Color Naming, Multidimensional Scaling, and Unique Hue Selections in English and Somali Speakers Do Not Show a Whorfian Effect

THESIS

Presented in partial fulfillment of the requirements for the degree Master of Science in the Graduate School of The Ohio State University

By

Ryan Lange, B.S.

Graduate Program in Vision Science

The Ohio State University 2015

Master's Examination Committee: Delwin T. Lindsey, PhD, Advisor

   Angela M. Brown, PhD
   Andrew Hartwick, OD, PhD
Abstract

The Whorfian linguistic relativity hypothesis is a highly contested model linking cognition and perception, in which substantial cultural-linguistic differences between languages are proposed, which significantly constrain in individual’s perception and world view. Due to the wide variation in color naming in cultures around the world, the constraints on naming imposed by color physics and the physiology of the visual system, and the near-ubiquitous nature of color communication, color naming has often been used as a model for testing linguistic relativity effects. However, by themselves, variations in color naming cannot give any useful information about whether color perception is affected by linguistic relativity. For this reason, we used color naming in conjunction with unique hue selection and multidimensional scaling (MDS) to test for the presence of Whorfian effects related to color naming for speakers of English and Somali, a language previously shown by our lab to show great inter-individual variation in both color naming and non-lexical measures of color perception.

In Experiment I we tested English subjects using a non-metric MDS paradigm with heteroluminant stimuli and found it to generally replicate fiducial orderings of stimuli in CIE UV space. Experiment II added a unique hue selection task, and English speakers’ unique hue selections in this task were concordant with those obtained from previous studies. We introduced a new method of MDS data collection, the binary sort
protocol, in Experiment III, which allowed us to quickly gather MDS data from English and Somali-speaking subjects. Somali color naming showed similar patterns to previous experiments by our lab, but we were unable to gather data from a sufficient variety of Somali informants to robustly test for Whorfian effects. Somali speakers’ MDS maps conformed more poorly to CIE UV space than English speakers’ maps, though analysis of stress indicated that Somali subjects may use two dimensions of unknown character in their judgments of color difference; additionally, numerous Somali speakers’ MDS maps and/or color naming showed evidence of purple-yellow affiliation, which is contrary to all previous color categorization and perception experiments. Experiment IV focused on English color naming, utilizing the gap statistic and the novel cluster stability analysis to find 15 shared chromatic categories of varying consensus and 4 distinct color-naming motifs among English informants, congruent with previous results from our lab. As in Experiment II and Experiment III, no evidence of linguistic relativity effects was found in our results, though we suggest that our color naming, unique hue, and MDS methodologies may have been insufficiently sensitive to detect these effects if present. We conclude that no definite linguistic relativity effects exist between English and Somali color naming, color difference perception, and unique hue selection, at least as measured by our paradigms.
Dedication

This thesis is dedicated to my family—my father Jeff, my mother Val, and my brother and sister David and Renae—and to my future family, including my fiancée Ellie, her parents Dan and Irene, and her brothers Tom and Joey.
Acknowledgements

I would like to thank my high school sciences teacher Bruce Vadnais and undergraduate thesis advisor Dr. John Watt, both of whom kindled my love of science, sharpened my mind, and curbed my tendency to blather.

Courtney Sexton’s tireless work in the lab made the completion of my projects inestimably easier.

Finally, Delwin Lindsey and Angela Brown have been as brilliant, patient, helpful, and fun as I could ever hope for mentors to be, and I cannot thank them enough for introducing me to the field of color science.
Vita

May 2008.......................................................... Shiloh Christian High School

May 2012.......................................................... Biology and Honors, The University of North Dakota

2012 to present.............................................. Candidate for Master of Science in Vision Science, College of Optometry, The Ohio State University

Publications


Fields of Study

Major Field: Vision Science
Table of Contents

Abstract ................................................................................................................................. ii
Dedication ............................................................................................................................ iv
Acknowledgements ............................................................................................................. v
Vita .......................................................................................................................................... vi
List of Figures ...................................................................................................................... xi
List of Equations ................................................................................................................ xiii
Introduction ....................................................................................................................... 1

Purpose ................................................................................................................................ 1
Color Perception and Appearance ...................................................................................... 1
  Color Physics, Color Mixing, and Trichromacy ................................................................. 1
  Chromatic Contrast and Opponent Processes ..................................................................... 5
Opponent Color Spaces and Chromaticity Diagrams ......................................................... 8
Color Dimensionality and Appearance Models ............................................................... 11
Color Ordering Systems ..................................................................................................... 12
Color Difference Experiments and Perceptual Space ....................................................... 13
Color Language and Color Categorization ....................................................................... 13
Project Overview ................................................................................................................ 19

Experiment I ....................................................................................................................... 21

Introduction and Study Design ........................................................................................ 21
Subjects ............................................................................................................................... 21
Apparatus and Stimuli ....................................................................................................... 22
Experiment III/IV MDS Stimuli ................................................................. 135
Unique Hue Stimuli .................................................................................. 136
Experiment IV English Color Naming Stimuli ........................................ 138
Experiment IV English Color Naming Stimuli (ctd.) ................................. 139
Experiment IV English Color Naming Stimuli (ctd.) ................................. 140
Appendix C: Details of Analysis................................................................. 142
Scree Plot Generation and Interpretation .................................................. 143
List of Figures

Figure 1. Hurvich & Jameson’s Opponent-Process Curves ........................................ 6
Figure 2. MacLeod-Boynton Cone Excitation Space .............................................. 10
Figure 3. Lab-Based Light Box .................................................................................. 23
Figure 4. Exp. 1 Stimulus Cards .............................................................................. 24
Figure 5. Exp. 1 MDS Solutions .............................................................................. 26
Figure 6. Unique Hues ............................................................................................... 30
Figure 7. Exp. 1 vs 2 MDS Solutions ...................................................................... 34
Figure 8. Exp. 2 Representative MDS Solutions ...................................................... 35
Figure 9. Exp. 2 Unique Hue Selections ................................................................... 36
Figure 10. Mobile Light Box Apparatus .................................................................. 41
Figure 11. WCS Color Chart .................................................................................... 42
Figure 12. Binary Sort Algorithm ............................................................................ 44
Figure 13. Exp. III English vs Somali Aggregate MDS Solutions ......................... 46
Figure 14. Exp. III Representative Somali MDS Solutions ..................................... 47
Figure 15. Exp. III Representative English MDS Solutions .................................... 48
Figure 16. Exp. 3 Scree Plots .................................................................................... 50
Figure 18. Exp. 3 Unique Hue Selection Frequency ............................................... 51
Figure 19. Exp. IV Modified Unique Hue Arrangement .......................................... 59
Figure 20. Exp. IV Color Naming Stimuli ................................................................. 60
Figure 21. Numbers of Raw Color Terms ................................................................. 68
Figure 22. High-Consensus Samples ........................................................................ 69
Figure 23. Raw Color Term Usage ........................................................................... 69
Figure 24. Optimal Number of Chromatic Clusters .................................................. 70
Figure 25. High-Consensus Clusters by $k$ ................................................................ 71
Figure 26. MUSTARD and “Gold” Variability ......................................................... 72
Figure 27. Modal Sample Glosses for Color-Naming Motifs .................................... 74
Figure 28. Exp. IV Unique Hue Selections ................................................................. 75
Figure 29. Exp. IV MDS Solutions ........................................................................... 76
Figure 30. Exp. IV vs. Exp. III Aggregate MDS Solutions ......................................... 77
Figure 31. Experiment I MDS Stimuli $L^*$ and $v^*$ Coordinates ............................ 88
Figure 32. Somali S5 cMDS Solutions with Full and Reduced Stimulus Sets ............. 91
Figure 33. Color Category and “Gold” Usage ............................................................. 94
Figure 34. Abney Effect-like Shift in English Color Naming .................................... 99
List of Equations

Eqn 1 ......................................................................................................................... 8
Eqn. 2 ........................................................................................................................... 9
Eqn. 3 .......................................................................................................................... 63
Eqn. 4 .......................................................................................................................... 63
Eqn. 5 .......................................................................................................................... 64
Eqn. 6 .......................................................................................................................... 65
Eqn. 7 .......................................................................................................................... 65
Introduction

Purpose

The purpose of this thesis is to investigate the link between color language and color perception. Specifically, we will be looking at linguistic color categories and their relationship to color difference perception as measured by multidimensional scaling of colors, and their relationship to the locations of unique hues in color space. This introduction is divided into three sections. The first section reviews the major theoretical developments leading to our classical understanding of color vision. The second section examines our current understanding of the link between color perception and color language. Finally, in the third section, an overview of the work described in this thesis is presented.

