence intervals were calculated using Bonferroni analysis. Table 4.3 shows the mean difference in the ranking of each pair of samples. A grey square indicates that the difference between the sample pairs is not significant, while a white square indicates that the two samples are statistically significantly different. It can be seen that there is no difference among reddish or blue-to-green samples. However, the blue-to-green samples are perceived as being significantly blacker than the reddish samples.

Table 4.3. Differences in Mean Rankings and their Significance.

	5 R	10 R	5 YR	10 YR	5 Y	10 Y	5 GY	10 GY	5 G	10 G	5 BG	10 BG	5 B	10 B	5 PB	10 PB	5 P	10 P	5 RP
10 RP	0.65	0.62	-0.10	-2.91	-2.32	-3.05	-3.55	-4.14	-4.97	-4.59	-4.38	-5.13	-4.47	-4.54	-3.95	-1.36	-0.64	-0.28	-0.04
5 RP	0.69	0.66	-0.06	-1.18	-2.28	-3.01	-3.51	-4.10	-4.93	-4.55	-4.34	-5.09	-4.43	-4.50	-3.91	0.00	-0.60	-0.24	
10 P	0.93	0.9	0.18	-0.94	-2.04	-2.77	-3.27	-3.86	-4.69	-4.31	-4.10	-4.85	-4.19	-4.26	-3.67	-1.08	-0.36		
5 P	1.29	1.26	0.54	-0.58	-1.68	-2.41	-2.91	-3.50	-4.33	-3.95	-3.74	-4.49	-3.83	-3.90	-3.31	-0.72			
10 PB	2.01	1.98	1.26	0.14	-0.96	-1.69	-2.19	-2.78	-3.61	-3.23	-3.02	-3.77	-3.11	-3.18	-2.59				
5 PB	4.60	4.57	3.85	2.73	1.63	0.90	0.40	-0.19	-1.02	-0.64	-0.43	-1.18	-0.52	-0.59					
10 B	5.19	5.16	4.44	3.32	2.22	1.49	0.99	0.40	-0.43	-0.05	0.16	-0.59	0.07						
5 B	5.12	5.09	4.37	3.25	2.15	1.42	0.92	0.33	-0.50	-0.12	0.09	-0.66							
10 BG	6.16	5.75	5.03	3.91	2.81	2.08	1.58	0.99	0.16	0.54	0.75								
5 BG	5.03	5.00	4.28	3.16	2.06	1.33	0.83	0.24	-0.59	-0.21									
10 G	5.24	5.21	4.49	3.37	2.27	1.54	1.04	0.45	-0.38										
5 G	5.62	5.59	4.87	3.75	2.65	1.92	1.42	0.83											
10 GY	4.79	4.76	4.04	2.92	1.82	1.09	0.59												
5 GY	4.20	4.17	3.45	2.33	1.23	0.50													
10 Y	3.70	3.67	2.95	1.83	0.73														
5 Y	2.97	2.94	2.22	1.10															
10 YR	1.87	1.84	1.12																
5 YR	0.75	0.72																	
10 R	0.65																		

It should also be noted that although the samples varied in chroma (with a three unit range), the grand mean rankings were modeled only based on the hue angle. The function representing the rank against hue angle is shown in equation 2. When the data are fit with a function of the cosine of the hue angle, the coefficient of correlation, R^2 value, of 0.94 is obtained. This is an excellent fit, given the inherent variability in testing with humans and especially those involving perceptions. Figure 4.5 shows the actual grand mean rank compared to the predicted rank, as calculated by the model.

$$Rank = 5.37 + 2.72 \cos h^{\circ} \tag{2}$$

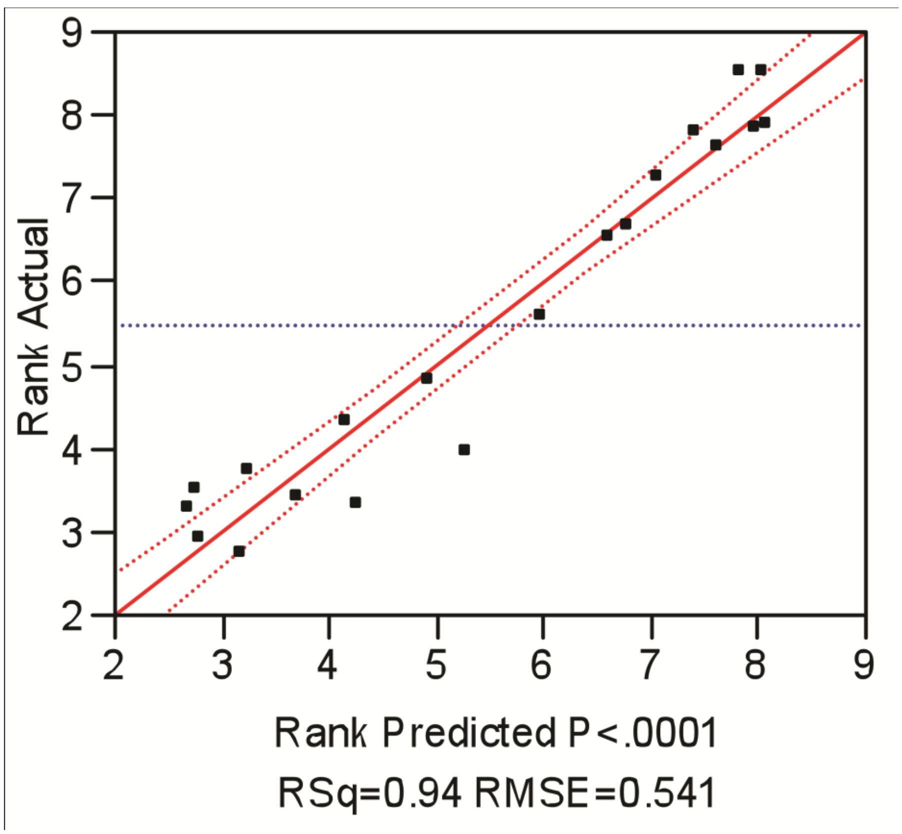

Figure 4.5. Actual grand mean rank vs. model predicted rank.

It should be noted that the chroma values of these samples were relatively high, and few observers were comfortable calling these samples black.

Conclusions

For the experimental conditions and observers used in this study, bluish to greenish blacks are considered “blacker” than colors of similar lightness and chroma with different hue angles. It is not clear whether a single preferred hue angle can be identified, due to the large variations amongst preferred colors and in particular the relatively large spaces between samples produced on a uniform hue circle. The results reported, however, are statistically supported. While the model developed does not include the effects of chroma and lightness, the idea was to use psychophysical data to demonstrate the impact of hue angle on the perception of blackness. Additional work aimed to elucidate the combined role of hue, chroma and lightness on perception of blackness and to enable the development of a blackness index is the subject of a separate study.

Acknowledgement

The authors thank all subjects who took part in the study and the Department of Textile Engineering Chemistry and Science at North Carolina State University for the Financial Support of this project.

References

1. Bode, R.; Ferch, H.; Koth, D.; Schumacher, W.. 1979. Schwarzgradskala für Pigmentruße. Farbe + Lack 85, 7-13.

2. Gläser, F., 1978. Probleme bei der Farbmessung geschönter Schwarzdrucke. Farbe + Lack 84, 579-584.

3. Lippok-Lohmer, K., 1986. Praxisnahe Schwarzmessungen. Farbe + Lack 92, 1024-1029.

4. Clonts, J. R.; Shamey, R.; Hinks, D., 2010. Effect of Colorimetric Attributes on Perceived Blackness of Materials. CGIV 2010 -- Fourth European Conference on Colour in Graphics, Imaging, and MCS/10 Vision 12th International Symposium on Multispectral Colour Science, 83-87.

5. Westland, S.; Cheung, T. L. V.; Lozman, O. R., 2006. A metric for predicting perceptual blackness. 14th Color Imaging Conference Final Program and Proceedings, 14-17.

6. Kuehni, R.G., Color Space and Its Divisions, Color Order from Antiquity to the Present, 2003, Wiley and Sons, New Jersey.

7. Shamey, R., Cárdenas L. M., Hinks D., Woodard R., "Comparison of naive and expert subjects in the assessment of small color differences," J. Opt. Soc. Am. A 27, 1482–1489 (2010).

8. Kuehni, R.G., "Variability in estimation of suprathreshold small color differences," Color Res. Appl. 34, 367–374 (2009).

5. Perception of Black Objects

Renzo Shamey[1]*, J.R. Clonts Haslup
[1]*Polymer and Color Chemistry Program, North Carolina State University, Raleigh, NC 27695-8301, USA*
[*]*Corresponding author: rshamey@ncsu.edu*

Abstract

Blackness may be considered a fundamental concept in vision and color science yet very little has been published on the topic. From a photometric stand point black is located at the zero point on the scale of luminance. In terms of apparent brightness it is also thought of as being located at the zero point on this subjective scale. Color atlases such as the Munsell Book of Color also locate black at the zero point on the Munsell Value scale and assume that it does not make a contribution to color appearance. However, it has been established that blackness certainly impacts the appearance of colors. Indeed there may be as many as several hundred shades of what may be considered nominally black. This work reports analysis of psychophysical experimental results aimed at elucidating the perception blackness as a result of variations in colorimetric attributes, namely hue and chroma, and presents the methodology employed for the development of a blackness index.

Keywords: Blackness, blackness index, acceptability volume, visual tolerance, psychophysical assessment of blackness.

OCIS codes: 330.1720, 330.5020.

1. Introduction

Despite the fact that blackness may be considered a fundamental concept in vision and color science it has surprisingly been largely ignored as a research topic. In a recent review the historical discussions surrounding the subject matter were examined and studies aimed at elucidating blackness from a physiological as well as psychological point of view were briefly discussed [1].

Theories about the perception of color at the turn of the twentieth century abounded; two of the well-known theories came from Ewald Hering [2] and Helmholtz [3]. Hering argued that black is a sensation similar to other colors and required contrast and reflectance from a surface to be perceived and that black was seen during a relaxation of the eye. Ladd-Franklin [4], supporting the work of Helmholtz, recognized that the sensation of black could be due to the cortical visual processes reaching a resting state, however, argued that the sensation of black was a constant background sensation, and not a distinct sensation caused by light. Ward stated that “the so–called sensation of the deepest black [Hering] has shown to be possible only under the influence of light” [5]. Titchener pointed out that black and white may be completely independent [6]. Ward stated that black is only sensed under the indirect action of light and thus black sensation must be preceded by a sensation of light [7]. Neifeld defined black as (“the psychic correlate of the absorption, by objects, of all the visible light–ray frequencies” and recognized that because of the Eigengrau, absolute black could only be seen in the context of spatial or temporal contrast with a bright light [8]. Michaels explained that shining a bright light on a black object causes the object to appear dark grey and supported Neifeld's statement that spatial or temporal contrast is required to obtain a perfect black [9]. Dimmick argued that transitions from black to white pass through a neutral grey point [10]. Rich pointed out that black and white could not be truly opposite colors since adaptation to one over a long period would eventually result in seeing a neutral grey [11]. Based on results of a series of experiments Dimmick supported Rich's hypothesis that a neutral grey existed as an elementary color [12]. Deane Judd argued that black and white apply to surfaces and stated that they are characterized by being completely opaque and matte, differing only in lightness [13-14]. He then suggested that the appearance of blacks was based on the surface mode, and that this was in agreement with the works of Helmholtz [3], Troland [15], and Neifeld [8]. Heggelund proposed a model to describe the achromatic colors in which black and "luminous" were opponent colors [16-17]. W.D. Wright pointed out that the separation of television color signals into a luminance channel and a chrominance channel, and their subsequent combination, would result in the regeneration of the original signal [18]. Wright also indicated that blackness is a subjective perception and that cannot be measured on any simple photometric scale [19]. Quinn, Wooten, and Ludman revived the issue of achromatic colors in 1985 and determined that grey is a mixture of black and white, which should not be treated as a separate elemental color [20]. Several workers also examined blackness from a spatial and temporal induction viewpoint [21-25]. It was found that blackness was a function of the luminances of the reference and inducing stimuli, regardless of the chromatic content of either. It was also found that induced blackness was additive [26-27]. Caivano defined five "primary cesias": transparence, specular reflection, translucence, diffuse reflection, and absorption and eight scales of cesia: white–black, specular–black, translucent–black, transparent–black, specular–white, transparent–translucent, specular–transparent, and white–translucent [28]. Stockman, Plummer and Montag used flicker photometry to attempt to elucidate the achromatic visual pathways [29]. Their research indicated that seven mechanisms are involved in the perception of luminance. In 2009, Bimler, Paramei, and Izmailov found hue and saturation shifts in spatially induced blackness [30].

1.1. Efforts Aimed at Development of a Blackness Index

A few one-dimensional scales of blackness have been proposed for use in various industries. The Kenya Bureau of Standards, for example, recommends a method for specifying the blackness of pencil leads based on reflectance measurement of thirty lines drawn 1.5 mm apart on a white drawing paper with a reflectance of 79±2% [31]. This method gives values generally between 60 and 80 percent reflectance. At least one patent recommends rating blackness as a function of the lightness of a sample [32].

In the 1970's and 1980's, several blackness indices were suggested for use in the German printing industry [33-36]. As these were created for specific applications they mostly dealt with blackness in terms of lightness, although a few also incorporated chromaticity terms. These formulae included an assumption that bluish blacks appear deeper than brownish blacks of similar lightness values. Westland, et al reported work on the density of the ink as a metric for blackness [37].

Earlier work from our group focused on the impact of hue on the perception of blackness [38]. This work determines the combined impact of chroma and hue on the perception of black objects. The luminance factor was kept low (below 20) to generate nominally black samples and the values were then kept relatively constant for the purposes of this experiment.

2. Method

A. Sample Preparation and Viewing Conditions

Several hundred black samples were produced using dyed wool and acrylic fabrics. It was found that exact variations in chroma and hue among nominally black woolen fabrics could not be effectively produced. Acrylic substrates were thus selected as the amount of dye used to saturate the fiber can be carefully determined and moreover highly chromatic sets of cationic dyes, needed to create variations in hue and chroma, are available for the coloration of acrylic. Two black dyes at two concentrations were employed to initially dye the fabric. The black fabric was then cut into smaller pieces and over-dyed in the Pyrotec laboratory dyeing machine with one of three concentrations of a trichromatic mixture to produce twelve nominally black color triangles. From the color triangles generated 30 samples with L* values between 10.5 and 12 were selected and mounted onto medium grey plastic backings and were precision cut to 2” × 2” square dimensions. The dye recipes and colorimetric values for samples are given in Table 5.1. Each dyebath contained 3 gL^{-1} sodium acetate, 50 gL^{-1} Glauber’s salt, and 5 mLL^{-1} glacial acetic acid in addition to the black or chromatic dyes. The cationic dyes used were Permacryl Black FB–L, Maxillon Black RM 200, Sevron Blue GBR, Sevron Red GBL (C.I. Basic Red 104), and Sevron Yellow 6DL (C.I. Basic Yellow 29).

Twenty-four samples were selected such that they comprised three concentric hue circles of eight evenly spaced samples each (C* ranges of 0.42-0.57, 0.89-0.97, and 1.58-1.86). Six additional samples, with C* between either 0.12 and 0.20 or 3.34 and 3.46, were also used to provide a few samples in an extended experimental range as control. The samples were divided into sets according to chroma: A (C* = 0.12-0.20), B (C* = 0.42-0.57), C (C* = 0.89-0.97), D (C* = 1.58-1.86), and E (C* = 3.34-3.46). All samples had similar texture and were displayed at the same orientation to avoid variations in psychophysical assessments due to potential variations in scattering of light.

Table 5.1. Dye recipes and colorimetric data for samples in this experiment. All concentrations are given as percent on weight of fabric.

Sample	Black Dye	Black	Red[a]	Blue[b]	Yellow[c]	L^*	a^*	b^*	C^*	$h°$	X	Y	Z
A 1	1	6.5	—	0.30	0.20	11.24	0.12	0.16	0.20	54.65	1.198	1.243	1.338
A 2	2	6.5	—	0.40	0.10	11.02	–0.12	0.01	0.12	174.50	1.222	1.276	1.352
A 3	2	6.5	0.15	0.45	0.15	10.70	0.15	–0.10	0.18	326.13	1.222	1.291	1.368
B 1	1	6.5	—	0.10	0.15	10.87	0.53	0.17	0.56	17.42	1.160	1.233	1.326
B 2	2	6.5	—	0.15	0.10	11.11	0.28	0.43	0.51	56.45	1.171	1.246	1.377
B 3	1	6.5	—	0.45	0.30	11.21	–0.14	0.42	0.44	107.87	1.107	1.171	1.306
B 4	2	6.0	0.45	—	0.30	10.80	–0.38	0.19	0.42	153.46	1.123	1.173	1.321
B 5	2	6.0	0.20	—	0.05	10.89	–0.42	–0.23	0.48	208.32	1.190	1.235	1.360
B 6	2	6.0	0.45	0.15	0.15	10.34	–0.21	–0.37	0.43	240.34	1.225	1.260	1.364
B 7	2	6.0	0.60	—	0.15	10.36	0.25	–0.51	0.57	296.31	1.242	1.286	1.337
B 8	1	6.5	—	0.15	0.10	10.81	0.53	–0.18	0.55	341.52	1.243	1.309	1.339
C 1	1	6.0	0.10	—	0.15	10.99	0.89	0.09	0.90	5.99	1.213	1.298	1.351
C 2	2	6.5	0.30	0.15	0.30	11.18	0.63	0.72	0.96	48.64	1.128	1.217	1.327
C 3	2	6.5	—	0.05	0.20	11.34	–0.05	0.97	0.97	93.13	1.138	1.221	1.371
C 4	2	6.5	—	0.45	0.30	11.26	–0.66	0.70	0.96	133.10	1.136	1.198	1.389
C 5	2	6.0	0.10	0.05	0.10	10.68	–0.96	–0.03	0.96	181.75	1.210	1.247	1.404
C 6	2	6.0	0.30	0.30	0.15	10.71	–0.75	–0.48	0.89	212.65	1.212	1.219	1.317
C 7	2	6.0	0.40	0.10	—	10.54	–0.06	–0.97	0.97	266.54	1.236	1.255	1.281
C 8	2	6.5	0.20	0.30	—	10.90	0.80	–0.52	0.96	326.78	1.266	1.335	1.298
D 1	2	6.5	0.60	—	0.15	10.70	1.73	0.12	1.73	3.98	1.158	1.265	1.302
D 2	1	6.0	0.30	—	0.45	10.96	1.38	1.00	1.71	35.92	1.154	1.274	1.376
D 3	1	6.5	—	0.15	0.60	11.52	–0.10	1.76	1.76	93.32	1.071	1.173	1.356
D 4	2	6.0	0.1	—	0.40	11.03	–1.47	0.87	1.70	149.49	1.162	1.218	1.468
D 5	2	6.0	—	0.30	0.20	11.09	–1.85	0.12	1.86	176.37	1.182	1.203	1.410
D 6	2	6.0	—	0.75	—	10.36	–1.51	–0.91	1.76	210.92	1.234	1.295	1.396
D 7	1	6.0	—	0.75	—	10.69	0.14	–1.58	1.58	275.08	1.196	1.348	1.276
D 8	1	6.5	0.20	0.30	—	10.58	1.26	–1.14	1.70	317.90	1.197	1.264	1.375
E 1	2	6.0	—	—	0.75	11.61	–2.70	2.17	3.46	141.17	1.163	1.219	1.336
E 2	1	6.0	0.75	—	—	11.00	3.02	–1.43	3.34	334.67	1.294	1.261	1.503
E 3	1	6.5	0.75	—	—	10.28	3.19	–1.25	3.43	338.53	1.201	1.163	1.374

Black Dyes Used: 1- Permacryl Black FB–L, 2- Maxillon Black RM 200
Trichromatic Dyes: a- Sevron Blue GBR, b- Sevron Red GBL (C.I. Basic Red 104), and c- Sevron Yellow 6DL (C.I. Basic Yellow 29).

A calibrated viewing booth, (Macbeth SpectraLight III, X-Rite), was illuminated with a filtered tungsten daylight simulating lamp with a correlated color temperature of 6500±200 K and constant illuminance of approximately 1400 lx in the middle of the display board. All the plastic components were uniformly spray painted to $L^*_{ab,10}$ of 74, which is approximately equivalent to Munsell N 7.25 and matches the interior of the viewing booth. All extraneous light was eliminated and the illumination conditions were carefully controlled during the experiment in order to minimize variability. The samples were displayed on a custom built 45° easel and viewing was set normal to the plane of the display.

B. Psychophysical Assessment Method

One hundred color normal subjects, having passed the Neitz test for color vision [39], including fifty men (mean age = 27) and fifty women (mean age = 24.5) completed two tasks three times each on separate sittings and with at least 24 hour gap between assessments. The average age of the entire group was 25.8. A total of 18,000 visual assessments were completed over a period of one year. Each subject wore a mid-gray laboratory coat and a pair of gloves to minimize color variability of the surround during the course of the experiment and to prevent damaging the samples. The samples were placed by the experimenter who also wore a mid-gray laboratory coat. At the beginning of the experiment, the subject viewed the empty illuminated viewing booth for 2 minutes to adapt to the light source, during which time the experiment was explained.

In the first task, viewers were randomly presented with each of the thirty samples and were asked to categorize each sample as either “black” or “not black.” Subjects were told that no restrictions existed in determining the number of black samples and that all, some, or none of the samples may be selected as black. In the second task, a reference black (an 'ideal black' that was essentially a light trap) was placed in the viewing booth and used to rate each sample on a custom scale. The reference black comprised a wooden hollow cube of 6” × 6” × 6” dimensions mounted at 45^{o} with a 2” square aperture cut in the center of the plane facing the subject. The interior of the cube was lined with black velvet and the exterior was painted grey to approximately Munsell N 7.25 to match the interior of the viewing booth. No, or minimal, light could escape the black box. The arrangement is shown in Figure 5.1. Subjects were asked to rate each of the thirty samples using a scale where the reference black was assigned a rating of zero. Subjects were instructed that a rating of 10 should be given to a borderline black sample and, consequently, that all samples that were not perceived as black should be rated 11 or higher. No endpoint was defined for the scale. Subjects were allowed to respond with zero if they felt the sample matched the blackness of the reference box perfectly. Samples were again presented individually in a different random order from the first task. The exact instructions given to observers are given in Appendix A.

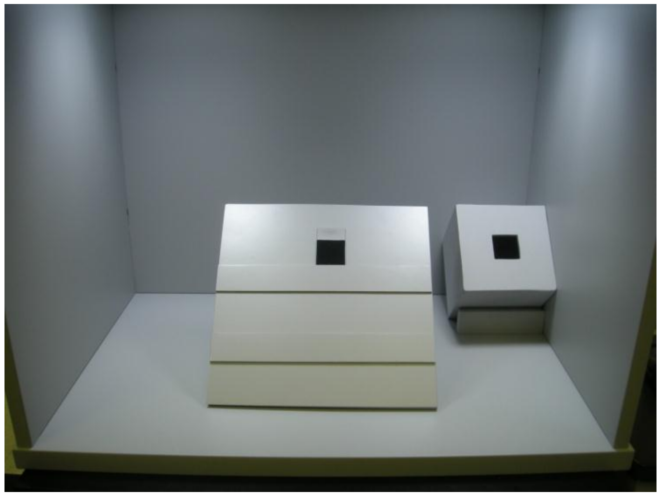

Figure 5.1. 45/0 illumination viewing geometry employed for psychophysical assessment of black samples and the box used as the “ideal black”.

Finally, a small trial was conducted to determine the lightness limits of blackness. Ten color normal observers were asked to classify thirty-one samples as black or not black. The samples were all neutral ($a^* = b^* = 0$), with lightness values ranging from 5 to 35 in one unit increments. The samples were presented in a calibrated monitor, one at a time, in random order. Each observer’s average limit for blackness was calculated, and a grand average was determined.

3. Results and Discussion

The results from the first task were analyzed using a binomial approach as well as a simple linear regression. A linear regression model was used to analyze the results of the second task. The standard deviations of the mean rating given in the second task as well as confidence intervals for the mean ratings were calculated. Each of these analyses gave a preliminary model for an index of blackness. Finally, the results from the two tasks were compared, and the multivariate model for mean rating was tested against the results from the first observation trial. Microsoft Office Excel was used for the calculations of standard deviation and confidence intervals. A traditional two tail paired t-test was used in this experiment, as the ratings of each sample were independent from each other. JMP software was used for the creation of the binomial and linear regression models.

The goal of task one was the elimination of all samples considered to be too distant from what would generally be considered black. Regardless of the response during the first task all thirty samples were presented during the second. For this reason, it was decided to treat the first task as a separate experiment to help determine the boundaries of blackness in color space. Figure 5.2 shows the percent of subjects identifying each sample black. The samples are divided by chroma group for clarity. The figure shows that samples that had relatively higher chroma were considered black less often than samples of similar hue at lower chroma, as expected. Several color difference models, including DECMC, CIE94 and more recently CIEDE2000 have incorporated functions to improve the performance of the models in achromatic regions and account for the reduced observer tolerance for variations in colorimetric attributes in this region. It can be seen that despite the relatively small C* range amongst samples tested, subjects' acceptability of blackness was highly influenced by small variations in chroma. It can also be seen that more subjects selected samples with hue angles (h°) between 200° and 270° as black compared to other regions of the CIELAB color space.

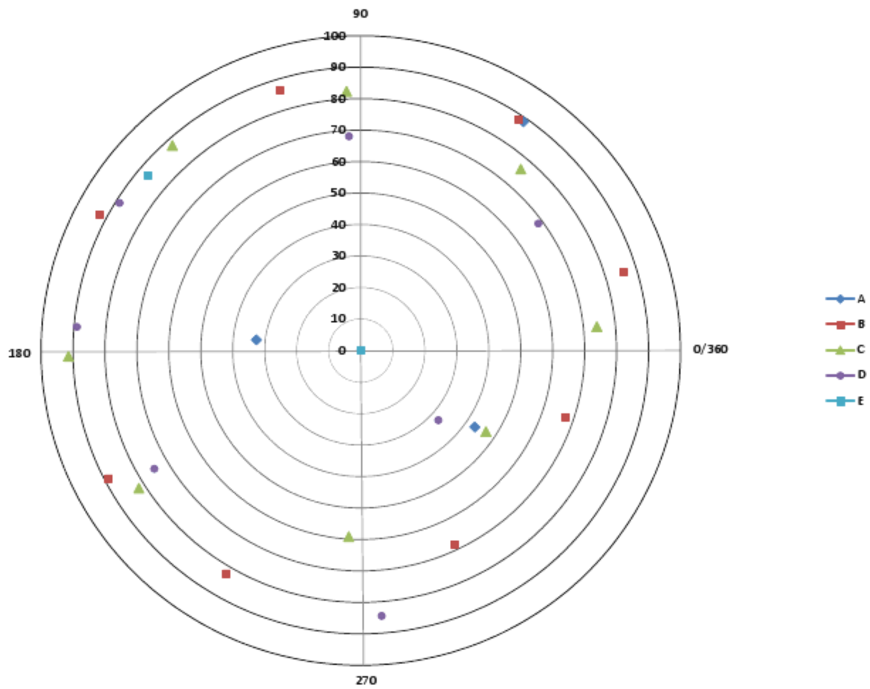

Figure 5.2. Percent of observers considering each sample black.

Task two enabled assessment of perceived blackness as a function of hue and chroma. Figure 5.3 shows a graph of blackness rating as a function of hue angle for chroma groups A-E. The lower the average blackness rating, the blacker a sample is perceived. Samples with relatively higher chroma were perceived to be less black than those with similar hues at lower chroma, and bluish-green samples were rated blacker than those with other hues, regardless of chroma.

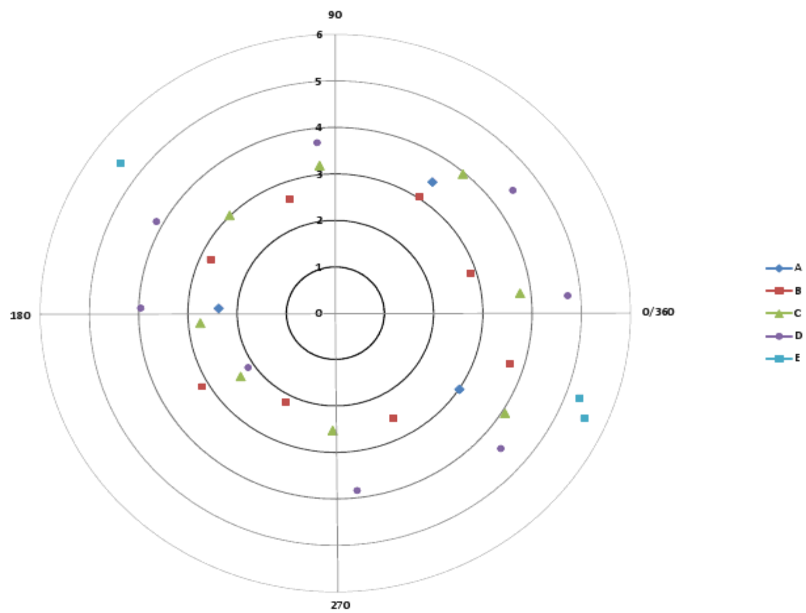

Figure 5.3. Mean perceived blackness ratings as a function of hue angle, *h*°, for chroma sets A-E.

Comparison of the two tasks in this experiment showed good agreement. Figure 5.4 illustrates the concordance between experiments. In this figure, dark gray (N=388, 92.16%) squares indicate sample pairs for which both tasks agreed—that is, the sample that was called black more often was assigned a lower mean rating. Light gray, (N=5, 1.18%) squares indicate sample pairs that tied in one or both of the tasks and therefore could not be compared. Medium gray, (N=28, 6.65%) squares indicate sample pairs where the tasks disagreed on. The white squares are placeholders so that each sample pair is considered only once.

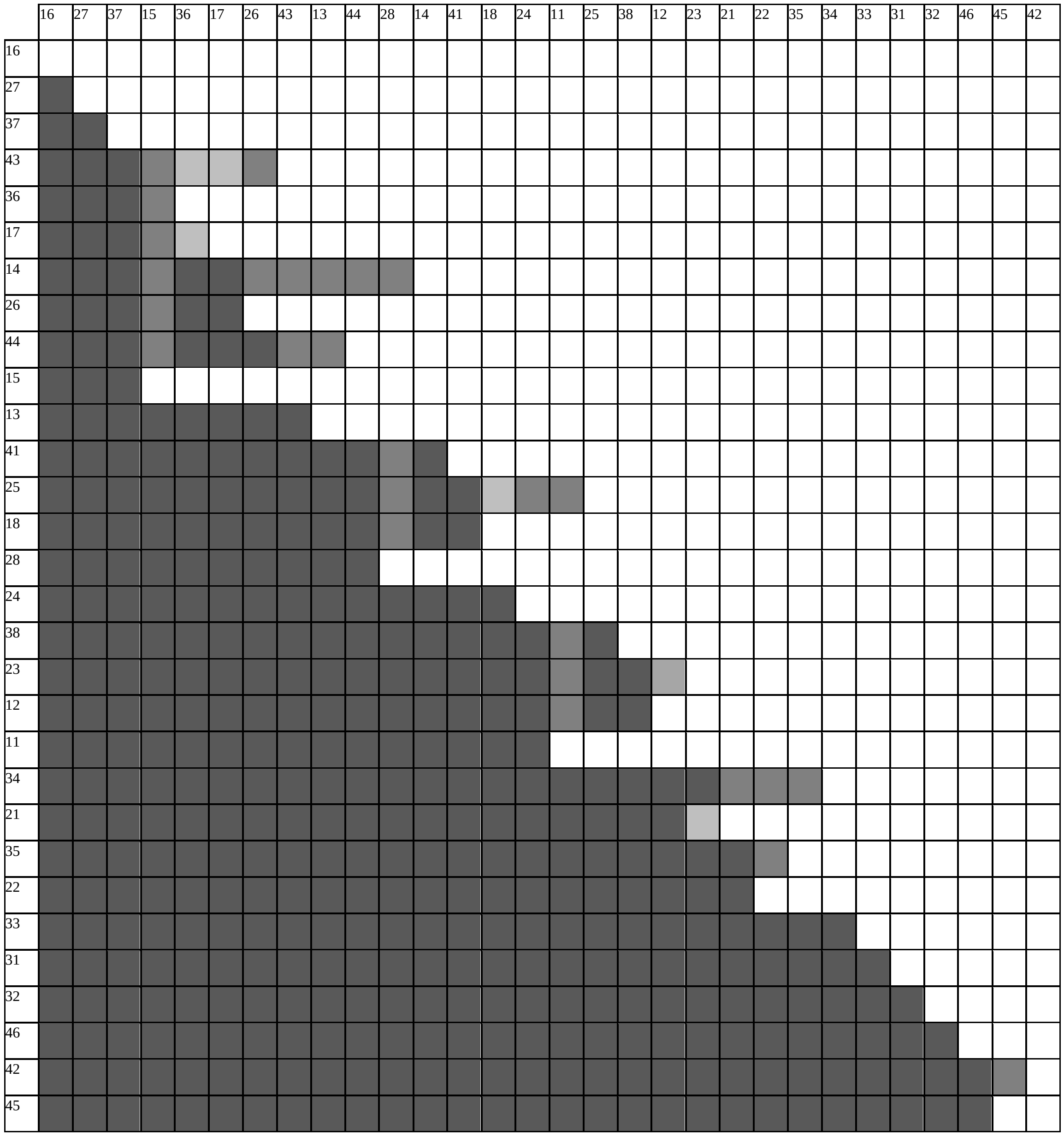

Figure 5.4. Agreement in blackness assessments between tasks one and two. Dark gray cells denote agreement in responses, light gray cells show pairs that tied in one or both of the tasks and therefore could not be compared, and medium gray cells indicate sample pairs where the results from two tasks disagreed.