Color Perception and Appearance

Color Physics, Color Mixing, and Trichromacy. The past half-millennium has seen a renaissance in understanding of both the physical and the perceptual nature of light and color. In antiquity, color was thought of as an intrinsic property of light and hue as a function of the ratio of white to black. One of the earliest known treatments of color physics was by Aristotle, who posited that hues were composed of combinations of white and black, and that all hues existed on a linear spectrum between the two (Crone, 1999). Later philosophers such as Avicenna, Al Tusi, and Forsius devised multilinear models in which multiple continuums through various colors existed between dark and light;
however, none of these models satisfactorily explained relationships between hues (Kuehni and Schwarz, 2008). Then, in the early 17th century, the Jesuit mathematician Francois d'Aguilon combined these multilinear models with a subtractive color mixing model (Briggs, 2013). This was the final incarnation of the Aristotelian linear view of hue relationships, as Sir Isaac Newton would soon supplant it with his two-dimensional additive model.

In his *Opticks* of 1704, Newton published the results of his historic prismatic dispersion and recombination experiments. In addition to establishing the color wheel, whose basic geometry of hue and saturation relationships are used to this day, the experiments described in *Opticks* provided three invaluable insights into the nature of color perception and appearance: the indivisible nature of spectral hues and composite nature of white light; the physiologic, rather than physical, basis of color; and finally, spectral metamerism and its subsequent extension to physiologic trichromacy (Newton, 1671, as cited by Mollon, 2003). These findings have profoundly impacted vision science and will be referenced throughout this thesis, so detailed discussion is merited. Newton’s three insights are all related. Contrary to Aristotle's view that white light is physically pure and that colors are byproducts of light mixing with darkness as it interacts with matter, Newton showed that white light could be dispersed into the spectral hues, which could not be further subdivided but could be recombined to form white or other hues. Working from this principle, Newton reasoned that if white light can contain many visible hues but not appear similar to any, then color could not be an innate property of the light, but rather a property of human sensation. (It should be noted that while others such as Descartes had suggested color fictionalism this before him, Newton was the first
to give it experimental treatment) (Nassau, 1997). This hypothesis was strengthened by Newton's observation of metamerism: the appearance of any hue could be replicated by the combination of other hues, even if the original hue was physically pure, which indicated a many-to-one mapping of physical to perceptual states. This many-to-one mapping contradicted the color realist view that color was an intrinsic property of light, and metamerism provided the perceptual framework for ensuing experiments which now define all relevant models of color perception. While Newton’s model explained metamerism in the general sense, it did not directly account for the perceptual trichromacy noticed by his contemporaries, most notably Tobias Mayer (1775, as cited by Mollon, 2003). In the century following Newton's color experiments, four scientists would develop a physiological model of tuned transducers to explain this phenomenon. In 1757, Mikhail Lomonosov theorized three anatomical transducers excited by discrete red, yellow, and blue light, and in 1777 George Palmer extended this hypothesis to propose prescient models of color blindness and complementary color afterimages. Like Lomonosov, though, Palmer failed to reconcile his discrete trichromacy with the continuous nature of visible light. Meanwhile, John Elliot, MD hypothesized that there were a number of retinal primaries tuned to various portions of the continuous visible spectrum; however, he never enumerated them. It was Thomas Young at the turn of the 19th century who merged these ideas into the first model of three tuned transducers with different but overlapping spectral sensitivity functions. He proposed that the activity of these tuned transducers, in various combinations, gave rise to the appearance of all visible colors. Importantly, Young's concept of the tuned transducer did not require that the physical spectrum must be discretely partitioned, nor that the phenomenologically
simple hues must excite only one class of tuned transducer. This insight, combined with his wave theory of light, allowed his model to accommodate the continuum of visible hues as well as psychophysical trichromacy and metamerism (Lomonosov, 1757; Palmer, 1777; Elliot, 1786; Young, 1802, all as cited by Mollon, 2003).

Hermann von Helmholtz and James Clerk Maxwell share credit for the final paradigm shift in trichromatic theory and color appearance. Helmholtz used prismatic color mixing to determine the proportions of complementary colors needed to metamerize to white; the diagram of these results became the first modern chromaticity diagram, similar in shape to the standard CIE chromaticity space. Soon afterward, Maxwell performed color matching experiments with a device that allowed him to compare mixtures of monochromatic light to sunlight; later refinements of his method by Guild and Wright produced the 1931 CIE standard chromaticity space, which will be discussed in greater detail later (Kaiser and Boynton, 1996). Iterations of these experiments with protanopic subjects, plotting the first dichromatic confusion lines, allowed Maxwell to derive estimates of the spectral sensitivity functions of the photoreceptors (Maxwell, 1857, as cited by Longair, 2008). Improvements on Maxwell’s psychophysical method by König, Nuberg, Wright, and others using normal, deuteranopic, and tritanopic subjects anticipated the first physical and electrophysiological recordings of the cone spectral sensitivity functions. (König and Dieterici, 1886; Nuberg and Yustova, 1955, all as cited by Mollon, 2003; Wright, 1952).

Over the next half-century, Young, Helmholtz, and Maxwell’s psychological model of trichromacy was given an anatomical and physiological basis. The initial experiments, carried out on fish and amphibian retinas, showed three separate cone
absorption spectra in situ first by spectrophotometric densitometry and then by
electrophysiological recordings (Hanaoka and Fujimoto, 1957; Kaneko and Hashimoto,
1969; Svaetichin and MacNichol, 1958). Similar experiments on human retinal cells soon
followed (Brown & Wald, 1964; Bowmaker and Hartnall, 1980). The absorption spectra
obtained in these experiments were very similar to those derived from psychophysical
experiments, showing cone classes with absorption and action maxima in the short,
medium, and long wavelength regions of the spectra (~ 420, 530, and 560 nm). For the
purpose of brevity, these short-wavelength sensitive, medium-wavelength sensitive, and
long-wavelength sensitive cones will be referred to as S cones, M cones, and L cones.
These experiments all but proved the existence of three classes of cones with differing
sensitivity spectra to be the basis of physiological trichromacy; however, several of these
same experiments also appeared to vindicate an opposing theory which had divided the
color perception community for three quarters of a century.

Chromatic Contrast and Opponent Processes. While the color-matching
experiments of Young, Helmholtz, and Maxwell all strongly supported their three tuned
transducers model of physiological trichromacy, other perceptual phenomena such as
negative color afterimages, acquired color deficiencies, and differential effects of
stimulus size on blue-yellow and red-green discrimination did not (Judd, 1949; Hartridge,
1949; Hering, 1878; 1880, as cited by Hurvich and Jameson, 1957). Ewald Hering, a
German physiologist, drew upon the phenomena of negative afterimages to create his
opponent-process hypothesis of color vision. He hypothesized that there were three
channels in the retina, each of which transmitted information about color in a one-
dimensional and mutually exclusive fashion. The channels he proposed were blue-versus
yellow, red-versus-green, and black-versus-white. Hering supported this model with the observation that no color can seem to be both yellow and blue, nor both red and green (Hering, 1878; 1880, as cited by Hurvich and Jameson, 1957).

Following from this observation, Leo Hurvich and Dorothea Jameson developed an experimental model known as hue cancellation to test for evidence of an opponent process. Subjects viewed, for example, a bluish light; then added a yellowish light until the mixture appeared neither bluish nor yellowish. This process was repeated with bluish light to cancel yellowish light, reddish light to cancel greenish light, and greenish light to cancel reddish light. The relative amount of complementary light required to remove the original sensation for each stimulus light was plotted against that stimulus light’s wavelength, giving the opponent process curves seen in Figure 1.