Two statistical models were obtained based on the results of this analysis. The first model gives the percent of observers who consider each sample to be black, which may be considered the acceptability percentage. The second gives a model of the mean assigned rating of each sample, which may be equivalent to degree of blackness, the higher the value the less black the sample. Both models are functions of luminance factor (L^*) and the colorimetric attributes a^* *and* b^*. The models are shown in Equations 1 and 2. The correlation coefficients for these models are 0.93 and 0.75 respectively.

$$\%BK = 142.261 - 5.4L^* - 4.4678a^* - 8.0715b^* - 4.5625a^{*2} - 1.4b^{*2} \text{ and } 0 \leq BK \leq 100 \quad (1)$$

$$NCSU_{BK} = 4.64L^* + 4.8436a^{*2} + 3.81b^{*2} \quad \text{where} \quad 0 \leq NCSU_{BK} \leq 100 \quad (2)$$

The models were created by performing linear regressions in JMP software. Models were tested using L^*, a^*, b^*, a^{*2}, b^{*2}, and a^*b^* values. While in both models, the b^{*2} term was determined to be insignificant, it was decided to incorporate b^{*2} term to enable inclusion of colorimetric attributes of samples to be tested. The a^*b^* term, however, was determined to be insignificant in both cases. The inclusion of second order terms implies that there is a maximum blackness at a given point in the CIELAB color space, and that moving away from that maximum in any direction will decrease the perception of blackness. The location of maximum blackness is not significantly different from the neutral point at a luminance factor of zero in either of these models. Using subjects' responses, a threshold value for blackness boundary was established.

The %BK shown in Equation 1 determines a value based on responses from 18,000 observations recorded in this study. The %BK model should be interpreted such that values above 100% are capped to 100%, and negative results should be capped to zero. When interpreted in this manner, the equation becomes zero around a lightness value of 26.5 for neutral samples. The results of the psychophysical assessments involving the determination of acceptable blackness boundary based on varying L* verify this conclusion: no sample with a lightness value above 27 was ever determined to be black.

The blackness ratings model, denoted as $NCSU_{BK}$, (where everything rated above 100 is considered to be not black) is also verified by the results of the third experiment. The grand average line between black and not black falls at a luminance factor (L^*) of 22. The model gives a value of 100 at L* of approximately 21.5, with everything lighter being rated above 100 (or not black).

Finally, the ratings model was used to predict ratings for the samples to test the fit. The results are shown in Figure 5.5. The general shape of the model fits relatively well, although the actual samples tend to be rated closer to absolute black than the model predicts. The samples' lightness values are low enough that they are perceived to be blacker than the model would predict, and this will be examined in future work.

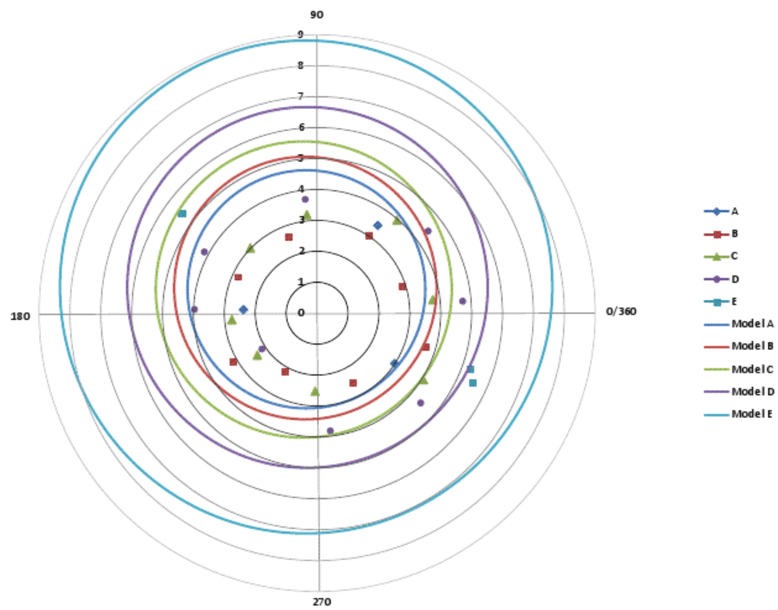

Figure 5.5. Fit of models to blackness data.

Conclusions

Blackness of objects is dependent on their colorimetric attributes, i.e., lightness, hue and chroma and not confined to a zero point on the luminance factor or apparent brightness scales. Small variations in lightness, chroma and hue significantly influence the perceptual magnitude of blackness. Of interest is the determination of a boundary of a "blackness volume" in the color space based on acceptability judgments from a statistically acceptable number of subjects. Currently no firm foundations for the determination of such a boundary based on observer responses exist.

Analysis of the data from 10 color normal subjects indicates that surfaces with a luminance factor greater than 27 are not considered black. Accordingly samples with $L^* > 27$ are considered to be outside of the blackness boundary in the CIELAB space. However, the exact boundaries of the blackness volume will be further tested in a future study.

In addition, results from 100 subjects providing over 18,000 assessments indicate that, irrespective of chroma, on average samples with hue angles between approximately 200° and 270° were perceived to be the most black, i.e., cyan to bluish-blacks. Blacks with hue angles above 315° or below 45° (reddish-blacks) were considered to be the least black and the ratings trended between the most and least blacks as a function of hue angle. In general, the blackness rating was inversely proportional to C* for the samples that were not greenish- to bluish-black. Hence, for the observers studied, increasing C* has a deleterious effect on perception of blackness for all samples except greenish blacks and bluish blacks. A preliminary index of blackness, $NCSU_{BK}$, is proposed that may be used to indicate the degree of blackness of surfaces with lightness values below 27.

References

[1] Shamey R, Clonts Haslup JR, Perception and Quantification of Blackness, Vision Research, *Submitted*

[2] Hering, E., 1964. Outlines of a Theory of the Light Sense; Harvard University Press: Cambridge, MA, 1964.

[3] Helmholtz, H. v., 1962. Handbook of physiological optics. J. P. C. Southall, (Ed. and trans.), Dover: New York.

[4] Ladd-Franklin, C., 1932. Colour and Colour Theories; Harcourt, Brace and Company: New York.

[5] Ward, J., 1905. Is 'Black' a Sensation? Brit. J. Psychol. 1, 407-427.

[6] Titchener, E. B., 1916. A Note on the Sensory Character of Black. Journal of Philsophy: Psychology and Scientific Methods 13, 113-121.

[7] Ward, J., 1916. A Further Note on the Sensory Character of Black. Brit. J. Psychol. 8, 212-221.

[8] Neifeld, M. R., 1924. The Ladd-Franklin Theory of the Black Sensation. Psychol. Rev. 31, 498-502.

[9] Michaels, G. M., 1925. Black: A Non-light Sensation. Psychol. Rev. 32, 248-250.

[10] Dimmick, F. L., 1920. A Note on the Series of Blacks, Greys, and Whites. American Journal of Psychology 31, 301-302.

[11] Rich, G. J., 1926. Black and Grey in Visual Theory. Amer. J. Psychol. 37, 123-128.

[12] Dimmick, F. L., 1933. The Psychophysical Determination of the Limits of Pure Gray. Amer. J. Psychol. 45, 313-314.

[13] Judd, D. B., 1940. Hue, saturation, and lightness of surface colors with chromatic illumination, J. Opt. Soc. Amer. 30, 2-32.

[14] Judd, D. B., 1941. The Definition of Black and White. J. Psychol. 54, 289-294.

[15] Troland, L. T. (1921). Thc enigma of color vision. American Journal of Physiological Optics 2, 317-337.

[16] Heggelund, P., 1974. Achromatic color vision I: Perceptive variables of achromatic colors. Vision Research 14, 1-1079.

[17] Heggelund, P., 1974. Achromatic color vision II: Measurement of simultaneous contrast within a bidimensional system. Vision Research 14, 1081- 1088.

[18] Wright, W. D., 1980. Perception of Blackness. Die Farbe 28, 161-166.

[19] Wright, W.D., 1981. The nature of blackness in art and visual perception. Leonardo 14, 236- 237.

[20] Quinn, P. C.; Wooten, B. R.; Ludman, E. J., 1985. Achromatic Color Categories. Perception and Psychophysics 37, 198-204.

[21] Werner, J. S.; Cicerone, C. M.; Kliegl, R.; DellaRosa, D., 1984. Spectral efficiency of blackness induction. J. Opt. Soc. Am. A 1, 981-986.

[22] Cicerone, C. M.; Volbrecht, V. J.; Donnelly, S. K.; Werner, J. S., 1986. Perception of Blackness. J. Opt. Soc. Am. A 3, 432-437.

[23] Fuld, K.; Otto, T. A.; Slade, C. W., 1986. Spectral responsivity of the white-black channel. J. Opt. Soc. Am. A 3, 1182-1188.

[24] Volbrecht, V. J.; Werner, J. S.; Wooten, B. R., 1989. Temporal Induction of Blackness--I. Color Appearance. Vision Res. 29, 1425-1436.

[25] Volbrecht, V. J.; Werner, J. S.; Wooten, B. R., 1989. Temporal Induction of Blackness--II. Spectral Efficiency and Tests of Additivity. Vision Res. 29, 1437-1455.

[26] Volbrecht, V. J.; Werner, J. S.; Cicerone, C. M., 1990. Additivity of spatially induced blackness. J. Opt. Soc. Am. A 7, 106-112.

[27] Shinomori, K.; Schefrin, B. E.; Werner, J. S., 1997. Spectral mechanisms of sptially induced blackness: data and quantitative model. J. Opt. Soc. Am. A 14, 372-387.

[28] Caivano, J. L., 1993. Appearance (Cesia): Variables, Scales, Solid. Farbe 39, 115-125.

[29] Stockman, A.; Plummer, D. J.; Montag, E. D., 2005. Spectrally opponent inputs to the human luminance pathway: slow +M and −L cone inputs revealed by intense long-wavelength adaptation. The Journal of Physiology 566, 61-76.

[30] Bimler, D. L.; Paramei, G. V.; Izmailov, C. A., 2009. Hue and saturation shifts from spatially induced blackness. J. Opt. Soc. Am. A 26, 163-172.

[31] Kenya Bureau of Standards, 2005. Black lead pencils - specification. KS 90:2005.

[32] Nakajima, Y.; Yukawa, Y.; Kamimori, I.; Yamanouchi, A.; Hoshida, Y., 2007. European Patent 1 270 624 B1.

[33] Schumacher, W., 1976. Photometrische Schwarzmessung. pp 582-587.

[34] Bode, R.; Ferch, H.; Koth, D.; Schumacher, W.. 1979. Schwarzgradskala für Pigmentrußе. Farbe + Lack 85, 7-13.

[35] Gläser, F., 1978. Probleme bei der Farbmessung geschönter Schwarzdrucke. Farbe + Lack 84, 579-584.

[36] Lippok-Lohmer, K., 1986. Praxisnahe Schwarzmessungen. Farbe + Lack 92, 1024-1029.

[37] Westland, S.; Cheung, T. L. V.; Lozman, O. R., 2006. A metric for predicting perceptual blackness. 14th Color Imaging Conference Final Program and Proceedings, 14-17.

[38] Clonts, J. R.; Shamey, R.; Hinks, D., 2010. Effect of Colorimetric Attributes on Perceived Blackness of Materials. CGIV 2010 -- Fourth European Conference on Colour in Graphics, Imaging, and MCS/10 Vision 12th International Symposium on Multispectral Colour Science, 83-87.

[39] Neitz, J. Manual: Neitz test of color vision (Western Psychological Services, 2001).

6. Assessment of inter and intra-instrument variability in the measurement of very dark surfaces

Reid Clonts Haslup, Renzo Shamey, David Hinks
Department of Textile Engineering Chemistry and Science, North Carolina State University, Raleigh, NC 27695-8301

Abstract

Little attention has been paid to date to the precision and accuracy of the measurement of very dark (black or near black) objects. When measuring very dark objects, the signal to noise ratio is low due to the low reflectance of these objects. Because of the commercial importance of black goods, it is crucial to adequately control their measurement. It is to be expected that black goods will be more variable in measurements than lighter, brighter colors.

In a three part experiment, the repeatability and reproducibility of the measurement of very dark objects were determined and tested. Thirty-one samples were measured on a total of eleven spectrophotometers, and the 95% confidence intervals for the measurements were calculated and verified. These confidence intervals for black samples were found to be larger than the confidence intervals for lighter, brighter samples, and one sample measured on two different spectrophotometers could show variability up to a ΔE_{ab} of two units. Confidence intervals were also calculated for spectral responses of the spectrophotometers every twenty nanometers in wavelength across the visible spectrum. The raw reflectance data may vary by approximately 0.2% between spectrophotometers for any given wavelength.

1. Introduction

Despite several decades of work and systematic study on color measurement, the measurement of very dark objects is still subject to significant error. With any analytical instrument, the signal-to-noise ratio (S/N) is used to compare the desired signal to the background noise inherent in a measurement.[1] When dealing with spectrophotometers, the signal may be interpreted as the light reflected from or transmitted through the object being measured, while the noise arises from stray light or absorption due to defects in the sphere coating or variations in the illumination and detection mechanisms. As a signal becomes weak the significance of noise increases, and if the signal is weak enough the S/N ratio approaches unity and measurements become highly unreliable. In reflectance spectrophotometers when the object reflects a small amount of light, such as in dark objects, the signal becomes relatively weak. This can have significant implications on the accuracy of measurements and reliability of data.[1] The usefulness of any data is highly dependent on reliability of measured values. Any calculation is only as reliable as its inputs.

Due to increasing significance of global sourcing in manufacturing colored goods and the need to accurately communicate color, and since black is one of the most widely produced colors, accuracy in determining the degree of blackness of colorants, inks and finished colored goods is of primary importance. While the determination of inter- and intra-instrument agreements and repeatability in the assessment of reflectance measurement of colored objects is always important, it can be critical for the measurement of dark objects, especially black materials.

Previously, a measurement study was conducted on spectrophotometers in a round robin inter-laboratory assessment[2]. This measurement study consisted of twenty-three sample pairs, only four of which had lightness values, L^*_{ab}, below 30. Of these four, three were highly chromatic, and only one black sample was tested. Furthermore, this study focused heavily on the confidence intervals found on ΔE^*_{ab} values, and not on the confidence intervals around the absolute colorimetric values. While for general spectrophotometric use the confidence intervals calculated in this study may be considered sufficient, for the absolute measurement of very dark objects additional examination would be required. The current study focused on the absolute measurement of nominally black samples (that is, samples that would be called black when viewed individually by a group of observers).

2. Experimental

Three separate experiments were undertaken to determine the repeatability and reproducibility of measurements taken of near-black objects. In the first experiment, twenty-one samples were measured ten times each on three spectrophotometers, including two bench top and one handheld unit, using specular included, UV excluded, 30 mm apertures for the desktop units and 20 mm aperture for the handheld unit, and CIE 1964 supplementary standard colorimetric observer conditions. In addition four spectrophotometers (using two desktops and two handheld units) were employed, using the same settings with the exception that the specular component of light was excluded. The second handheld spectrophotometer was only capable of specular excluded measurements.

In the second experiment, twenty-six samples comprising thirteen sample pairs were measured eight times each on six desktop spectrophotometers using specular included as well as excluded (separately), UV excluded, and 1964 supplementary standard colorimetric observer conditions. Aperture size was kept constant at 30 mm, with the exception of one unit that utilized a 20 mm aperture. Spectra were recorded for each trial in this experiment, in addition to colorimetric values.

In the third experiment, four samples were measured on nine desktop spectrophotometers once each, one spectrophotometer twice, and one spectrophotometer three times to check the repeatability and reproducibility parameters determined in the first two experiments. Again, spectra as well as colorimetric values were recorded in this experiment.

For the first experiment, twenty-one nominally black samples of varying materials were employed. The samples varied in hue, lightness, and chroma, but were all considered to be black when viewed individually in a Macbeth Spectralight III viewing booth under a calibrated daylight simulator. The general specifications of the samples are given in Table 6.1.

Table 6.4. Sample identification for the first and second experiments.

Sample Code	**Sample Identity**
A1	Commercial single-faced polysuede fabric
A2	Commercial double-faced polysuede fabric
B1	Commercial boiled woven wool fabric
E1	"Domino" paint chip from Sherwin Williams
B2	Commercial wool-cashmere felt fabric
H1	Commercial gold-selvedge velvet fabric
I1	"Greenblack" paint chip from PPG
N	Woven nylon, dyed in-house (Only in Experiment 1)
J1	Glossy automotive black paint on metal substrate from DuPont
J2	Glossy automotive black paint on plastic substrate from DuPont
I2	Commercial black posterboard
K1	Metallic automotive black paint on metal substrate from DuPont
K2	Metallic automotive black paint on plastic substrate from DuPont
G2	Commercial stretch velour
M1	Glossy black calibration tile from Datacolor
M2	Black calibration trap from Datacolor
E2	"Tricorn" paint chip from PPG
C2	Commercial black vinyl
H2	Commercial white-selvedge velvet fabric
L1	Woven wool, dyed in-house (A)
L2	Woven wool, dyed in-house (B)
C1	Woven cotton broadcloth, dyed in-house (Only in Experiment 2)
D1	Woven cotton corduroy, dyed in-house (Only in Experiment 2)
D2	Knit cotton fabric, dyed in-house (Only in Experiment 2)
F1	Woven cotton duck fabric, dyed in-house (Only in Experiment 2)
F2	Woven cotton sateen fabric, dyed in-house (Only in Experiment 2)
G1	Woven cotton flannel, dyed in-house (Only in Experiment 2)

The commercial fabrics were obtained from various fabric stores in small quantities while some of the fabrics were also dyed in-house. The paint samples were ordered from various manufacturers. A glossy black tile as well as a calibration light trap was also used in the study. The automotive paint samples were paired; that is, the glossy paint was the same on both the metal and plastic substrates and these samples visually matched. The wool samples dyed in house were not paired, and did not visually match.

The measurement protocol involved a single operator measuring each sample in ten different sittings, with a minimum of two hours (while the spectrophotometer was turned off) between each sitting. Samples were measured in a random order after each instrument was allowed to warm up and calibrated with the calibration set provided by the manufacturer. Samples were measured four times each with 90° rotation between each measurement, and their average colorimetric values, L*, a*, and b*, were recorded. The mean colorimetric values for each sample are given in Table 6.2.

Table 6.5. Colorimetric values for all samples from the first experiment.

Sample	**Specular Excluded**			**Specular Included**		
	L*	**a***	**b***	**L***	**a***	**b***
A1	9.09	-0.29	-2.95	19.07	-0.28	-3.02
A2	6.20	1.09	-1.20	16.14	1.08	-1.28
B1	12.98	0.29	0.21	12.97	0.29	-0.28
E1	7.79	1.03	-1.71	27.22	-1.04	-1.82
B2	12.05	0.26	-0.95	12.08	0.24	-0.97
H1	2.56	0.11	-0.24	2.72	0.10	-0.23
I1	30.14	-1.03	-0.80	29.03	-1.05	-0.83
N	21.77	-2.87	-2.06	21.84	-2.85	-2.09
J1	25.11	-0.09	-0.52	3.34	-0.17	-0.68
J2	25.14	-0.09	-0.52	3.05	-0.15	-0.63
I2	29.93	0.17	-1.51	29.31	0.15	-1.67
K1	26.87	0.17	-0.26	7.71	0.57	0.01
K2	26.86	0.17	-0.22	8.00	0.55	-0.16
G2	12.37	1.03	-1.30	12.54	0.90	-1.29
M1	3.91	0.04	0.12	0.72	-0.04	0.00
M2	0.22	-0.03	0.02	0.29	-0.04	0.07
E2	27.37	0.18	-0.76	26.03	0.25	-0.82
C2	25.88	0.04	-0.48	24.34	-0.01	-0.73
H2	3.96	0.78	-0.38	3.98	0.18	-0.36
L1	13.72	-0.02	-0.71	13.80	0.00	-0.75
L2	13.96	0.26	-0.99	14.07	0.27	-1.02

The majority of the samples used in the first experiment were also used in the second experiment, although one sample, N, was omitted due to unavailability. Six additional knitted and woven cotton fabric samples, dyed in the same dyebath and cut into two inch squares and mounted onto medium grey PVC backings, were also used as shown in Table 6.1. The average colorimetric values (L*, a*, and b*), along with the average color difference values (ΔL^*_{ab}, Δa^*, Δb^*, and ΔE^*_{ab}) are given in Tables 6.3 and 6.4, respectively, for specular excluded and specular included conditions. The color difference values were calculated for each pair of samples, using sample 1 as the standard and sample 2 as the batch (i.e., A1 is the standard and A2 is the batch.) Figures 6.1 and 6.2 illustrate the average spectra obtained for each sample in the second experiment.

Table 6.3. Absolute colorimetric values obtained in the second experiment.

	Specular Excluded			**Specular Included**		
Sample	**L***	**a***	**b***	**L***	**a***	**b***
A1	18.88	-0.39	-2.93	18.85	-0.38	-2.88
A2	15.97	1.00	-1.06	15.99	1.04	-1.05
B1	13.07	0.20	-0.18	13.03	0.23	-0.15
B2	11.87	0.17	-0.89	11.85	0.18	-0.85
C1	18.02	0.81	-0.82	17.97	0.81	-0.80
C2	24.23	-0.04	-0.78	25.80	0.02	-0.51
D1	13.29	0.49	-1.65	13.21	0.52	-1.60
D2	13.85	0.50	-0.94	13.83	0.52	-0.88
E1	27.22	-1.07	-1.81	27.65	-1.03	-1.72
E2	26.20	0.17	-0.82	27.23	0.15	-0.75
F1	18.32	0.85	-1.61	18.32	0.85	-1.56
F2	16.62	0.81	-0.45	16.61	0.83	-0.44
G1	14.12	0.77	-0.17	14.12	0.80	-0.15
G2	12.39	0.92	-1.14	12.14	0.93	-1.12
H1	2.44	0.11	-0.20	2.39	0.09	-0.21
H2	3.96	0.15	-0.36	4.02	0.19	-0.35
I1	29.11	-1.07	-0.87	29.95	-1.03	-0.81
I2	29.30	0.15	-1.63	29.86	0.19	-1.50
J1	4.15	-0.20	-0.98	24.95	-0.06	-0.51
J2	3.19	-0.18	-0.84	24.93	-0.07	-0.45
K1	8.53	0.53	-0.25	26.66	0.18	-0.22
K2	8.44	0.49	-0.40	26.64	0.17	-0.21
L1	13.36	-0.10	-0.72	13.31	-0.08	-0.68
L2	13.88	0.19	-0.97	13.80	0.21	-0.93
M1	0.47	-0.03	-0.21	23.65	0.06	0.06
M2	0.12	0.01	-0.03	0.07	0.02	-0.01

Table 6.4. Color Difference colorimetric values for the second experiment.

	Specular Included				Specular Excluded			
Sample Pair	**ΔL^*_{ab}**	**Δa^***	**Δb^***	**ΔE^*_{ab}**	**ΔL^*_{ab}**	**Δa^***	**Δb^***	**ΔE^*_{ab}**
A	-2.86	1.42	1.84	3.70	-2.91	1.39	1.88	3.74
B	-1.18	-0.05	-0.71	1.39	-1.20	-0.03	-0.71	1.41
C	7.83	-0.80	0.28	7.88	6.22	-0.85	0.04	6.28
D	0.62	0.00	0.72	0.98	0.56	0.01	0.71	0.95
E	-0.43	1.18	0.97	1.59	-1.01	1.24	0.98	1.88
F	-1.70	-0.02	1.13	2.05	-1.70	-0.05	1.16	2.06
G	-1.98	0.13	-0.97	2.28	-1.73	0.14	-0.97	2.05
H	1.63	0.10	-0.14	1.64	1.53	0.03	-0.16	1.54
I	-0.09	1.21	-0.69	1.40	0.17	1.23	-0.75	1.46
J	-0.02	0.00	0.03	0.09	-0.96	0.03	0.14	1.05
K	-0.03	-0.01	0.01	0.11	-0.09	-0.04	-0.15	0.55
L	0.49	0.29	-0.25	0.67	0.53	0.29	-0.25	0.66
M	-23.59	-0.04	-0.07	23.59	-0.35	0.04	0.18	0.54

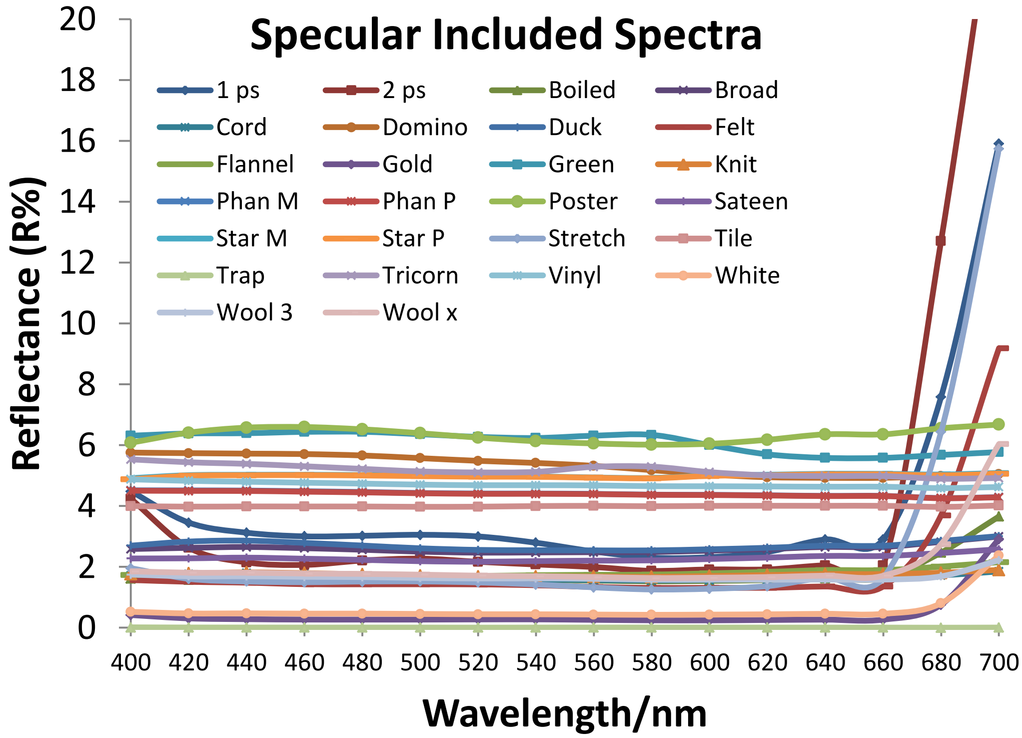

Figure 6.4. Specular Included Spectra from the Second Experiment.

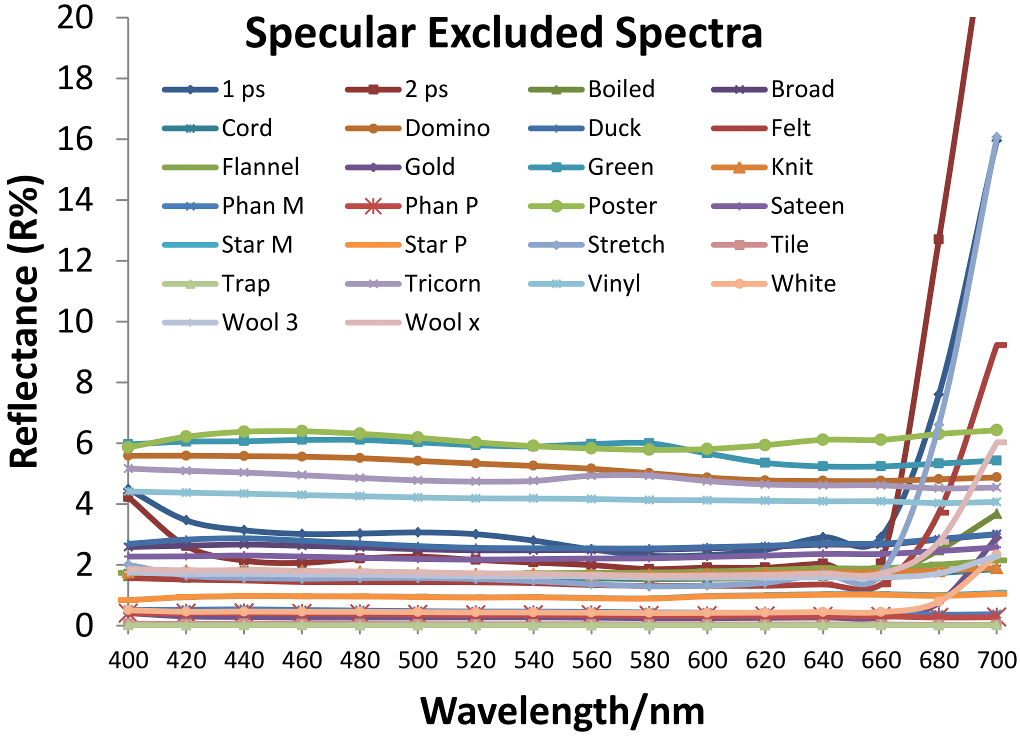

Figure 6.5. Specular Excluded Spectra from the Second Experiment.

The four samples used in the third trial were purchased from retail stores. Their average colorimetric values are given in Table 6.5. The samples were measured using the specular included mode only.

Table 6.5. Colorimetric values for the third experiment.

Sample	L*	a*	b*
1	17.66	0.02	-3.65
2	18.53	0.56	-5.54
3	19.09	0.47	-4.11
4	18.34	0.16	-1.26

3. Results and Discussion

ASTM 691 was used for data analysis.[3] It provides a method for analyzing repeated test measurements to determine repeatability and reproducibility of the measurement system.

Although its methods are typically applied to the results of an inter-laboratory study, conducting this analysis on measurements from a single laboratory will give an indication of the lowest possible variability, as operator and environmental conditions will be held constant. The study then tests the repeatability and reproducibility of the instruments themselves. The repeatability is defined as how well a single instrument produces the same output data for the same sample on different days, while the reproducibility is defined as how well different instruments give the same output data for the same sample.