Figure 1. Hurvich & Jameson’s Opponent-Process Curves. The x-axis indicates wavelength of stimulus light, while the y-axis indicates the relative amount of the complementary light used to cancel hue for that experimental wavelength. Open triangles below x-axis indicate yellow light used to cancel blueness, open above axis blue light to cancel yellowness; shaded triangles above axis indicate green light to cancel redness, shaded below axis red light to cancel greenness.
At about this time, electrophysiological experiments on fish and amphibian retinas by Svaetichin and MacNichol (1958) provided the first hints at the cone-opponent processes suggested by Hurvich and Jameson. However, the next piece in the opponent process puzzle lay not in the retina, but in the lateral geniculate nucleus (LGN) of the macaque. In 1966, De Valois, Abramov, and Jacobs measured the response of LGN cells to diffuse monocular stimuli of various spectral hues. In addition to “non-opponent” cells which increased or decreased firing rate in response to stimuli of all wavelengths, they also found “opponent” cells which increased firing rate in response to some wavelengths but decreased in response to others. They thus tentatively identified four categories of opponent cells: the first two, dubbed +B –Y and +G –R, gave excitatory responses to a range of short wavelengths and inhibitory responses to long wavelengths; the second two, +Y –B and +R –G, were excited by long and inhibited by short wavelengths. Initially, these four opponent cell classes appeared to satisfy all criteria for Hering’s opponency: their excitative and inhibitory maxima and minima sat cleanly in the blue, green, yellow, and red categories; their cross-over points coincided with intermediate colors such as cyan (blue-green), chartreuse (yellow-green), and orange; and the model even appeared to predict the Bezold-Brucke shift, a color appearance phenomenon wherein non-unique reddish and greenish hues appear to become more yellowish or bluish, respectively, as their luminance is increased (De Valois, Abramov, and Jacobs, 1966; Bezold, 1873 as cited by Pridmore, 1999).

However, several key data from their experiments did not quite fit, and their conclusions were not entirely empirical. For example, many of their long-wavelength-excited, short-wavelength-inhibited cells actually showed bimodal excitation maxima,
with a secondary maximum in the short-wavelength portion of the spectrum. It is interesting that these cells were included in the aggregate analysis, as they could be interpreted to provide a neurological basis for a purple versus mid-wavelength-encoding channel (or an extraspectral red versus green channel) rather than the spectral red versus green channel as interpreted by De Valois et al. Secondly, their classes’ maxima and minima differed from Hurvich and Jameson’s unique hues as derived from hue cancellation; for example, maxima for +R –G and minima for +G –R occurred at approximately 640 nm, while Hurvich and Jameson's unique red was extraspectral, requiring blue to cancel out any “yellowness.” Finally, the choice of four categories of opponent cells was based partially on \textit{a priori} knowledge: while a chi-squared test was used to rule out the possibility that there were only two classes, no further statistical treatment such as principal components analysis (PCA) was used to determine whether more than four categories would be a more appropriate model. The four categories seem to have been chosen primarily due to parsimony and conformation to existing theory.

\textit{Opponent Color Spaces and Chromaticity Diagrams.} The next step in linking the physiology of the visual system to color perception and appearance was the creation of opponent color spaces, which plot color by the opponent activity of the visual system. Donald MacLeod and Robert Boynton were the first to create such a model in 1979 when they detailed their cone excitation space. In this space, all colors were described as lying within isoluminant planes, each comprised of linear combinations of activity of multiple cone types as follows:

\[ S + M + L = 1 \]  \hspace{1cm} \textit{Eqn 1}
Equation 1 illustrates a hypothetical isoluminant plane in which all three cones are assumed to contribute to luminance; all colors in this plane would fall within a triangle, with each corner representing isolation of one cone type. This plane, however, does not reflect reality: when results from empirical studies such as Smith and Pokorny's flicker photometry (1972; 1975) are analyzed using this model, the triangle becomes skewed due to the S-cones' minimal contribution to luminance. Thus, MacLeod and Boynton's color space discounted S-cone contribution to luminance and used the following equation to define the isoluminant plane:

$$M + L = 1$$

Eqn. 2

In Equation 2, the S-cones are assumed not to contribute to luminance; thus, chromaticity along the abscissa of this isoluminant plane varies only as a function of relative excitation of the L- and M-cones, whereas the ordinate varies only as a function of the S-cone excitation. Wavelength is plotted as a function of excitation of the S-cones and relative excitation of the L- and M-cones, creating a curve enclosed within the plane (See Figure 2). This two-dimensional plane and the three-dimensional space to which it was generalized have the important property that when monochromatic spectral colors are additively mixed, the resulting chromaticity and luminance are simple vector additions of the original values when weighted by the contributions of L and M cones to luminance, and the contributions of all three cones to chromaticity. Additionally, in the range of monochromatic wavelengths between green and red, where the S-cone is modeled to have no contribution to chromaticity, only the value of $r$ for relative chromaticity contributions of the L- and M-cones is necessary to model any wavelength.
(MacLeod and Boynton, 1979). Figure 2 shows a visualization of MacLeod and Boynton’s cone excitation chromaticity space.

![Figure 2. MacLeod-Boynton Cone Excitation Space](image)

Figure 2. MacLeod-Boynton Cone Excitation Space. “r” on bottom right indicates relative L cone excitation, and “g” on bottom left indicates relative M cone excitation. x-axis scale is for r. \( r + g = 1 \) for any spectral hue, by definition. “b” on left indicates relative S cone excitation, which does not contribute to luminance, or to color appearance between ~520 and ~700 nm in this model. Lines represent lines of confusion for protanope and deuteranope through this constant-luminance space.

Expanding on MacLeod and Boynton’s cone-excitation chromaticity model and Smith and Pokorny's cone fundamentals, Krauskopf and others (Krauskopf, Williams, and Heeley, 1982; Derrington, Krauskopf, and Lennie, 1984) created a chromaticity model with axes for luminance, invariant S cone activity, and invariant L-vs-M cone activity, which allowed them to test independently for the presence of downstream S cone modulated and L-vs-M cone modulated chromaticity channels in the macaque. These experiments showed the presence of cell groupings in the lateral geniculate nucleus.
(LGN) which responded preferentially to signals from the L versus M cones, and from
the combined L and M cones versus S cones. However, these opponent axes do not
coincide with Hering’s opponent primaries of red-green and blue-yellow (Webster,
Miyahara, Malkoc, and Raker, 2000). More recent experiments have attempted to
characterize the transformations of opponency signals encoded in LGN into the
psychological opponent colors of Hering, and to determine their loci (Brouwer and
Heeger, 2013; Conway, 2001; Conway and Tsao, 2006; Horwitz, Chichilinsky, and
Albright, 2007). In general, these attempts have failed (see Lindsey & Brown, 2014 for a
review of the literature). It is striking that after more than one hundred years of research,
color vision scientists still do not have a satisfactory physiological explanation for color
opponency or unique hues.

Color Dimensionality and Appearance Models. Since electrophysiological
experiments have been unable to adequately characterize the dimensions of color
appearance, we must turn to psychophysical models in order to better understand these
aspects of color perception. Among the first of many psychophysical models of
perceptual color spaces were provided in the 1800s by Helmholtz and Maxwell, who
created rudimentary chromaticity diagrams based on the results of their trichromatic color
mixture experiments (Helmholtz, 1855; Maxwell, 1860, as cited from Mollon, 2003).
Grassman (1853) showed the mathematical peculiarity that any three experimental lights
can be matched to a stimulus light, though in many cases one of the experimental lights
must be added to the stimulus light in lieu of adding a negative amount of it to the match.
Because a multiplicity of experimental lights could be used for matching functions, it is
not obvious from first principles what the primaries for a trichromatic chromaticity
diagram ought to be. In 1931, the Commission internationale de l'éclairage (CIE) solved this problem by using purely mathematical primaries called "X," "Y," and "Z" to optimize their chromaticity space, known simply as the CIE model. These X, Y, and Z primaries correlate roughly to broadband mid-spectrum cone excitation, S-cone excitation, and luminance, respectively (Kaiser and Boynton, 1996). While this model, known as CIE XYZ, is mathematically tractable for specifying and matching colors of various sources, it is not perceptually uniform—colors that are equidistant in the Euclidean space of CIE XYZ are not necessarily judged as perceptually equidistant by the standard observer. In 1975, the CIE LUV model was created based on a set of transformations of the original CIE XYZ space; this space has been shown to be far more perceptually uniform than CIE XYZ. For this reason CIE LUV will be used to model color appearance within this thesis (Witt, 2007).

Color Ordering Systems. While color appearance models such as those developed by the CIE are excellent for modeling color matches or perceptual distance, they are limited in everyday application. Even colors of identical CIE XYZ or CIE LUV coordinates can give different sensations under different viewing conditions; conversely, natural objects show perceptual hue constancy within a surprising range of viewing conditions, though their objective colorimetric values may change dramatically (Abramov and Gordon, 1994). Color ordering systems, such as the Munsell and Color-aid schemes, use standardized physical stimuli such as pigmented cards or chips to facilitate color comparison and communication in situ (Munsell, 1905). These ordering systems generally vary along dimensions of hue, lightness, and saturation, and are often calibrated similarly to color appearance models in order to maximize perceptual uniformity within
these dimensions. The Munsell and Color-aid systems are well established in both color naming and color perception experiments, and will thus be the systems used in our experiments (Berlin and Kay, 1969; Kay et al., 2009; Lindsey and Brown, 2014; submitted; Shepard and Cooper, 1992; Bimler and Uuskula, 2014).