The analysis provides two measures of repeatability, r (repeatability) and k (the within-laboratory consistency statistic), as well as two measures of reproducibility, R (reproducibility) and h (the between-laboratory consistency statistic). The h and k values are used to ensure that all instruments used were properly calibrated and that no instrument reported consistently different values (i.e. bias) from the remaining instruments in the group. In these experiments, no instrument was consistently flagged as being different from the other instruments, and because of this and the prohibitively large amount of h and k data calculated, these values are not reported in this paper. The r and R values were calculated at 95% confidence intervals, and represent the range expected in a measurement repeated on different days; that is, two measurements repeated on different days will be within these limits 95% of the time.

3.1 Colorimetric Data

From the first experiment, the ASTM 691[3] analysis indicates that there is more variability in colorimetric values obtained from handheld spectrophotometers than those obtained from desktop models. Moreover, handheld spectrophotometers exhibit greater variation from the grand mean colorimetric values. From the second experiment, ASTM 691 analysis indicates that the tested spectrophotometers produced consistent results. All spectrophotometers used in this trial were desktop models, and all were properly calibrated according to the procedures recommended by the American Association of Textile Chemists and Colorists.[4]

To determine the final confidence intervals, repeatability (within one instrument, designated *r*) and reproducibility (between instruments, designated *R*) were determined for each of the first two data sets. These values were then combined in a weighted average, based on the number of observations used, to calculate the intra and inter-instrument agreement for each data set. For the first experiment 210 assessments and for the second experiment 208 assessments were employed. The average repeatability and reproducibility limits are given in Table 6.6.

Table 6.6, *r* and *R* limits, weighted averages.

	Specular Included		**Specular Excluded**	
	r	***R***	***r***	***R***
L*	±0.55	±0.93	±1.15	±1.71
a*	±0.37	±0.43	±0.17	±0.25
b*	±0.20	±0.63	±0.27	±0.68
ΔL^*_{ab}	±0.63	±0.69	±0.20	±0.27
Δa^*	±0.11	±0.12	±0.15	±0.17
Δb^*	±0.20	±0.28	±0.29	±0.31
ΔE^*_{ab}	±0.62	±0.65	±0.86	±0.92

The calculated values are higher than previously published confidence intervals for colorimetric values that represented various regions of the color space.[2] The previously published confidence intervals are given in Table 6.7, and unfortunately only include reproducibility values for specular included measurements.

Table 6.7. Previously published reproducibility values for materials of a variety of colors.

	Reproducibility
L*	±0.16
a*	±0.15
b*	±0.22
ΔL^*_{ab}	±0.02
Δa^*	±0.06
Δb^*	±0.07

As spectrophotometers measure light reflected from an object’s surface, and dark objects reflect much less light, there is a much smaller signal-to-noise (S/N) ratio for these measurements. As the S/N ratio approaches unity, the precision of an instrument decreases, and when the S/N ratio is between zero and one, the measurement is not reliable at all. Thus there is more inherent variability when measuring very dark objects compared to those that reflect. As such the methodology employed for the calibration and measurement of dark objects requires a higher degree of scrutiny.

The data from the third experiment was used to verify the confidence intervals set for measurements based on specular included conditions. All measurements were within the confidence intervals determined, indicating that the calculated confidence intervals are wide enough to be robust. The reproducibility values (intra-instrument variability) for each sample are given in Figure 6.3. Figure 6.4 shows the repeatability of assessments (inter-instrument agreement) among the three spectrophotometers tested. In Figure 6.4, Sample 1-2 refers to Sample 1 measured on Spectrophotometer 2. The confidence intervals presented here are meant to be interpreted as plus/minus confidence intervals; that is, R=±0.93 for L* indicates that measurements are subject to variability with a total range of 1.86 centered on the mean measurement. Thus a sample with a measured L* value of 10 on one spectrophotometer will, with 95% confidence, give an L* value between 9.07 and 10.93 when measured on a second spectrophotometer.

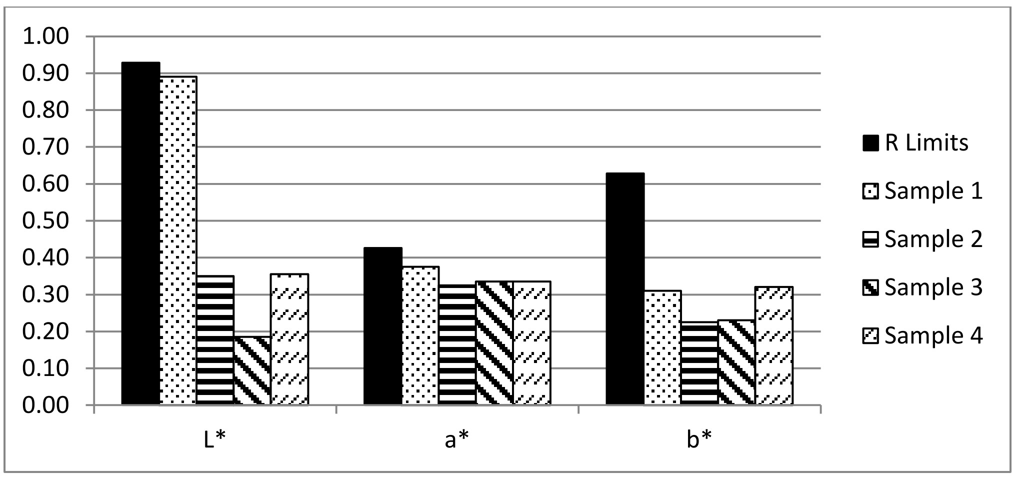

Figure 6.6. R limits and reproducibility of data from third experiment.

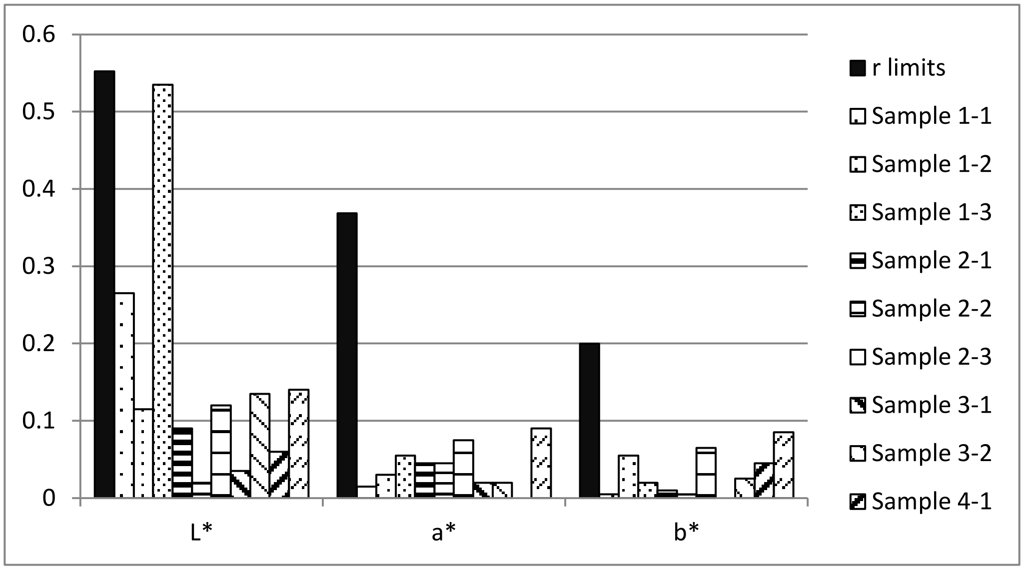

Figure 6.7. r limits and repeatability of data from third experiment.

3.2 Spectral Data

Although in the second experiment, the spectra were collected every 10 nm from 350 to 750 nm and all data points were used in the calculation of colorimetric values, only the readings every 20 nm from 400 to 700 nm were analyzed in the spectral analysis. ASTM 691 analysis was applied to the spectral data as well. Reflectance values varied from 0.00 to 25.37%, with very few readings above 6%. Repeatability and reproducibility limits were calculated for every 20 nm wavelength between 400 and 700 nm for all samples and spectrophotometers used in experiment two. The averaged limits are given in Table 6.8.

Table 6.8. Repeatability and Reproducibility limits by wavelength.

	Specular Included		Specular Excluded	
Wavelength	**r**	**R**	**r**	**R**
400	0.136	0.335	0.164	0.330
420	0.121	0.257	0.144	0.247
440	0.112	0.214	0.138	0.231
460	0.109	0.197	0.132	0.211
480	0.108	0.185	0.131	0.206
500	0.105	0.164	0.129	0.191
520	0.103	0.159	0.125	0.186
540	0.100	0.173	0.122	0.189
560	0.098	0.172	0.120	0.189
580	0.097	0.158	0.117	0.183
600	0.097	0.166	0.119	0.188
620	0.096	0.173	0.118	0.190
640	0.101	0.186	0.124	0.204
660	0.101	0.186	0.124	0.204
680	0.212	0.326	0.242	0.373
700	0.363	0.527	0.443	0.627

The r and R limits increase at the ends of the spectrum. This is likely due to the fact that these wavelengths of light are bent further inside the instrument due to the diffraction grating, and thus are more susceptible to noise. Furthermore, the wavelengths at the high end of the spectrum are where the reflectance values between 6% and 26% are found. These higher reflectance values have wider repeatability and reproducibility confidence intervals, but the uncertainty in the measurements is a smaller fraction of the total response of the spectrophotometer. When only data points with a reflectance below 6% are considered, the r and R values may be averaged to be 0.11 and 0.189 respectively for specular included conditions and 0.138 and 0.213 respectively for specular excluded conditions, across all wavelengths measured.

Conclusions

Due to the small signal to noise ratio in the measurement of dark objects their measurement produces significantly larger variability measurement of lighter objects. The variability, both between and within instruments, for the handheld spectrophotometers studied was higher than in desktop spectrophotometers studied. Variations in spectrophotometers may result in as much as two color difference units, determined by ΔE^*_{ab}, for the same sample.

Acknowledgment:

We are grateful to Roland Connelly for the provision of technical data for Experiment 3 and for helpful discussions.

References

1. Skoog DA, Holler FJ, Nieman TA. Principles of Instrumental Analysis. Philadelphia: Harcourt Brace and Company; 1998. 849 p.

2. Clonts R, Thangavelu R, Hinks D, Dunn J, Guzman P, Laidlaw A. Inter-Instrument Agreement in the Colorimetric Measurement of Textile Materials. AATCC Review 2006; 8:45-48.

3. ASTM 691. Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method. ASTM International; 2008.

4. AATCC Evaluation Procedure 6, 2010. Instrumental Color Measurement. American Association of Textile Chemists and Colorists.

7. Conclusions

From the observation trials, it can be concluded that bluish and greenish blacks are considered perceptually blacker than reddish and violetish blacks. Analysis of the data indicates that, irrespective of chroma, on average samples with hue angles between approximately 200° and 270° were perceived to be the most black, i.e., cyanish-blacks to bluish-blacks. Blacks with hue angles above 315° or below 45° (reddish-blacks) were considered to be the least black and the ratings trended between the most and least blacks as a function of hue angle. From analysis of confidence intervals, it can be seen that there is no difference among reddish samples, and no difference among blue-to-green samples, but that the blue-to-green samples are perceived as being significantly blacker than the reddish samples. In general, the blackness rating was inversely proportional to C* for the samples that were not greenish to bluish black. Hence, for the observers studied, increasing C* has a deleterious effect on perception of blackness for all samples except greenish blacks and bluish blacks.

It may also be said that using a rating metric, instead of a ranking one, provides better data for the formulation of a blackness index. Ratings values may be translated between data sets, while ranks are only valid within the data set in which they were collected. The use of an ideal black as an endpoint for the ratings increases the robustness of these data sets.

Due to the small signal to noise ratio in the measurement of dark objects their measurement produces significantly larger variability measurement of lighter objects. The variability, both between and within instruments, for the handheld spectrophotometers studied was higher than in desktop spectrophotometers studied. Variations in spectrophotometers may result in as much as two color difference units, determined by ΔE^{*}_{ab}, for the same sample. Specular excluded measurements are subject to more variation than specular included measurements.

8. Future Work

The immediate next step is to determine the contribution of lightness to the perception of blackness. This should be done by obtaining (through printing or dyeing) uniformly textured samples that vary in all three dimensions of color space. The samples should be focused in the blue-green region of color space, as that is the area that has been identified as being perceptually blacker than other regions. If possible, samples should range in lightness from 6 to 25, although samples between 10 and 12 are not necessary as these have already been tested. They should be rated in terms of blackness by comparison to an ideal black (the "black box" used in the second experiment described in this work) by a minimum of 26 color normal observers (13 males and 13 females), using the scale from the second experiment, where 0 is an ideal black and 10 is the farthest from ideal that would still be called black. The ratings assigned in this trial should be combined with the ratings assigned in the second experiment of this work, and an index should be postulated. The index's form should be determined by the data, finding the best fit of the data in L*C*h°, L*a*b*, or Yxy space.

Following the proposal of the index, the index should be tested with "real world" samples obtained from various sources. The samples should all be fairly uniform in texture, although some variation will have to be accepted. Many appropriate samples have already been generated or collected, and thus the sample preparation step for this experiment should be trivial. The samples should be rated, again by a minimum of 26 color normal observers, and their ratings should be checked against the proposed index. If the index is found to be robust, it should be published. If the model requires extensive modification, those changes should be made and the model should be tested again with new "real world" samples.

Additionally, it is reasonable to assume that other groups will be publishing their own blackness indices in the near future. Work verifying those indices, possibly with the goal of combining them, will be required. This work should be collaborative in nature, and should determine a final blackness index to be proposed to the CIE. No observer tests should be required, as data from the formulation of one index can be used to verify a second index. However, if it is found that the methodologies for obtaining each index vary significantly, the models should be tested with at least 26 color normal observers. The major expected difference between models is the use of different endpoints of the scale (different definitions of ideal black), and rating data could be converted mathematically to match the scale in use. For example, the current work dictates that the lower a sample is rated, the blacker it is perceived. Earlier German indices dictated the opposite relationship; that is, the higher a sample was rated, the blacker it was. By determining the endpoint of the German scale and reversing the direction of data from the current work, these two data sets may be compared or even combined.

Following the publication of a blackness index, further work could be done to ascertain the contribution of texture to the perception of blackness. Yarns should be dyed black and used to knit or weave samples of varying textures. This will ensure that all samples are colorimetrically uniform, avoiding the problems of uneven dye uptake due to fabric structure features. These samples should then be rated against the ideal black, and a texture term should be added to the index if it is found to be appropriate.

Furthermore, results from these experiments (showing that bluish blacks are preferred) could be combined with results from whiteness experiments (showing that bluish whites are preferred) to possibly reshape color space. If blue undertones are universally preferred along the entire achromatic spectrum, then perhaps the L*a*b* color space should reflect this in the position of the achromatic axis. Work on determining preferred grey samples would be an appropriate first step here. Grey samples of varying lightness, chroma and hue should be prepared, and observers should be asked to rate the neutrality of the samples. No ideal grey should be presented, as no single ideal grey exists (there is no clear endpoint to the grey scale, as there is with whiteness and blackness).

There is also room for improvement in the presentation of digital samples to observers. Using a radiospectrophotometer, color uniformity across a monitor should be characterized. Once the color reproduction capabilities of the monitor are determined, it would be helpful to build a lookup table of RGB inputs, position on the monitor, and L*a*b* outputs. With this data in hand, it might be possible to use digital samples instead of physical ones for the formulation and verification of the blackness index.

REFERENCES

AATCC Evaluation Procedure 6, 2010. Instrumental Color Measurement. American Association of Textile Chemists and Colorists.

ASTM 691, 2008. Standard Practice for Conduction an Interlaboratory Study to Determine the Precision of a Test Method. ASTM International.

Berger, A., 1959. Weissgradformeln und ihre praktische Bedeutung. Farbe 8, 187-202.

Bimler, D. L.; Paramei, G. V.; Izmailov, C. A., 2009. Hue and saturation shifts from spatially induced blackness. J. Opt. Soc. Am. A 26, 163-172.

Bode, R.; Ferch, H.; Koth, D.; Schumacher, W.. 1979. Schwarzgradskala f ür Pigmentruße. Farbe + Lack 85, 7-13.

Caivano, J. L., 1991. Cesia: A System of Visual Signs Complementing Color. Color Res. Appl. 16, 258-268.

Caivano, J. L., 1993. Appearance (Cesia): Variables, Scales, Solid. Farbe 39, 115-125.

Caivano, J. L., 1994. Appearance (cesia): construction of scales by means of spinning disks. Color Res. Appl. 19, 351-362.

Caivano, J. L., 1996. Cesia: Its Relation to Color in Terms of the Trichromatic Theory. Farbe 42, 51-63.

Cicerone, C. M.; Volbrecht, V. J.; Donnelly, S. K.; Werner, J. S., 1986. Perception of Blackness. J. Opt. Soc. Am. A 3, 432-437.

Clonts R.; Thangavelu R.; Hinks D.; Dunn J.; Guzman P.; Laidlaw A., 2006. Inter-Instrument Agreement in the Colorimetric Measurement of Textile Materials. AATCC Review 8, 45-48.

Clonts R.; Shamey, R., 2008. "Perceptual Assessment of Blackness", ISCC Special Topics Meeting: Black & White, Inter-Society Color Council, 20-22 Nov, 2008, Portland, OR, USA.

Clonts, J. R.; Shamey, R.; Hinks, D., 2010. Effect of Colorimetric Attributes on Perceived Blackness of Materials. CGIV 2010 -- Fourth European Conference on Colour in Graphics, Imaging, and MCS/10 Vision 12th International Symposium on Multispectral Colour Science, 83-87.

Croes, A. W., 1959. Simple Formula for Calculating Whiteness from Photoelectric Tristimulus Data. J. Opt. Soc. Am. 49, 830-831.

Dacey, D. M.; Lee, B.B., 1994. The 'blue-on' opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. Nature 367, 731-735.

Dartnall, H. J. A.; Bowmaker, J. K.; Mollon, J. D., 1983. Human Visual Pigments: Microspectrophotometric Results from the Eyes of Seven Persons. Proceedings of the Royal Society of London. Series B, Biological Sciences 220, 121.

Dimmick, F. L., 1920. A Note on the Series of Blacks, Greys, and Whites. Amer. J. Psychol. 31, 301-302.

Dimmick, F. L., 1929. Gray and the Color Pyramid. Amer. J. Psychol. 41, 284-290.

Dimmick, F. L., 1933. The Psychophysical Determination of the Limits of Pure Gray. Amer. J. Psychol. 45, 313-314.

Dimmick, F. L., 1941. Black and White. Amer. J. Psychol. 54, 286-289.

Evans, R., 1974. The Perception of Color. Wiley: New York.

Friele, L. F. C., 1959. Measurement and Grading of Whiteness of Optical Bleached Materials. Farbe 8, 171-186.

Fuld, K.; Otto, T. A.; Slade, C. W., 1986. Spectral responsivity of the white-black channel. J. Opt. Soc. Am. A 3, 1182-1188.

Ganz, E., 1976. Whiteness: photometric specification and colorimetric evaluation. Applied Optics 15, 2039-2058.

Gläser, F., 1978. Probleme bei der Farbmessung geschönter Schwarzdrucke. Farbe + Lack 84, 579-584.

Griesser, R., 1973. Instrumental assessment of whiteness. Ciba-Geigy Review 1.

Griesser, R., 1981. Instrumental Measurement of Fluorescence and Determination of Whiteness: Review and Advances. Review of Progress in Coloration and Related Topics 11, 25-36.

Hardy, A. C., 1930. US Patent 1 780 231.

Heggelund, P., 1974. Achromatic color vision I: Perceptive variables of achromatic colors. Vision Research 14, 1-1079.

Heggelund, P., 1974. Achromatic color vision II: Measurement of simultaneous contrast within a bidimensional system. Vision Research 14, 1081- 1088.

Helmholtz, H. v., 1962. Handbook of physiological optics. J. P. C. Southall, (Ed. and trans.), Dover: New York.

Hering, E., 1964. Outlines of a Theory of the Light Sense; Harvard University Press: Cambridge, MA, 1964.

Hesselgren, S., 1967. The Language of Architecture. Studentlitteratur: Lund, Sweden.

Hunter, R. S., 1958. Description and Measurement of White Surfaces. J. Opt. Soc. Am. 48, 597-605.

Hurvich, L. M. and Jameson, D., 1957. An opponent-process theory of color vision. Psychological Review 64, 384-404.

International Commission on Illumination, 1986. CIE Technical Report on Colorimetry 15.2, 36-38.

Judd, D. B., 1935. Method for Determining Whiteness of Paper. Paper Trade Journal 100, 40-42.

Judd, D. B., 1936. Method for Determining Whiteness of Paper II. Paper Trade Journal 103, 38-44.

Judd, D. B., 1940. Hue, saturation, and lightness of surface colors with chromatic illumination, J. Opt. Soc. Am. 30, 2-32.

Judd, D. B., 1941. The Definition of Black and White. J. Psychol. 54, 289-294.

Judd, D. B., 1961. A Five-Attribute System of Describing Visual Appearance. ASTM Special Technical Publication 297.

Kenya Bureau of Standards, 2005. Black lead pencils - specification. KS 90:2005.

Kipphan, H., 2001. Handbook of Print Media: Technologies and Production Methods; Springer: Berlin.

Kuehni, R.G., Color Space and Its Divisions, Color Order from Antiquity to the Present, 2003, Wiley and Sons, New Jersey.

Kuehni RG, “Variability in estimation of suprathreshold small color differences,” Color Res. Appl. 34, 367–374 (2009).

Kulp, T. D.; Fuld, K., 1989. Black spectral responsivity. J. Opt. Soc. Am. A 6, 1233-1238.

Ladd-Franklin, C., 1932. Colour and Colour Theories; Harcourt, Brace and Company: New York.

Lippok-Lohmer, K., 1986. Praxisnahe Schwarzmessungen. Farbe + Lack 92, 1024-1029.

MacAdam, D. L., 1934. Specification of Whiteness. J. Opt. Soc. Am. 24. 188-191.

Makarenko, A. V.; Shaykevich, I. A., 2000. Dependence of the Whiteness of Paper on Surface Roughness and Illumination Conditions. Color Res. Appl. 45, 170-175.

Michaels, G. M., 1925. Black: A Non-light Sensation. Psychol. Rev. 32, 248-250.

Nakajima, Y.; Yukawa, Y.; Kamimori, I.; Yamanouchi, A.; Hoshida, Y., 2007. European Patent 1 270 624 B1.

Neifeld, M. R., 1924. The Ladd-Franklin Theory of the Black Sensation. Psychol. Rev. 31, 498-502.

Neitz J., 2001. Manual: Neitz Test of Color Vision, Western Psychological Services.

Patton, T. C., 1973. Pigment Handbook; Wiley: New York.

Pope, A., 1949. The Language of Drawing and Painting. Harvard University Press: Cambridge.

Quinn, P. C.; Wooten, B. R.; Ludman, E. J., 1985. Achromatic Color Categories. Perception and Psychophysics 37, 198-204.

Rich, G. J., 1926. Black and Grey in Visual Theory. Amer. J. Psychol. 37, 123-128.

Schumacher, W., 1976. In Photometrische Schwarzmessung. pp 582-587.

Selling, H. J.; Friele, L. F. C., 1950. Whiteness relations and their applications. Applied Scientific Research B1, 453-476.

Shamey R, L. M. Cárdenas, D. Hinks, and R. Woodard, “Comparison of naive and expert subjects in the assessment of small color differences,” J. Opt. Soc. Am. A 27, 1482–1489.

Shinomori, K.; Nakano, Y.; Uchikawa, K., 1994. Influence of the illuminance and spectral composition of surround fields on spatially induced blackness. J. Opt. Soc. Am. A 11, 2383-2388.

Shinomori, K.; Schefrin, B. E.; Werner, J. S., 1997. Spectral mechanisms of sptially induced blackness: data and quantitative model. J. Opt. Soc. Am. A 14, 372-387.

Skoog D.A.; Holler F.J.; Nieman T.A., 1998. Principles of Instrumental Analysis. Harcourt Brace and Company: Philadelphia.

Society of Dyers and Colourists, 1987. Colour Index International; The Society: Bradford, West Yorkshire.

Stockman, A.; Plummer, D. J.; Montag, E. D., 2005. Spectrally opponent inputs to the human luminance pathway: slow +M and −L cone inputs revealed by intense long-wavelength adaptation. The Journal of Physiology 566, 61-76.

Titchener, E. B., 1916. A Note on the Sensory Character of Black. Journal of Philsophy: Psychology and Scientific Methods 13, 113-121.

Titchener, E. B., 1916. A Further Word on Black. Journal of Philsophy: Psychology and Scientific Methods 13, 649-655.

Uchida, H., 1998. A New Whiteness Formula. Color Res. Appl. 23, 202-209.

Vaeck, S. V., 1966. On the Topology of Whiteness in Colour Space. Annales textiles 2, 184-200.

Volbrecht, V. J.; Werner, J. S.; Wooten, B. R., 1989. Temporal Induction of Blackness--I. Color Appearance. Vision Research 29, 1425-1436.

Volbrecht, V. J.; Werner, J. S.; Wooten, B. R., 1989. Temporal Induction of Blackness--II. Spectral Efficiency and Tests of Additivity. Vision Res. 29, 1437-1455.

Volbrecht, V. J.; Werner, J. S.; Cicerone, C. M., 1990. Additivity of spatially induced blackness. J. Opt. Soc. Am. A 7, 106-112.

Volbrecht, V. J.; Kliegl, R., 1998. Color Vision: Perspectives from Different Disciplines, in: Backhaus, W. G. K., Kliegl, R. and Werner, J. S., (Eds.), The Perception of Blackness: An Historical and Contemporary Review. Walter de Gruyter & Co.: New York, pp 187-206.

Ward, J., 1905. Is 'Black' a Sensation? Brit. J. Psychol. 1, 407-427.

Ward, J., 1916. A Further Note on the Sensory Character of Black. Brit. J. Psychol. 8, 212-221.

Westland, S.; Cheung, T. L. V.; Lozman, O. R., 2006. A metric for predicting perceptual blackness. 14th Color Imaging Conference Final Program and Proceedings, 14-17.

Werner, J. S.; Cicerone, C. M.; Kliegl, R.; DellaRosa, D., 1984. Spectral efficiency of blackness induction. J. Opt. Soc. Am. A 1, 981-986.

Wright, W. D., 1980. Perception of Blackness. Farbe 28, 161-166.

Wu, S. M., 2010. Synaptic Organization of the Vertebrate Retina: General Principles and Species-Specific Variations: The Friedenwald Lecture. Invest. Ophthalmol. Vis. Sci. 51 1264-1274.

Xin, J. H.; Shen, H.-L.; Lam, C. C., 2005. Investigation of Texture Effect on Visual Colour Difference Evaluation. Color Res. Appl. 30, 341-347.

Zollinger, H., 2003. Color Chemistry: Syntheses, Properties, and Applications of Organic Dyes and Pigments; Wiley-VCH: Switzerland.

APPENDICES

Appendix A. Psychophysical Assessment Instructions

1. Today's testing will have two stages. In the first part, I will show you thirty samples and ask you to tell me if they are black or not. A yes or no will suffice for the answer. You'll see each sample by itself in a random order. It's okay if you think they're all black, or that none of them are black. Color vision is a very subjective thing, and so there are no right or wrong answers.

2. (As the ideal black box was presented) Now I'll show you the same thirty samples again. They're in a different random order this time, and your previous answers don't matter. The color that you see in the hole in the lid of this box is an ideal or perfect black. You're going to rate the samples on a scale using this color as a reference point. Consider this color to be zero. Any color that you would call black should be given a rating between 0 and 10, inclusive. If it matches the color you see in the hole exactly, it's a zero. If it's close but not a match, rate it maybe 2 or 3. If it's barely black, give it a 10, and if it's almost black but not quite, give it an 11. Do you understand the instructions?

Appendix B: Whiteness

Many similarities exist between the models for blackness and whiteness. In both cases, there is a theoretical “perfect” color: complete absorption or reflectance of light. People encounter myriad objects in their daily lives that they call white or black, and a cursory examination shows that these objects may have a rather large range of hues and shades. There are essentially no “perfect” whites or blacks in most people’s experiences. In the early twentieth century, color scientists devoted much effort to quantifying whiteness, most with encouragement from the paper industry. Because of the connection between white and black, and the lack of research on black objects, a good starting point for an investigation into a blackness index is this early work on whiteness indices.

A.1 MacAdam’s 1934 Work

One of the earliest papers on whiteness came from MacAdam in 1934 (MacAdam, 1934). He recognized that color graders were making judgments based on brightness or reflectance and purity. Since many white objects are slightly yellowish, he focused on cotton fabric samples with dominant wavelengths between 574 and 576 nm. The samples were laundered, in some cases with iron salts or carbon black, to vary their brightness and purity. The salts and carbon did not change the dominant wavelengths of the samples. Thirty observers were asked to force rank 24 samples, ranging from 70% to 90% in brightness and 0% to 10% in purity, in order of whiteness. He found little consistency between observers or between trials with the same observers. However, there was a general pattern to all of the rankings. It was evident that contours of equal whiteness could be drawn on a plot of brightness vs. purity for the samples. At low purities, the brightness was the most significant factor in whiteness, but at higher purities brightness was less important. MacAdam stated that “any one-parameter whiteness classification must be restricted to the class of samples for which it is devised,” (MacAdam, 1934) indicating the need for either a larger class of samples or multiple indices.

A.2 First Judd Whiteness Equation

The next year, Deane Judd began working on whiteness, using paper samples (Judd, 1935). Judd again recognized that near-white objects may vary from ideal white in either brightness or chromaticness. He also acknowledged that whiteness might be measured as the departure from an ideal white standard, defined as a layer of magnesium oxide at least a half millimeter thick. Furthermore, he identified the central theoretical question to be answered: “How much achromatic departure (grayness) is equivalent to a given amount of chromatic departure (say yellowness)?” Observers asked to rank samples in order of whiteness must answer this question for themselves, and it is this answer that leads to much of the variability between observers. MacAdam had restricted his samples such that they varied only in greyness and yellowness, but Judd wondered about deviations in other chromatic directions. To measure this, he used the distance on a uniform chromaticity scale between the standard white point and the point representing the sample in question. Between MacAdam’s publication and Judd’s, MacAdam had sent Judd a formula for whiteness:

(A.1) $$Whiteness = \sqrt{(R - KS^2)},$$

where R = reflectivity, K = 6700, and S = distance. Judd replaced MacAdam's purity with his distance measurement, and found agreement between his modified formula and MacAdam's results. Judd measured thirty samples to determine their achromatic and chromatic departures from the MgO standard, and calculated whiteness values for them.

In the second part of Judd's paper, he had fifteen observers judge the original samples. To test the formula, three working standards were used: MgO, "natural paper white", and "intermediate", found experimentally (Judd, 1936). This caused the value of *S* in the formula to shift. Judd also varied the value of *K*, which varied the impact of the chromatic departure. He found that the formula did not hold true for samples that were clearly not white, giving output values below 0. It was determined that the observers could be split into groups depending on how well their rankings agreed with the calculated whiteness values obtained with each of the three working standards. Several observers seemed to be using "natural paper white" as their standard, while others used MgO. It should be noted that the "natural paper white" was calculated by finding the weighted average chromaticity coordinates of five observers' whitest three samples. It was assumed that this standard approached the whitest paper possible without the addition of a blue or red tint to the pulp. The "intermediate" standard was calculated based on four observers in a similar manner. It was also found that changing the value of *K* from 6700 to 5400 or 8000 did not impact the results significantly, although setting *K* to zero or infinity decreased the agreement with visual results. Regardless, Judd recommended the use of his revision of MacAdam's equation with MgO as the standard and *K* set to 6700.