*Color Difference Experiments and Perceptual Space.* In conjunction with these models of color appearance and ordering, behavioral experiments have been used to determine the relationships between color appearance, dimensionality, and communication. The classic model for characterizing dimensions and differences in color perception, which will be used in this thesis, is multidimensional scaling (MDS). In MDS protocols, measures of similarity, dissimilarity, affiliation, or confusion between $n \times (n - 1)/2$ pairs of stimuli are sorted into a similarity or dissimilarity matrix which is used to calculate the eigenvalues of the comparisons and their corresponding eigenvectors. Then, for an $m$-dimensional solution, the $m$ largest eigenvalues are multiplied by their eigenvectors to give the $m \times n$ coordinates in the modeled perceptual space. MDS methodologies have been used extensively to model color and other perceptual spaces from the 1960s onward (Wickelmaier 2003; Shepard and Cooper 1992; Indow and Kanazawa, 1960).

**Color Language and Color Categorization.**

The interplay between color perception and the higher-level mechanisms of color cognition and language present unique challenges and opportunities to the vision scientist, the linguist, and the philosopher. Carefully modeling the links between the mechanisms of perception and behavior allows us to make inferences about the nature of cognition. However, linking propositions between the two are often difficult to find, and
some such as Saunders and van Brakel (1997) believe that the pathways from color perception to cognition to language are insufficiently constrained for non-trivial, universal inferences about the visual system to be made based on color naming. These doubts are usually rooted in a hypothesis known as linguistic relativity, an aggregate of observations and informal philosophies synthesized and popularized primarily by the linguists Edward Sapir and Benjamin Lee Whorf in the early 20th century. Neither Sapir nor Whorf ever explicated this hypothesis; the first concise definition was put forward by Roger Brown and Eric Lenneberg (1954), and a slightly modified version of this definition is generally used to this day. The hypothesis as they defined it is that "the world is differently experienced and conceived in different linguistic communities" and that "language is causally related to these psychological differences." Thus, significant differences in cognition of the natural world are expected across cultural-linguistic communities. One of the oldest models for testing the Whorfian linguistic relativity hypothesis is color naming. Surveys of the world’s color naming systems by sociologists and anthropologists, which date back to the 19th century, document an array of lexical partitions of color space. For example, a large number of the world's languages do not separate green and blue, while some have categorical terms only for "dark," "light," and "red." In some extreme cases such as the Karam language of Papua New Guinea, no term denoting "color" in the abstract has been found, though experiments have shown some degree of lexical partitioning of the spectrum (Kay et al, 2009; Bulmer, 1968). However, these phenomena do not demand a relativist interpretation. Some have argued that "limited" color lexicons reflect regionally common physiological constraints on color perception (such as heavy macular pigmentation screening short wavelengths and thus
impeding their discrimination) (Lindsey and Brown, 2002), or that the observed differences in color naming are indicative of the evolution of color terms towards some universal set, rather than differences in color perception. Under either assumption, the primary limitations on color cognition are not socio-linguistic, but physiological, and the evolution of color lexicons proceeds toward some maximum set of categories defined by those physiological limitations (Berlin and Kay, 1969). This position, known as linguistic universalism, posits that though syntactic and semantic variations do exist between languages, these variations do not affect cognition of the world in any significant way.

Brent Berlin and Paul Kay were among the first to investigate cross-cultural differences in color categorization from the perspective of color universalism. In their seminal study of color lexicons in twenty different languages, they identified eleven Basic Color Terms (BCTs): in English, these are black, white, brown, red, pink, orange, yellow, blue, green, purple, and gray. These basic color terms are said to encompass all consensus categorical color words in any given language, though not all languages should be assumed to use all of them. They also proposed stages in which these color terms come to be used by cultures: first, light and dark colors are separated; second, a category for reddish colors is added; third, a category for yellowish or greenish colors; fourth, yellowish and greenish colors are separated; fifth, bluish colors are separated from greenish ones; sixth, brown is categorized; and finally, purple, pink, orange, and/or gray categories are added (Berlin and Kay, 1969).

It should be noted that Berlin and Kay's initial study was flawed in a number of areas: most of the subjects were bilingual university students, which may have biased their responses to be more similar to those of monolingual English speakers (Saunders
and van Brakel, 1997). Additionally, certain ambiguous color terms appear to be interpreted to conform to a priori defined color category universals; and by using samples which varied only by hue, saturation, and lightness, Berlin and Kay were unable to investigate other dimensions of color categorization which may have been more salient to these speakers (Lucy, 1997). More importantly, though, it would be inappropriate to attempt to prove or disprove the hypothesis of linguistic relativity using color lexicon experiments only; indeed, the commonalities of basic color terms and their foci only appear to show that color naming (and thus, it is assumed, color cognition) is not entirely unconstrained.

Studies conducted in the wake of Berlin and Kay’s findings have attempted to use both lexical and non-lexical tasks in order to test for relationships between color language and cognition, with mixed results to date. The most robust evidence for Whorfian effects on color vision has concerned cross-cultural studies of color memory in which words and colors are linked (e.g., Roberson et al., 2000). Other studies have involved strictly non-verbal protocols such as measurements of discrimination thresholds (Witzel and Gegenfurtner, 2013) and dissimilarity judgements (Kay and Kempton, 1984), and have shown support for either universalism or a mixed model in which color naming contributes to a cognitive, rather than sub-cognitive, bias in color perception. The most extensive tests of the Whorfian hypothesis to date, however, have involved visual search (see Eckstein, 2011 for a contemporary review of the visual search paradigm). In this paradigm, subjects are asked to locate, as quickly as possible, a “target” item embedded in an array of “distractor” items within the search space. The independent variables in these experiments are theoretically important visual differences between target and
distractor items, while the crucial dependent variable is reaction time (RT). When color differences between otherwise identical target and distractor items are large, RTs are very short, regardless of the number of distractor items present in the display.

Tests of the Whorfian hypothesis, however, generally employ small target-distractor color differences and reaction times tend to be longer. Here, the critical parameter is whether target and distractor lie in the same or different categories. An effect of language on color perception is indicated when RT is faster for different-category (say, bluish-green vs greenish-blue) target/distractor conditions than for same-category conditions (say, both bluish-green or greenish-blue). Results with this paradigm have been mixed. RTs measured in early studies seemed to reveal Whorfian effects (Daoutis et al., 2006), even suggesting a localization of the effects to the left “language” hemisphere (Gilbert et al., 2006). However, other studies (e.g., Lindsey et al., 2010; Brown et al., 2011) failed to find Whorfian effects in visual search. While both Lindsey et al. and Brown et al. found systematic differences in RT with search condition, they were able to account for their results with a quantitative model based on low-level color vision without invoking language-related effects on visual search. A detailed discussion of the conceptual difficulties in testing the Whorfian hypothesis using the visual search paradigm may be found in Brown et al. (2011).

Meanwhile, the original color naming protocol employed by Berlin & Kay (1969) has been expanded with greater rigor in the World Color Survey (WCS), a repository of naming data for 330 colors from 110 mostly unwritten languages, which appears generally to support the findings of Berlin and Kay’s original study (Kay et al., 2009). Problematic methodology still abounds—for example, the vast majority of color naming
study designs are based on the implicit assumption that hue is the most salient dimension in color categorization across cultures, though empirical results suggest otherwise (Lucy, 1997). Yet, these studies’ results and the universalist inferences drawn from them form a more comprehensive and robust model of the relationships between color physics, color cognition, and color communication than any other model yet put forward (Kay and McDaniel, 1978; Witzel and Gegenfurtner, 2013).