A.3 Selling and Friele's Review

A preliminary review of the work on whiteness was published in 1950 by Selling and Friele (Selling and Friele, 1950). They looked at studies on unbleached and blued textiles and papers, ignoring optically brightened materials with the assumption that the results would be unchanged while the experimental difficulties would be increased. The authors recognized that the MgO standard may not be ideal and that a sample which was slightly bluer or yellower than MgO may be perceived as whiter. Since no such material exists, and because MgO was so much whiter than commercially available standards, they stated that any slight difference between MgO and the true ideal standard could be ignored. Selling and Friele first asked if the whiteness rankings from multiple observers were comparable. They conducted a test with 20 paper samples and 34 naïve observers, which showed that the different observers gave responses that were positively correlated with each other. They determined that 34 observers were too few to make absolute grading decisions, but that the group was sufficiently large to be reliable for their comparative study of whiteness indices. The one sample that gave the largest variability in rankings was somewhat greenish, and they concluded that it was more difficult to rank a sample which was not yellowish or bluish.

The first method they tested for use as a unidimensional whiteness scale was the brightness, measured as the percent reflectance at 450 nm. This method gave a correlation coefficient of 0.86 with visual gradings, but papers treated with a blue dye were problematic. The blue dye lowered the reflectance at 450 nm, and thus these papers were ranked lower according to the brightness method than visual grading. Their second method was lightness, using Y values. This gave a correlation coefficient of 0.73, performing worse than brightness. This method also failed for the blued papers, as dyeing will always decrease the lightness of a sample. Selling and Friele then investigated the proposed whiteness relations of Stephansen and Harrison. These relations, respectively, were:

(A.2) $$W = 2B - R$$

and

(A.3) $$W = 100 - R + B,$$

where B is the percent reflectance at 450 nm and R is the percent reflectance at 670 nm. When compared to visual assessments, Stephansen's formula had a correlation coefficient of 0.91, while Harrison's had a correlation coefficient of 0.81. Stephansen's formula also failed for blued papers, grading them whiter than visual observations did. Judd's whiteness formula (given in a previous section) was tested, and found to have a correlation coefficient of 0.85 with visual gradings. The primary advantage to Judd's formula, however, was its ability to determine an optimal dye level for a blued paper without failing. For their fifth method, two of Selling's own formulae were tested. These were

(A.4) $$W = 100 - \sqrt{\left(10\Delta\sqrt{Y}\right)^2 + k(\Delta s)^2}$$

and its simplification,

(A.5) $$W = 100 - \sqrt{(\Delta Y)^2 + k'(\Delta s)^2}.$$

The first equation used a root lightness scale and both equations used the saturation of the sample measured on the MacAdam transformation of the UCS diagram. Values of k and k' were determined using a series of small disks of a blend of lamp black and titanium dioxide along with ten paper samples. These formulae were found to have correlation coefficients of

0.95 and 0.94, respectively, and were theoretically capable of finding an optimum whiteness value for blued papers. The final method tested a second relationship proposed by Judd:

(A.6) $$W = 100 - \sqrt{\left(10\Delta\sqrt{Y}\right)^2 + \tfrac{49e7}{c^2}\sqrt{Y}(\alpha^2 + \beta^2)},$$

where c represented the color difference between black and white (taken to be 20 in Judd's publication), $\sqrt{\alpha^2 + \beta^2}$ represented the chromaticity difference between the sample and MgO on Hunter's uniform chromaticity scale, and a root lightness percent scale was utilized. The resemblance between Judd's work and Selling's was immediately recognized. Judd's formula gave a correlation coefficient of 0.85 with visual data, and Selling and Friele found that changing the value of c from 20 to 70 gave the formula a correlation coefficient of 0.95 with the visual data. After this preliminary testing, the researchers prepared a series of blued papers. The brightness, lightness, Stephansen and Harrison methods all immediately failed to agree with visual observations. Observers were then asked to choose the whitest of these blued papers, and these results were compared to the three Judd and two Selling formulae. It was finally concluded that only the modified second Judd and second Selling formulae were consistent with visual data. Selling and Friele also proposed a method for determining the optimum concentration of bluing agent. By plotting $10\sqrt{Y}$ vs. $\sqrt{k\Delta s}$ and finding the tangential circle, the optimum concentration could be found by interpolating between the two nearest points. The relative effectiveness of two blue dyes could be found on the same plot.

A.4 Hunter's 1958 CIE Review

In 1953, The American Ceramic Society began to study white surfaces to create a standard white for appliances to ensure uniformly colored kitchens in American households, but they were not interested in the input of the Inter-Society Color Council, and their project was eventually abandoned. The committee formed by the ISCC for this purpose continued its work on whiteness, however, and in 1958 Hunter published a summary of their efforts (Hunter, 1958). He first pointed out that preferred whites varied according to observers, and that for textile goods, most observers prefer bluish whites. He presented the reflectance curves of four white objects to illustrate how they differ from each other and show that many near whites have lower reflectances in the blue region of the spectrum than in the yellow region. Curves for each of these four objects were also obtained using a goniometer, and Hunter stated that the specular reflection could make these objects more difficult to reliable measure colorimetrically.

After a discussion of measurement instruments and color spaces that may be used to determine whiteness, he explained that green or blue reflectance was used in certain industries where the deviations from ideal whiteness are in the reddish or yellowish directions. Organic materials such as cellulose (a major component of both the paper and textile industries) tend to have relatively low blue reflectances. For this reason, several early whiteness quantification methods used lightness and yellowness as opposed to brightness and purity or chromaticness. With the assumption that the Lab color space was perceptually uniform, he suggested using the following equation to describe whiteness:

(A.7) $$W = 100 - \sqrt{(100 - L)^2 + (a^2 + b^2)}.$$

Hunter plotted the equiwhiteness lines provided by this equation on the same axes as MacAdam's earlier empirical equiwhiteness lines, and showed that for yellow samples the equation gave whiteness values higher than observers did.

A.5 Croes' Response to Hunter

Croes responded to Hunter's paper in 1959 (Croes, 1959). He simplified the formula that had been in use in the wheat industry, which Hunter felt was too complex for general use. The simplified version of the formula was given as

(A.8) $$W = G - A + B.$$

In addition to simplifying the formula, Croes also translated it from the tristimulus values *A*, *B*, and *G*, giving expressions for calculating whiteness using x, y and Y or u, v and V. He tested these three formulas on wheat milling products, and found them to be equivalent to visual gradings, but the data for this experiment was not presented. He finally cautioned that these equations were only to be used on samples which were yellowish.

A.6 Friele's 1959 Work

1959 also saw two papers on whiteness published in *Die Farbe*. The first of these was by Friele, and focused on optically brightened samples (Friele, 1959). He selected 23 paper samples (several of which contained a whitening agent) and asked 28 naïve observers to observe them pairwise. From these results, an average ranking from most to least white was obtained. He plotted equiwhiteness lines for his observers and for Selling's 1950 formula and noted that they did not match for samples containing a whitening agent. Realizing that observers were likely making determinations of whiteness based predominantly on either lightness or blueness, he separated out the observers using blueness as their criteria. Their equiwhiteness lines were also compared to the average equiwhiteness lines as well as the formula used by the Technical Association of the Paper and Pulp Industry, and again found no match. Friele suggested that a whiteness formula would have to be nonlinear to fit the data, but rejected Judd's earlier formula because of its symmetry around a neutral axis. He concluded that using brightness as a measure of whiteness was the best option for the time, although he did suggest a hyperbolic model based on tristimulus values. This model was given only as an example to spark discussion, as Friele stated that "the confusion with respect to the subject [was] already intolerable" (Friele, 1959).

A.7 Berger's Formula

Berger began his 1959 work with a review of, essentially, Selling and Friele's 1950 paper (Berger, 1959). He found two primary flaws with all of the formulae presented there: they failed when optically brightened materials were tested, and they required the measured X, Y, and Z values to be converted into α and β before the whiteness could be calculated. He suggested that a whiteness formula should take the form $W = f(\bar{Y}) + (\bar{Z} - \bar{X})$. By including a constant in the *f(Y)* term, the proper relationship between lightness and hue could be set. One advantage of this form was that it favored bluish whites. Berger tested the formula

(A.9) $$W = {}^{\bar{Y}}/_{3} + (\bar{Z} - \bar{X})$$

against the visual data of Selling and Friele, and found a correlation coefficient of 0.83 compared to their correlation coefficient of 0.85. Furthermore, he tested his formula against Selling's for a small series of optically brightened papers. His formula agreed with the visual data, while Selling's failed. He then prepared a new set of samples and tested his formula again. With the new samples, Berger obtained a correlation coefficient of 0.93. He found that some samples had larger observer variability than others, and that these samples tended to be reddish or greenish, or to be significantly lighter than other similarly yellow papers. He stated that his formula performed better than Selling's, although he did not give a correlation coefficient for Selling's formula and the new visual data. From these results, Berger concluded that it was impossible to define a normal observer for the assessment of whiteness. He pointed out that it would be unfeasible to find an equation that matched visual data perfectly, but that proposed whiteness formulae should be judged on fitting the visual data better or worse than other formulae. The proposed formulae should also be considered limited in scope, and should not necessarily be universal for all materials.

A.8 Griesser's 1973 Review

In 1973, Griesser gave an overview of the instrumental assessment of whiteness (Griesser, 1973). He highlighted the repeatability of instrumental measurements due to their consistent lighting. He advocated using reflectance measurements at one or two wavelengths for assessing the whiteness of samples before and after bleaching, as the lack of fluorescence and bluing in these cases simplifies matters. These methods unfortunately failed for bluish samples, however, and for samples with reflectance curves that did not follow the same basic shapes. He discussed the importance of consistent lighting for the measurement of fluorescently treated samples, and mentioned Ganz's calibration method as a possible way to deal with any inconsistencies. Griesser discussed several different types of proposed whiteness formulae, including those based on B, G and A values from tristimulus photometers, L, a and b values in the Hunter coordinate system, and functions of wavelength and/or lightness. He stated that many of these formulae were developed empirically and that they often fail when samples or observers change, and pointed to the work of the CIE towards identifying a "standard observer for whiteness". While this was perhaps a reasonable undertaking at the time (although Berger had already stated one would not exist (Berger, 1959)), it would not be many years before the CIE accepted that no such standard observer exists (International Commission on Illumination, 1986). Griesser then discussed Ganz's work analyzing various whiteness formulae, but Ganz's next step required the standard observer for whiteness at the time of writing. Griesser concluded by discussing measurement methods and techniques.

A.9 Vaeck's Topological Approach

Vaeck took a topological approach to whiteness, published in 1966 (Vaeck, 1966). He argued that mapping XYZ values to a single whiteness value was only valid if it could be reduced to a fairly simple linear vector combination, and that the available empirical whiteness formulae were wrong. He maintained that MacAdam's early work was flawed due to the lack of fluorescent samples and the use of illuminant C instead of D_{65}. MacAdam, Vaeck stated, relied on two assumptions which were incorrect: that a perfect diffuser was the ideal white, and that for a given luminance the neutral point was the best possible white. As late as 1955, MacAdam was arguing that blueness should be penalized in a good whiteness formula, although consumers were already of the opinion that a bluish white was better than a neutral one. In 1959, the Zeiss Elrepho came onto the market and allowed researchers to measure fluorescence, and this was the basis of Friele and Berger's work. This pair, however, only used commercial white paper and textile samples, and thus left a large portion of the "white" region unexplored.

Vaeck began his research in 1964, looking at fluorescent and nonfluorescent samples that varied in the red–green direction. Based on these results, he created a topology of whiteness that covered the entire white region. When tested against all available data, he considered it a great success. Vaeck claimed that most researchers in the area of whiteness accepted his view, although some felt that simpler equations would be as successful as his complex mapping algorithms. In the course of his mapping, he dropped the idea of an ideal white point altogether, as there seemed to be one point for natural whites and another for fluorescent whites. Because luminance is less important than chromaticity in the perception of whiteness, he felt this was plausible. Further work, however, showed that because natural whites and fluorescent whites were placed in the same topological framework, some near-whites that were treated with both fluorescent whitening agents and blue dye could not be properly analyzed. Thielert and Schliemann proposed a second formula to correct this problem, but they assumed that an optimal white point existed and that the equiwhiteness lines were the same shape at all luminance levels. Vaeck attributed the deficiencies of their method to the fact that they only used white papers, while whiter textile samples could be obtained. They also seem to have used D_{60} illumination for their measurements (which may have been done incorrectly anyway) while basing their formula off of Vaeck's earlier work, which relied on illuminant C. Vaeck then pointed out that his equiwhiteness lines explained Thielert and Schliemann's data better than their own formula did. He concluded from all of this that the equiwhiteness lines must change their shape according to the luminance of the samples. For relatively low luminance values, the optimum white point should be near neutral, and for high luminance values, the optimum white point is so far towards blue that it is completely theoretical. His final remark was that measurements and evaluations on a carefully produced set of white samples were needed to accurately map the white region.

A.10 Ganz's 1976 Review

Ganz began his 1976 paper with a thorough review of the difficulties in accurately measuring fluorescent white samples (Ganz, 1976). He went into detail about the differences between real light sources and the standard illuminant D_{65}, and gave several techniques to convert colorimetric values to the standard illuminant. He then discussed existing whiteness formulae, dividing them into two groups: formulae that set the color of a perfect diffuser as a standard white in a uniform color space, and formulae that were based on reflectance values under different filters. He cautioned readers that although whiteness formulae all define equiwhiteness regions, they are only to be used in the near-white region, and that their potential failings for chromatic colors should not be considered a flaw. He recognized that most whiteness formulae gave consistent results on the yellowish side of neutral, but that there was much disagreement on the bluish side of neutral. This is especially noticeable when fluorescent samples are evaluated, and is particularly true of formulae that treat the perfect diffuser as the ideal white. He went on to discuss the issues of preferred whites and hue preferences, stating that the formulae previously discussed do not have inherent hue preferences, although those based on reflectance values may be given hue preferences by the selection of values for their constant parameters.

Ganz then suggested that a more universal whiteness formula could be constructed that depended on the relevant preferred white that would take the shape of halves of hyperboloids or tilted hyperboloids in the CIE Y, x, y color space. He used a set of paper samples prepared by Berger to evaluate the effects of reddish or greenish tints on whiteness assessment. Observers rejected samples that were too chromatic to be white and then sorted the remaining samples into eleven tint groups—five red groups, five green groups, or blue. The results from this experiment indicated that a whiteness formula based on hyperboloids that are symmetric about a yellow-blue hue line would be acceptable. Ganz found that tint was independent of the UV content of the light source, and that a perfect diffuser has no tint. He concluded that the older linear whiteness formulae were all roughly equivalent, and that the choice amongst them should be based on convenience and available instrumentation. Furthermore, his hyperbolic formula and tint scale needed further experimentation.

A.11 Griesser's 1981 Review

In 1981, Griesser presented a review of whiteness formulae, focusing on fluorescent materials (Griesser, 1981). He reminded readers that no consensus had been reached as to which formula was the best, and that all of the formulae were biased in one direction or another. Formulae based on one or two reflectance measurements and those based on tristimulus values were reviewed. These formulae tended to give results inconsistent with visual observation when fluorescent whitening agents were present, and many of them were only valid for the class of samples used in their developments. Furthermore, several of the tristimulus value-based formulae have strong hue preferences—Berger's and Croes's showed a marked blue-green preference, while Stensby's was useful for assessing samples that had violet tints. Ganz proposed a formula that showed a more neutral hue preference, which was intended to be more universally applicable. Vaeck proposed his graphical system utilizing equiwhiteness lines in a chromaticity diagram, but it was difficult to reconcile with the light source of the measuring instrument, and it did not match well with visual observations. Thielert and Schliemann had proposed a similar graphical method using equiwhiteness ellipses in the chromaticity diagram, but this suffered the same problems as the Vaeck system. Griesser pointed out that calculations of the whiteness of samples treated with fluorescent whitening agents based on fixed parameters (such as k=6700 in Judd's first formula) were dependent on the instruments used for measurement, as each lamp contained slightly different UV contents. The fixed parameters also had the effect of setting a hue preference for the formulae, and this might be undesirable in some situations. To deal with these problems, Ganz created a whiteness formula based on x, y, and Y that gave equiwhiteness planes with parameters that could be changed in order to tune the bias of the formula. This equation, also known as the Ciba-Geigy method, is

(A.10) $$W = (D \cdot Y) + (P \cdot x) + (Q \cdot y) + C,$$

where

(A.11) $$D = \left(\frac{\delta W}{\delta Y}\right),\ P = -\left(\frac{\delta W}{\delta S} \cdot \frac{cos(\varphi+\eta)}{cos(\varphi)}\right),\ Q = -\left(\frac{\delta W}{\delta S}\frac{sin(\varphi+\eta)}{cos(\varphi)}\right),$$

and

(A.12) $$C = [W_0 \cdot (1 - D)] - (P \cdot x_L) - (Q \cdot y_L).$$

This formula was highly dependent on the reference dominant wavelength (of the chosen reference white), as well as the hue preference for the observer. It did, however, show good agreement with visual observations for cotton and plastic samples. The values of the parameters for this formula could be found graphically or mathematically, utilizing a standard physical white scale. Griesser went on to discuss the effect of a tint deviation in the violet–red or blue–green direction, and a method for calculating the tint relative to an equiwhite step of the physical white scale used. Finally, he questioned the introduction of the D_{65} illuminant as a representation of sunlight, recognized the difficulty in accurately measuring the spectral power distribution of a light source, and discussed Ganz's proposal for a hyperbolic whiteness formula.