One language which has shown unexpected behavior in color naming experiments is Somali, a Cushitic language thought to use five (Berlin and Kay, 1969) or six (Maffi, 1990) BCTs, glossing to *madow* (black), *cadaan* (white), *guduud/casaan* (red), *huruud/cawl* (yellow), *cagaar* (green), and *buluug* (blue, not found by Berlin and Kay). Previous studies (Lindsey and Brown, 2012; Brown, Isse, and Lindsey, 2015, submitted) have shown that Somali informants show great diversity in color naming, particularly in response to blue and green samples. Some informants partitioned blue and green similarly to English speakers (BLUE-GREEN users), while others used one color term to describe all blue-to-green colors (GRUE users), and still others were GRAY or DARK users. Additionally, many Somali informants used terms which glossed to YELLOW-OR-ORANGE to name purple and lavender samples, even though the two regions of color space are disjoint in perceptual models of color ordering. This behavior is intriguing, as categorical grouping of discontinuous regions of color space is proscribed even by supporters of linguistic relativity (Roberson, Davies, and Davidoff, 2000). These color naming patterns were largely corroborated by behavior in a free-sorting task, in which numerous Somali informants showed high affiliation of blue and green samples, along with purple-to-lavender and yellow-to-orange samples. Due to the variability of
Somali color naming, we wished to develop a non-verbal model of color perception to which we could compare Somali color naming results. This model would have to give metric results which could be scaled to known hue dimensions such as CIE LUV and accurately capture individual differences in color perception.

**Project Overview**

In order to directly test the linguistic relativity hypothesis that across-culture differences in color naming are correlated with differences in color perception, we wished to develop MDS and unique hue selection paradigms which could be used with English and Somali-speaking informants, then compare the results of these protocols to color-naming data. Experiments I and II tested and refined a non-metric MDS protocol employed by Shepard and Cooper (1992) in their study of English-speaking subjects. Experiment II also added a unique hue selection task, and correlational analyses were used to compare individual differences in unique hue selections and MDS solutions. Experiment III introduced a novel method of non-metric MDS data collection, the binary sort, which we used to more rapidly test English and Somali speakers. Experiment III also used a constrained naming task to collect color naming from Somali speakers. We originally hoped to test approximately equal numbers of Somali BLUE-GREEN, GRUE, GRAY, and DARK motif users in order to test our linguistic relativity hypothesis; however, for a variety of reasons we were not able to do so. Thus, Experiment IV focused on testing a more limited version of the linguistic relativity hypothesis, analyzing classical MDS, unique hues, and color naming data from English subjects only. This experiment also sought to extend Lindsey and Brown’s (2014) discussion of the English color lexicon, particularly with respect to the usage of high-consensus but non-basic
chromatic categories. To this end, the gap statistic analysis from Tibshirani, Walther, and Hastie (2003), as introduced by Lindsey and Brown (2006; 2009), was used to determine the number of shared categories among English informants. Also used was a novel treatment called cluster stability analysis, which determined the consensus of categorical clusters across numerous trials. Our analyses showed no evidence of linguistic relativity effects between English informants’ color naming, classical MDS, and unique hue selections; however, they corroborated Lindsey and Brown’s 2014 analysis of English color naming.
Experiment I

Introduction and Study Design

This pilot experiment tested the viability and robustness of a non-metric MDS paradigm using pairwise comparisons of both psychological primary and intermediary hues. It also sought to test the fit of the resulting MDS solutions to the samples’ fiducial positions in CIE LUV color space using Procrustes analysis. The experimental protocol was inspired by Shepard and Cooper (1992), who had subjects sort a stack of 36 stimulus pair cards consisting of all pairwise combinations of 9 sample colors which spanned the range of colors found on a classical color wheel. Cards were sorted from least to most dissimilar on the basis of the perceived differences between the colors on each card. An important distinction from Shepard and Cooper’s study is that our subjects viewed color pairs on a neutral gray background, rather than Shepard and Cooper’s bright white background. We also used 11 rather than 9 samples spanning the classical color wheel.

Subjects

Nine subjects participated in this study. Eight (four males) were first-year optometry students, and the other (male) was an optometry professor. All participants were native English speakers; one (ID 7) was natively bilingual, also speaking Arabic. Subjects were recruited on a voluntary basis. Copies of the voluntary participant consent forms and HIPAA privacy notices can be found in Appendix A. All subjects were required to pass the Farnsworth D-15 color arrangement test, give written consent and
HIPAA authorization, and show the ability to perform all protocol techniques. All protocols used in this and succeeding experiments were approved by the Ohio State University biomedical research IRB and were in accordance with Declaration of Helsinki principles.

**Apparatus and Stimuli**

A light box was used for all protocols performed in Experiment I. This light box was 143 cm wide by 75.5 cm tall by 71 cm deep and lit by four 121.92 cm Philips F40T12 Spectralite bulbs (Koninklijke Philips N.V., Amsterdam, Netherlands) rated at 5000K, 90 CRI, and 2200 lumens apiece. These light sources were mounted 57 cm above the table top, giving an average illuminance of 1970 to 2216 lux (Lindsey and Brown, 2014). The walls, floor, and ceiling of the light box were lined with white foam-core board, and a 94.5 cm by 60.5 cm Color-aid® Gray 4.5 sheet (Color-aid Corp., New York, NY) mounted onto the same foam-core board served as the experimental area. A 143 cm by 35.5 cm partial wall of white poster board was mounted to the top front of the apparatus in order to shield both experimenter and participant from direct illumination by the light sources. Figure 3 shows the author seated at the light box.
Stimuli for the dissimilarity sort protocol consisted of all pairwise combinations of green, cyan, blue, violet, purple, magenta, red, red-orange, orange, yellow, and chartreuse samples. Each sample was a 2.2-cm square, and sample pairs were mounted, axially centered and with 1.6 cm between edges, onto stimulus cards. Stimulus cards were made from Color-aid Gray 4.5 sheets, cut into 9.0 cm by 5.0 cm rounded-corner rectangles. See Figure 4 for several representative stimulus cards. With 11 colors represented, the total number of stimuli was \((11 \times 10) \div 2 = 55\). All cards were given unique two-letter IDs. The first letter of each ID corresponded to the first color compared, and the second letter to the second color compared. Letter-color keys ranged from “A” for green to “K” for chartreuse, following the arrangement given above. Each card was affixed with a barcode with its two-letter ID, and sort order was recorded by these IDs.

**Figure 3. Lab-Based Light Box.** The author seated at the light box used in this experiment, performing the color dissimilarity sorting task.
Methods

Participants were told that they were going to be sorting cards based upon the degree of dissimilarity of each card’s color pair. All stimulus cards were given to the participant in an unordered pile, and the participant was asked to take cards from the pile *ad libitum* and sort them into as many columns as necessary. Cards were to increase in dissimilarity from participant’s left to right between columns and from back to front within columns. Instructions were repeated or clarified as requested by the participant. Each participant was instructed at the beginning of the task that she would be able to make any amendments she deemed necessary to the sort ordering at any time during or at the end of task, but that no instruction would ever be given on the “proper” ordering of the stimuli.

Analysis

Participants’ color dissimilarity sorts were scanned into an Excel file using an LZ410-WDP barcode scanner (Worth Data, Santa Cruz CA) and the ranking data were
analyzed in a custom MATLAB script. The script sorted the ranking into a dissimilarity matrix, then performed non-metric MDS in two and three dimensions on this matrix and saved the stresses of the solutions. Procrustes analysis was then applied to rotate and isotropically scale the MDS coordinates to best match the stimulus fiducial points, as plotted in CIE LUV color space. Procrustes analysis gave transformed coordinates in UV (hue only) and LUV (hue and standardized luminance) space and goodness-of-fit $d$ to UV and LUV space, which were saved and used in further analyses.

**Results**

Generally, MDS produced well-organized perceptual hue maps, in which the order of the hues in two dimensions, as determined by MDS, corresponded to their perceptual order in color space. Reversals of adjacent hues were common in the individual MDS maps, however, and the red-orange and orange coordinates were uniformly displaced toward red and yellow, respectively. See Figure 5(A-C) for several representative hue maps. Kruskal stress values ($\mu = 0.0796; \text{std} = 0.0218$) are generally low, indicating that the MDS maps reflect consistent dissimilarity judgments by subjects. Procrustes goodness-of-fit $d$ ($\mu = 0.1380; \text{std} = 0.0613$) values indicated generally fair projections of scaled maps onto CIE LUV uniform color space. When sorts were aggregated (averaged) before MDS and Procrustes analyses, both stress and $d$ decreased (stress = 0.03671, $d = 0.0841$), as would be expected due to population norm effects. However, the displacement of orange and red-orange was still apparent. Figure 5(D) shows Procrustes-scaled MDS maps for aggregate sorts in two dimensions.
Figure 5. Exp. 1 MDS Solutions. Representative MDS solutions from Exp. 1. Triangles connected by lines represent stimulus fiducial points; circles represent MDS solutions. A. Best-ordered solution (stress=0.053; d=0.098). B. Standard solution (stress=0.088; d=0.189). C. Worst-ordered solution (stress=0.106; d=0.211) with numerous reversals of colors. D. MDS solution of aggregated sorts (stress = 0.03671, d = 0.0841). Note separation of orange and red-orange solutions, clustering of red-orange with red and purple with violet, and stretching along the CIE-v* axis.