A.12 CIE Recommended Whiteness Formula

In 1986, the CIE recommended a set of formulae for calculating whiteness, standardizing the evaluation of white goods (International Commission on Illumination, 1986). They called for the use of illuminant D_{65} with either the 2° or 10° standard observer. The committee cautioned that the formulae were only to be used on near-white samples, and that the results were relative. The formulae that they recommended are:

(A.13) $$W = Y + 800(x_n - x) + 1700(y_n - y),$$

(A.14) $$W_{10} = Y_{10} + 800(x_{n,10} - x_{10}) + 1700(y_{n,10} - y_{10}),$$

(A.15) $$T_w = 1000(x_n - x) - 650(y_n - y),$$

and

(A.16) $$T_{w,10} = 900(x_{n,10} - x_{10}) - 650(y_{n,10} - y_{10}),$$

where W is the whiteness, x, y, and Y are the tristimulus values, and T is the tint factor for the 2° standard observer (the subscript *10* indicates the same values for the 10° standard observer). It should be noted that the CIE rejected the calls for a hyperbolic whiteness formula in favor of a simpler, linear version.

A.13 Uchida's Proposed Modification to the CIE Formula

In 1998, Uchida proposed a modification of the CIE whiteness formula (Uchida, 1998). It had been determined that if the formula was applied to samples containing fluorescent whitening agents of various tints, its results did not correspond well to visual observation. Uchida gave a brief review of problems with the formula before suggesting its modification.

Sanders and Wyszecki had shown that the subjective brightness of bluish samples was higher than that of yellowish samples with the same luminance factors. The original CIE formula gave a baseline for whiteness, preferring samples without reddish or greenish tints. As most samples treated with fluorescing agents move along this baseline instead of away from it, Uchida decided to use the baseline in his modification. He recognized that each observer has his own preferred white, and that the preferred white is often fluorescent tinged with red, blue, or green. His first modification to the CIE formula, however, ignored the hue preference of the observer, and simply added a term for tint (that is, distance from the baseline whiteness). He then looked at excitation purity and the maximum whiteness allowed for each luminance factor according to the CIE. He first suggested changing the limit of application for the formula from *5Y–280* to *5Y–275*. This still gave samples on one side of the maximum whiteness that were defined as white, while samples on the other side were not white. As this is illogical, he proposed a formula divided into two parts, one for "in-base" (that is, bluish or yellowish) samples and one for "out-base" (reddish or greenish) samples. For in-base samples, he recommended his first modification:

(A.17) $$W_{10} = W_{CIE,10} - 2\left(T_{w,10}\right)^2.$$

For out-base samples (with the 10° observer), he used $P_{w,10}$ to represent whiteness as follows:

(A.18) $$P_{w,10} = (5Y_{10} - 275) - [800(0.2742 + 0.00127(100 - Y_{10}) - x_{10})^{0.82} + 1700(0.2762 + 0.00176(100 - Y_{10}) - y_{10})^{0.82}].$$

This change took into account the effect of tint for samples that vary primarily in the direction of excitation purity. Equiwhiteness lines calculated with the new formula are ellipsoid and therefore more correlated to visual assessments. The correlation was tested and verified on both paper and fabric samples.

A.14 Surface Roughness and Whiteness

In 1999, Makarenko and Shaykevich published their work on the relation between paper whiteness and surface roughness (Makarenko and Shaykevich, 1999). They had noticed that glossy white papers were sometimes measured to be less white than matte or rough greyer papers, although visually the lighter glossy sample appeared whiter. They attributed this to the fact that only diffuse reflections were measured in accordance with the ISO 11475 standard procedure. They used the standard CIE formula for whiteness under D_{65}/10° conditions. This formula is based on tristimulus values, which are based on percent reflectance values, which are in turn based on the amounts of light scattered and reflected from a surface. The researchers focused on surfaces that had either small (0.0–0.08 µm) or large (>1 µm) irregularities. They measured the papers in three methods: the ISO standard with only diffuse reflectance, a modification of the ISO standard that included diffuse and specular reflectance, and using a cloudy sky as a light source while measuring diffuse and specular reflectance. The differences in Y_{10}, whiteness, and tint values were all small, regardless of irregularity size and illumination method, and the chromaticity coordinates of the samples did not vary by more than 0.15% for any sample. Makarenko and Shaykevich concluded that the ISO standard measurement method gave whiteness values that were artificially low for even slightly rough samples. The chromaticity coordinates of samples were independent of the measurement technique or the surface roughness.

A.15 Differences between Black and White

Although there are many parallels between the study of whiteness and the study of blackness, several factors important to the measurement of whiteness are insignificant for the evaluation of black samples. For example, the issue of fluorescence is not a factor in black materials as high concentrations of black dyes or multi-dye combinations absorb any fluorescence emitted by low levels of fluorescent brightening agents. The effect of surface roughness on blackness perception and measurement is also likely to be different from effects of this variable on whiteness.

Appendix C: Current Practices in Coloration

There are many industrially important black colorants in use today. The selection of black colorant is dependent on the material to be colored and the desired properties of the finished good. Dyes are colored compounds that are soluble in the medium in which they are applied, and are generally used on textile and paper goods. Pigments are colorants which are insoluble in the medium in which they are applied, and are commonly used in printing inks, paints, and materials such as ceramics and plastics.

B.1 Black Pigments

The earliest known black pigments were lampblacks, created by burning lamps under a cool surface, on which the pigment could collect. These lampblacks are part of a large category of pigments known as carbon blacks, which are amorphous impure carbon solids obtained by burning hydrocarbons. Other carbon blacks include acetylene black, thermal black, channel black, and furnace black (Patton, 1973). They are categorized based on their manufacture method. These pigments vary widely in particle size and depth of shade. Typical carbon blacks have published depth values ranging from 58 to 257 when evaluated visually on a scale from 0-260, although the method for determining these values is unknown.

Other common black pigments include graphite (pure carbon), natural black or micaceous iron oxide (obtained from magnetite), copper/chrome complex black (copper(II) chromate), aniline black, and logwood black lake (hematein). Carbon black accounts for nearly 50% of all printing inks in the United States, and yet the primary use of carbon black is in tires

(Patton, 1973). Carbon black is also used in paints, plastics (adding stability, UV protection, electrical properties, and physical reinforcement), and papers. Graphite is used in applications where carbon black may oxidize, or where lubrication is needed. It also improves electrical conduction more than carbon black. Iron oxides are used mostly in paints and coatings, especially in metal protective paints and metal primer coatings. Copper/chrome complex blacks are used in ceramics due to their high temperature stability. Aniline blacks are used to give deep matte blacks due to their high absorbance coupled with a low scattering coefficient. Logwood black lakes are relatively uncommon, and are mostly employed to color leather goods.

For the most part, black pigments are used alone to color goods. As discussed previously, black pigments are often tinted with small amounts of blue pigments to obtain rich blacks. Printers often add small amounts of cyan, magenta, and yellow inks to get obtain the desired hue (Kipphan, 2001). These small additions, however, are usually made on top of a black pigment, and do not comprise a large portion of the total pigment loading. It is possible to obtain very dark colors without the use of black ink in typical printing processes, but the use of a black ink greatly reduces ink use and thus cost.

B.2 Black Dyes

It is difficult to synthesize a dye molecule that will absorb all incident light. In fact, no “ideal” black dye has been produced to date (Zollinger, 2003). Many dyes absorb a large portion of incident light, however, and are nearly black. Two of the best “black” colorants are Aniline Black and Sulfur Black T, which are both polymeric dyes of somewhat unknown structure. The polymeric nature of these dyes allows them to absorb a very wide range of wavelengths of light. Zollinger suggested the addition of “compensating” dyes when using typical black dyes. The exact dye used will depend on the absorption spectrum of the black dye; a black that has low absorbance in the yellow region will give best results when a yellow dye is added. This is similar to the addition of blue pigments to carbon blacks mentioned above. One almost ‘ideal’ black colorant was reported synthesized in 1998. This dye is a monoazo dye with broad absorption peaks covering nearly the entire visible spectrum. Its structure is given in Figure B.1.

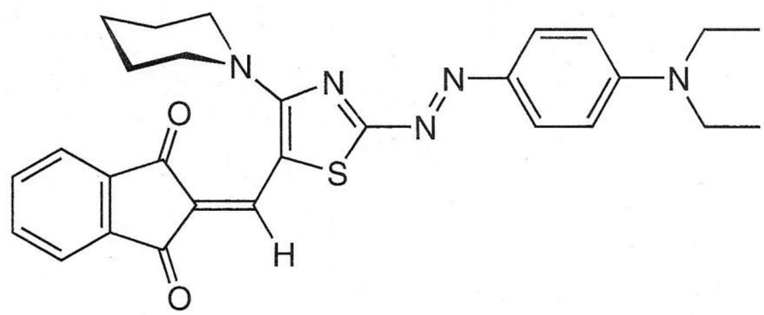

Figure B.1. Structure of the almost ideal black colorant.

B.3 Trichromatic Mixtures

Dyes and pigments give color to materials by absorbing certain wavelengths of light and reflecting all other wavelengths. Mixing these colorants has the effect of adding their absorbance curves, such that a mixture of a red, yellow, and blue colorant will result in a black mixture at sufficiently high concentrations of the individual dyes. Although the majority of black pigments seem to be single pigments, most commercial black dyes are blends of three or more chromatic dyes.

Figure B.2 shows TLC plates that were spotted with four different basic black dye mixtures. These dyes are commercially prepared mixtures, and their manufacturer's information is unfortunately unavailable. Although separation was not complete in some cases, the presence of red, blue and yellow dyes can be seen. It is evident that none of these dyes contain black colorants. In this case, TLC was performed in a 2:1 methylene chloride/methanol mobile phase. The dyes used are, from right to left, Maxillon RM 200, Permacryl Black FBL, Sevron Black M, and Astrazon Black FSW.

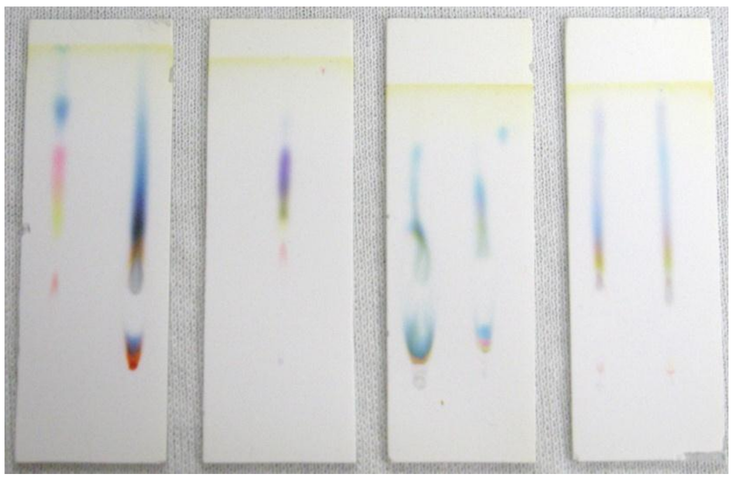

Figure B.2. TLC Plates of four basic black commercial dyes.

The Colour Index lists over 7000 dyes and pigments (including acid, azoic, basic, direct, disperse, mordant, pigment, reactive, solvent, sulphur and vat classes). Of these, only 667 are nominally black. Over half of these colorants are red, yellow or blue (Society of Dyers and Colourists, 1987). There are eight cataloged black basic dyes, and the commercial dyes separated above clearly do not include any of these. Numerous patents may be found detailing sets of dyes of various classes that are useful for trichromatic mixtures.

Appendix D. Failed Sample Generation Techniques

Plan A: Digital Design Lab Printer

Using Photoshop, squares of color approximately 3" on each side were made. The colors of the squares were designated using RGB values, and started at (0, 0, 0). Originally, the squares went up to (60, 60, 60) in steps of 5 values each, and it was eventually decided that samples needed to be closer to each other and black. Squares were then made that went from (0, 0, 0) up to (20, 0, 0), (0, 20, 0), (20, 0, 0), (20, 20, 0), (20, 0, 20), (0, 20, 20), and (20, 20, 20) in steps of 1 unit. The sets of samples up to values of 10 were printed on glossy paper using the digital printer in the Digital Design Lab. These all appeared to be uniformly black, and were undesirably glossy. Matte paper was obtained, and a set of 21 samples varying only in the red direction—from (0, 0, 0) to (20, 0, 0)—were printed on the new paper. These samples were left face up in the office for over a week when it was noted that they were fading to green. The fading was such that the samples were no longer nominally black.

Plan B: Photo Quality Home Printer

Since the ink used in the digital printer faded in combination with the matte paper, it was decided to try a typical photo quality ink jet printer. A set of samples including black, +10 in all directions, and +20 in all directions was printed. This was streaky, but this was likely due to low ink in the printer. Half of each sample was masked with posterboard and left in the office for a week. At the end of the week, the mask was removed, and it was noted that the samples had faded as well. The fading was less noticeable than with the digitally printed samples, but it was nonetheless present.

Plan C: Paint Chips

68 paint chips were obtained from Pittsburgh Paints, Duron, Sherwin Williams, and Benjamin Moore. These were classified as “black” or “not black” based on rough visual assessment. All chips were measured using the spectrophotometer in specular included and specular excluded geometries, and their L*, a*, and b* values were recorded. Generally, the L* values were around 30. From these data, it was decided (somewhat arbitrarily) that the measured cut-offs for “blackness” would be L* (specular excluded) < 30, |a*| and |b*| < 2.5, and |a*| + |b*| < 3.0. Using these criteria, 27 samples passed as black. However, these samples were not uniformly distributed in hue, and thus could not be used to identify the contribution of hue angle to perceived blackness.

Plan D: Matte Munsell Paper Samples

A hue circle of forty samples around a Munsell value of 3 were ordered from Munsell. This was the lowest value that a complete circle of matte samples could be produced. Unfortunately, these samples were all too light and highly chromatic to be considered black. For this reason, the first experiment was conducted using twenty glossy samples from Munsell.

Initial Dye Trials

A set of 12 black wool samples has been dyed in the Ahiba Texomat. The dyes used were Telon Yellow FG-01, Telon Red 2BN, Telon Navy AMF, and Isolon Black. These samples were placed in the dyebaths, agitated for approximately 15 minutes, and then heating began.

Heating was slow at first, reaching only 58° C after 50 minutes, and the heat rate was increased. After 105 minutes, the baths had reached 100° C and 1 mL glacial acetic acid was added to each. The samples continued to agitate for an additional hour before being allowed to cool. After dyeing they were washed using Apolloscour (2 g/L water) and acetic acid (1 g/L water) in 45° C water and dried at 120° F. The dried samples were measured with the specular reflectance included and excluded, and the differences in the colors of the dyed samples were exceedingly small.