Mean stress and d values for men (stress $\mu = 0.0789$; d $\mu = 0.1111$) and women (stress $\mu = 0.0805$; d $\mu = 0.1715$) were compared using Mann-Whitney U tests, which found no significant difference in means for either stress ($U = 0.713$) or d ($U = 0.270$). A Spearman rank-order test showed no significant correlation between stress and d ($\rho = 0.3$, p = 0.433).
Discussion

Inspection of perceptual hue maps, along with the generally low stress and $d$ values, indicated that the color dissimilarity sort and non-metric MDS created robust perceptual color maps, which generally replicated the results of Shepard & Cooper (1992) and recapitulated the metric properties of test samples, as specified by UV hue coordinates. This is significant, because most recent MDS experiments, such as those performed by Boehm et al. (2014), have used isoluminant stimuli, concerned that luminance effects would severely alter hue scaling if heterochromatic stimuli were used. To the contrary, our results show that hue ordering is generally well preserved when MDS is applied to sorts of heterochromatic stimuli. However, the cases of orange and red-orange are concerning, and two participants’ sorts (22.2%) gave $d$ values above 0.2, indicating poor fits to CIE LUV space. Additionally, numerous participants noted that the red-orange stimulus appeared “off” against the gray background. The red-orange sample and gray background were nearly isoluminant (RO-EX Y = 94.1; Gray 4.5 Y = 90.58); this likely contributed to the unusual appearance of this sample, and to the scaling effect noted in Results. Finally, the gray background was roughly median in luminance compared to dark and light samples (magenta Y = 47.77; yellow Y = 302.3), which we suspected might cause luminance to be weighted as a factor in dissimilarity sorting. For these reasons, we wished to alter the stimulus cards in order to reduce the effects of luminance on scaling, hoping this would improve the fit of sorts to CIE LUV space. To do this, we followed the example of Shepard and Cooper (1992) and changed the background to white, so as not to provide an easy reference for luminance comparisons across samples.
Experiment II

Introduction and Study Design

In order to minimize the luminance effects seen in the MDS solutions in Experiment I, this second portion of the pilot experiment tested the same non-metric MDS paradigm with a modified set of eleven hues displayed on a white background rather than the gray background used in the first experiment. A unique hue protocol was also introduced, in order to investigate the relationship between MDS solutions and unique hues. Our hypothesis was that MDS coordinates for the psychological primaries (red, yellow, green, and blue) and unique hue selections would co-vary (both within and between MDS maps and unique hue selections), which would be indicative of systematic differences in color perception between subjects.

Subjects

Thirteen first and second year optometry students were tested, seven of them female. Criteria for participant recruitment and inclusion were identical to those in Experiment I. Four subjects (two male) had also participated in Experiment I.

Apparatus and Stimuli

Participants performed all protocols in the light box described in Experiment I. Stimuli for the color dissimilarity sort protocol were similar to those selected in Experiment I, with several notable differences. Stimulus cards were constructed from “Ice White” Canford cardstock (L = 500.2; x = 0.341; y = 0.370) (Daler-Rowney Ltd.,
Bracknell, United Kingdom) and were the same dimensions as gray cards used in Experiment I, while color stimuli were 2.5-cm diameter circles rather than rectangles. Several color stimuli, such as green, blue, and magenta, were altered slightly in order to decrease lightness differences between samples, and red-orange and violet samples were removed in favor of yellow-chartreuse and gold to give 11 stimuli (55 stimulus cards) with more evenly distributed perceptual hue and lightness in the UV plane of CIE LUV. A table of the Color-aid samples used in the construction of stimuli, their Munsell equivalents, and their CIE LUV colorimetric data can be found in Appendix B.

Unique hue stimuli consisted of 40 Munsell glossy color samples (Munsell Color Corp., Grand Rapids, MI). Four subjective criteria were used in the selection of samples: first, the hues of samples were chosen to provide uniform differences in hue between adjacent samples; second, the overall contrast in value was minimized; third, the contrast in value between adjacent samples was made as close to uniform as possible; and fourth, chromas were selected to approximate those of our MDS stimuli. See Figure 6 for the unique hues in CIE UV space and arranged as in the protocol, and Appendix B for colorimetric data for unique hue samples.
**Methods**

Except for the changes made to the stimulus cards, the protocol for the color dissimilarity sort was the same as in Experiment I.

In the unique hue phase of an experimental session, the experimenter first arranged the stimuli into a rough ovoid or bean-like shape. This arrangement was chosen to prevent participants from drawing upon the spatial relationships of a “color wheel” to inform their unique hue choices. After the unique hue stimuli had been arranged, each participant was asked to select “a red that appears neither yellowish nor bluish, a yellow that appears neither reddish nor greenish, a green that appears neither bluish nor yellowish, and a blue that appears neither greenish nor reddish.” Participants were encouraged to hold samples close together if they felt it would help them in the task.

**Analysis**

Non-metric MDS and Procrustes analysis were used to visualize subjects’ dissimilarity sorts as in Experiment I.
Our primary hypothesis to be tested in this experiment was that MDS coordinates and unique hue selection coordinates ought to co-vary in angular position, both within themselves and with each other, in some predictable fashion; we termed this hypothetical covariance perceptual relativity. We envisioned two possible types of relativity effects: the first, hue space shifts, would be indicated by positive correlations in angular space; and the second, hue space compressions/expansions, would be indicated by negative correlations in angular space. For example, if MDS coordinates for red and green were consistently positively correlated, such that positive angular deviations from the mean for red MDS coordinates were usually paired with positive angular deviations from the mean for green MDS coordinates, and negative deviations with negative deviations, this would be considered a hue shift effect along the red-green axis for MDS maps. On the other hand, if for example unique hue coordinates for blue and green were consistently negatively correlated, such that a positive angular deviation from the mean for unique blue were generally paired with a negative angular deviation from the mean for unique green, and vice versa, this would be considered a hue compression/expansion effect along the blue-green axis for unique hue selections. We searched for evidence of hue shifts and hue compressions/expansions within and between MDS and unique hues.

First, we tested for perceptual relativity effects within unique hue selections. Because of the low number of subjects and total unique hue samples selected and the discrete nature of unique hues selected, we used polar rank, rather than polar angle, in these tests. Lists for unique red, unique yellow, unique green, and unique blue were converted into rank-coded lists, where rank was indicated by relative directionality around the Munsell color circle shown in Figure 6(A) (p. 30), such that samples
clockwise to violet were of increasing wavelength. For each set of unique hue selections, the most clockwise sample selected was coded as 1, and the most counter-clockwise was coded as $r_s$, the range of samples selected. Spearman rank-order tests were then performed on all combinations of the rank-coded lists (red/yellow, red/green, red/blue, yellow/green, yellow/blue, and green/blue). Null hypotheses for all tests were that hue selection lists were independently distributed, and the null was rejected for $p < 0.00417$ (two-tailed test with correction for six comparisons).

For any pairs of unique hue lists found to be significantly correlated, a permutation test (Nichols and Holmes, 2002) was performed on one list, and the Spearman $\rho$ was found for all permutations of this list with the unaltered second list. The number of permutations giving a $\rho$ magnitude equal to or greater than that of the fiducial comparison was divided by the total number of permutations, and the probability of the two data sets being dependently distributed was determined by this ratio.

To test for perceptual relativity effects within MDS and between MDS and unique hues, we performed circular correlation analysis on residual hue angles. First, informants’ red, yellow, green, and blue MDS coordinates, as well as their unique hue selections, were converted into CIE UV hue angles (with UV = 0,0 as the origin for the polar plot), and across-subject means were calculated for MDS and unique hue angles. Then, residual hue angle lists were calculated by subtracting each participant’s MDS and unique hue angles from their respective mean hue angles. To test for hue space distortions within MDS, we performed circular correlations for all combinations of MDS residual hue angle (RHA) lists (red RHAs/orange RHAs, red RHAs/gold RHAs, …, violet RHAs/blue RHAs), with the null hypothesis that the two lists were independently distributed, using
the MATLAB CircStat toolbox’s circ_corrc test of circular correlation (Philipp Berens, Max Planck Institute for Biological Cybernetics, Tübingen, DE). Circular correlation measures covariance of data lists similarly to traditional correlation, but adjusts for the proximity of angles near 0 and 360 degrees, which are close in angular space but far apart on the real number line. Correction was made for multiple comparisons. Then, to test for relativity effects between MDS and unique hue selections, each list of residual MDS angles was compared using circ_corrc with the list of residual unique hue angles, with the null hypothesis again that the two lists were independently distributed.

Results

MDS produced hue maps with qualitatively better fit and less reversal of adjacent hues than in Experiment I. Mean Kruskal stress values ($\mu = 0.0815$, std = 0.0242; 11 “good” stress < 0.1; 2 “fair” < 0.2) and $d$ ($\mu = 0.1106$, std = 0.0554) were comparable to Experiment I; however, only 1 MDS coordinate set out of 13 (7.70%) fit poorly to fiducial points ($d > 0.2$), indicating that MDS maps of participants’ sorts adhered comparatively well to CIE UV space. MDS maps, two-dimensional stress, and $d$ for repeat subjects were very similar to values from Experiment I. When sorts were aggregated and subjected to MDS, the resultant stress was 0.03163 and goodness-of-fit $d$ was 0.0741. See Figure 7 for comparison of the aggregate results from the new and old
protocols and Figure 8 for representative hue maps. Finally, circular correlational analysis showed no systematic covariance among MDS colors’ residual hue angles.

Figure 7. Exp. 1 vs 2 MDS Solutions. A. MDS solution of aggregated sorts from Experiment I. B. MDS solution of aggregated sorts from Experiment II. Note the presence of nearly-conflated samples in Experiment I (purple and magenta, red and red-orange), which is not seen in the results of Experiment II. However, in Experiment II as in Experiment I, we do see the displacement of orange and red, as well as the apparent stretching along the v* axis.
Unique hues were selected from very limited ranges, with blue having the largest range of selections at five. Modal unique hues were 7.5R 5/12 (red), 5Y 8/12 (yellow), 2.5G 5/10 (green), and 2.5PB 5/10 (blue). See Figure 9 for modal unique hue selections in CIE UV space, and a histogram of selection distributions for all four hues. Spearman rank-order analyses of the ranked unique hue lists initially showed a significant negative correlation between red and green lists ($\rho = -0.7407; p = 0.0224$). Of all 1,512
permutations of the list of ranked unique green selections, 84 showed the same
correlation to the red list as did the unaltered green list ($\rho = -0.7407$). Thus, the
probability that the red and green lists were not independently distributed was calculated
to be $84/1512 = 0.0556$, slightly above the criterion for significance for two-tailed testing
($p < 0.025$).

**Figure 9. Exp. 2 Unique Hue Selections.** A. Modal unique hue selections in UV space. Circles: modal
unique hue selections for red, yellow, green, and blue. Lines: connect u*v* coordinates of MDS stimuli. B.
Histogram of unique hue selection frequency

Circular correlation analyses of MDS and unique hues’ residual angles showed no
significant correlations between MDS maps and unique hue selections, indicating no
systematic perceptual shifts or compressions.

**Discussion**

As in Experiment I, the color dissimilarity sort protocol and non-metric MDS
produced perceptual hue maps that closely fit CIE UV coordinates of the sort stimuli.
MDS solutions based on sorts of the new protocol fit marginally (though not
significantly) better to CIE UV than in Experiment I, with a lower percentage of “poor”
fits with Procrustes goodness-of-fit d > 0.2. Additionally, there appeared to be fewer instances of color reversals in MDS solutions from the new protocol. However, two trends from Experiment I persisted in this protocol. First, red and orange MDS solutions were relatively distant in MDS solutions as compared to fiducial coordinates; and second, overall luminance effects appeared to still have some degree of salience, as evidenced by stretching of solutions along the v* axis, where luminance differences between yellowish and purplish colors appear to have biased sorting to some degree.

Hue values and dispersion of unique hue selections agreed well with the established literature (Webster et al., 2000). Our experiments showed no systematic shifts, compressions, or expansions within subjects’ unique hue spaces; nor did they show any significant relationships between unique hue selections and MDS results. However, we hoped to continue to improve our protocols with further experiments, and to test more subjects for greater statistical power of these comparisons.
Experiment III

Introduction and Study Design

In this experiment, we used MDS analysis to compare English and Somali perceptions of color differences. Previous studies (Lindsey & Brown, 2014; Brown et al., submitted) have shown that many Somali informants lexically categorize colors very differently from English ones. Of particular interest here are the differences in the colors that are named “purple,” “green,” and “blue” in English. Somalis often use the same name for purple (and/or lavender) and yellow color samples (Lindsey and Brown, 2014; Brown et al., submitted). Somalis also vary significantly in how blues and greens are categorized. While some lexically distinguish between these colors in an English-like way (GREEN-BLUE motif), other Somalis often use single terms that gloss to the universal categories GRAY, DARK, or GRUE (GREEN-OR-BLUE). Given these differences in color naming between Somali and English informants, Experiment III was designed to test for an association between differences in color naming and perception of color dissimilarity in these two cultures. Specifically, the expectation was that MDS would reveal that Somalis who use the same names for pairs of colors would perceive these colors as less dissimilar than English subjects who assign the colors different names. The original experimental design called for data collection on English informants and approximately equal numbers of Somali informants with GREEN-BLUE, GRAY, DARK and GRUE color naming systems; however, due to difficulties with our
interpreter we were only able to test a limited number of Somali participants. Only one Somali informant tested for this experiment used the GRUE color naming system, and all others used the GREEN-BLUE system. Additionally, due to the limited amount of Somali data that could be gathered, a fourth protocol (Experiment IV) would be used to test the link between color naming and color dissimilarity perception, using only English-speaking subjects.

Because we planned to collect color naming as well as MDS and unique hue data in the field, we wanted the dissimilarity-sorts for MDS analysis to be as time-efficient as possible. Therefore, in Experiment III, we adopted a novel binary sort methodology to the MDS protocol and used a reduced stimulus set of eight colors. Both Somali and English participants used this binary sort protocol, which proved to be an improvement over previous methodologies.

**Subjects**

Seventeen subjects participated in this study: seven English speakers (four OSU College of Optometry students and three OSU Mansfield undergraduates; five female) and ten native Somali speakers recruited from the Columbus, OH area (six female). Somali speakers varied in age (25-72 years; mean = 50.5), but mean ages for men and women were not significantly different. All Somali speakers used the standard dialect, and most spoke other languages minimally (generally English and/or Arabic), though one subject was conversant in English. On average, Somali speakers had lived in the United States for 9 years prior to testing (range = 3-14 years). One male Somali participant (ID 6) was tested before the unique hue selection protocol was developed, and thus only MDS and color naming data are included for this subject. Optometry students were tested
in the OSU College of Optometry. Somali subjects were tested in the field at a Somali Community Center and undergraduates were tested at the Ohio State University Mansfield campus, both using a mobile light box described in Apparatus. All subjects were compensated with ten-dollar gift cards from Target® or Kroger® as desired. Criteria for inclusion were the same as in Experiment II, with the exception that Somali subjects would use an interpreter. Paid consent forms in English and Somali, and the HIPAA form in Somali, can be found in Appendix A.

**Apparatus and Stimuli**

Optometry students were tested at the lab-based light box described in the Apparatus section of Experiment I. Somali and undergraduate subjects were tested at a mobile light box was used which duplicated all of the experimentally important characteristics of the laboratory-based apparatus. The participant and experimenter sat at a folding table, 71.5 cm tall with a 65 by 46 cm approximately rectangular top, covered by a 56 by 30.5 cm Color-aid® Gray 4.5 sheet. The light source, consisting of four Philips TL90 F17 T8/TL950 bulbs rated at 5000K, 98 CRI, and 850 lumens apiece, was suspended 50 centimeters above the table top, giving an average illuminance of ~890 lux. Two pieces of gray felt approximately 35 by 63 cm were attached to each side of the light source in order to mask the experimenter from the participant and to shield both experimenter and participant from direct illumination by the light sources. Figure 10 shows a photograph of the mobile light box.
To shorten the duration of the dissimilarity sort protocol for Somali subjects, a subset of the stimulus set from Experiment II was used, representing all pairwise combinations of red, orange, yellow, chartreuse, green, cyan, blue, and purple. See Appendix B for a list of samples’ Color-aid designations, along with their colorimetric data and Munsell equivalents. With eight colors represented, the total number of stimuli (color pairs) was \((8 \times 7) ÷ 2 = 28\).

Somali subjects performed a color-naming protocol using a 23 sample subset of colors from the WCS chart, as employed in a recent study by Lindsey et al. (2015) comparing color naming by Hadza (Tanzanian nomadic hunter-gatherers), Somali, and English informants. Figure 11 shows these samples within the WCS chart.
Figure 11. WCS Color Chart. Red dots indicating samples used in the Somali color naming experiment, along with our names for these colors.

Unique hue stimuli and their presentation were unchanged from Experiment II.

Methods

As in Experiment II, color dissimilarity sorting was performed first for all participants. In dissimilarity sort sessions, participants were told that they were first going to be judging the differences between pairs of colors on a card, then comparing those differences between two cards. First, the experimenter showed a participant a single card and instructed the participant to consider the degree to which the two colors on the card appeared different, with special emphasis placed on judging the appearance, rather than the connotations or feelings evoked by the colors. Next, the experimenter showed a second card and gave the same instructions. Finally, the two cards were placed side-by-side, and the participant was instructed to point to the card whose two color samples appeared more different from one another. This process was repeated as necessary if a participant asked for help with the task, or if she visibly compared colors across rather than between cards, for example by pointing to a color on one card and then to a color on the other card. In order to help avoid this confusion of within-card versus between-card comparison, the first several cards were chosen so as not to duplicate any colors between
cards; for example, if the first card chosen was orange and red, the next card would not include red or orange.

Following this initial comparison, a binary sort procedure was used to rank each stimulus pair. In this procedure, each new stimulus card was compared first to the middle-ranked reference card; the stimulus card was then moved halfway “up” or “down” the remaining search space depending on whether its colors were judged more or less different than the reference card, and compared to a new reference card. Each comparison divided the search space in half, until a final judgment placed the comparison card between a reference card and an endpoint (either the end of the list or a previously-used reference card). Figure 12 shows an example of this binary sort procedure. Because each comparison divides the search space in half, the maximum number of comparisons $c_m$ to be made for a given number of reference cards $n$ is $\text{floor}(\log_2 n) + 1$. This is the lowest possible $c_m$ for any non-Bayesian sort algorithm. Bayesian modification was deemed inappropriate, as it would imply “correct” and “incorrect” responses, possibly introducing response bias.
Figure 12. Binary Sort Algorithm. Stimulus card (small, orange/yellow) is first compared to a reference card median in the search space (here, red/blue). If the subject indicates that the colors on stimulus card are more dissimilar than those on reference card, the stimulus card is moved halfway “down” the remaining search space; if stimulus color are less dissimilar, the stimulus card is moved halfway “up” the remaining search space. In this example, thick solid arrows indicate card movements as dictated by the hypothetical subject; dotted arrows indicate card movements for the opposite responses. Figure shows the maximum number of comparisons possible (4) for this number of reference cards (12).

In color naming sessions, Somali participants were first instructed that they would be providing one-word color names for 23 colored samples. They were further instructed to use color names which would be applicable to any object, and which they would employ in everyday conversation to describe the color of an object to another person.
(after Lindsey and Brown, 2014; Berlin and Kay, 1969; Kay et al., 2009). If a participant provided more than one color word for a sample (such as “buluug-cagaar”—blue-green) or used relational modifiers (such as “[like] cagaar”), he was asked to rename the sample with a monolexemic color word. “Don’t know” was an allowed answer in this protocol.

Unique hue selections were performed in the same manner as in Experiment II, with the caveat that Somali informants would not be required to select both unique cagaar and buluug if they did not understand the distinction.

**Analysis**

English and Somali-speaking informants’ dissimilarity sorts were analyzed using non-metric MDS and Procrustes analysis as in Experiments I and II. The resulting solutions were also searched for hue reversals, defined as any orderings of MDS solution in angular UV space that do not match the order of the fiducial points in the same space.

Analyses of unique hue selections for all informants were performed in the same manner as in Experiment II. English informants’ MDS solutions and unique hue selections were compared via circular correlation in the same fashion as in Experiment II. Somali color naming was compared qualitatively to non-metric MDS solutions.

**Results**

*Non-Metric MDS.* Non-metric MDS on binary sorts by English-speaking informants generated well-organized perceptual maps, with mean stress = 0.0569 and mean Procrustes goodness-of-fit to UV space $d = 0.1346$. Neither median stress nor $d$ was significantly different than in Experiment II ($U = 0.0324$ and $U = 0.9368$, respectively; two-tailed tests). English and Somali MDS solutions were compared with two-tailed Mann-Whitney U tests and found to be non-significantly different in median
two-dimensional stress \((U = 0.0297)\) and significantly different in median \(d\) \((U = 0.0066)\). Four English speakers showed “good” stress < 0.1 and three showed “fair” stress < 0.2, and all but one “good” to “fair” \(d\). Somali speakers showed mean stress = 0.1089 (6 “good” stresses; 3 “fair” stresses; and 1 “poor” stress) and mean \(d = 0.4652\) (1 “good” fit, 10 “poor” fits). See Figure 13 for aggregate MDS solutions for English and BLUE-GREEN-using Somali informants, Figure 14 for representative English speakers’ MDS solutions, and Figure 15 for representative MDS solutions for BLUE-GREEN-using Somali informants, as well as the MDS solution for the GRUE-using Somali informant.

![Figure 13. Exp. III English vs Somali Aggregate MDS Solutions.](image)

**Figure 13. Exp. III English vs Somali Aggregate MDS Solutions.** Triangles connected by lines represent stimulus fiducial coordinates; circles represent MDS solutions. A. MDS solution of aggregate English sorts (stress=0.021; \(d=0.052\)). B. MDS solution of aggregate (minus S10, “GRUE” speaker) Somali sorts (stress=0.050; \(d=0.069\)). Note the persistence of \(v^*\) axis stretching and separation of red and orange in the English aggregate, and the lack of such \(v^*\) stretching in the Somali aggregate. Compare also the Somali aggregate to individual Somali speakers’ solutions, which generally show much poorer fit of solutions.
Figure 14. Exp. III Representative Somali MDS Solutions. A. Best-ordered solution (stress=0.047; d=0.076). While there are no hue reversals, purple very nearly crosses the v* axis. B. Standard solution (stress=0.076; d=0.407) with displacement of chartreuse and purple across the u* and v* axes, respectively. C. Worst-ordered solution (stress=0.152; d=0.659) with general disorganization. D. GRUE speaker’s solution (stress=0.0750; d=0.596) with blue-cyan and green-chartreuse couplings reversed.
Figure 15. Exp. III Representative English MDS Solutions. Triangles connected by lines represent stimulus fiducial coordinates; circles represent MDS solutions. A. Best-ordered solution (stress=0.071; d=0.057). B. Good stress, fair d solution (stress=0.045; d=0.144) with reversal of blue and cyan. C. Fair stress, fair d solution (stress=0.175; d=0.100) with no focal reversals but unusual shape. D. Worst-ordered solution (stress=0.198; d=0.415) with reversals of blue with cyan and displacement of yellow across the v* axis, as well as near-conflation of green and chartreuse. Note separation of orange and red in all solutions.

Of the seven English-speaking informants, three (IDs 3, 6, and 7) showed hue reversal, with one informant (7) showing displacement of the yellow MDS solution across the v* axis. Meanwhile, only one Somali informant’s solution (ID 3) showed no hue reversals. All other solutions showed multiple conflations, most noticeably disorder among blue, cyan, green, and chartreuse (all except IDs 1 and 3), and displacement of
purple or yellow across the v* axis (IDs 1, 2, 5, 6, 7, 9, and 10). Despite the frequency of these displacements, the general disorganization of Somali informants’ MDS solutions, and the high average $d$ values for fit to CIE UV, aggregation of Somali informants’ sorts gave a very well-ordered perceptual mapping with no reversals, two-dimensional stress = 0.0503, and $d$ = 0.0692. As in Experiments I and II, English informants’ perceptual hue maps uniformly showed greater-than-expected dispersion of orange and red solutions, likely due to the difference in lightness between the samples. See Figures 14 and 15 (pp. 47-48) for illustration of these effects in English and Somali MDS maps. Because nearly one-half of Somali speakers’ solutions showed only “fair” fit in two dimensions and most showed poor fit of Procrustes $d$, we sought to further characterize the number of dimensions appropriate for modeling MDS solutions.

In order to determine whether a two-dimensional solution was sufficient to model each participant’s sort we generated scree plots, which are used to visually assess the amount of variability explained by components of a solution (see Wickelmaier, 2003 and Appendix C for further discussion on usage and interpretation of scree plots). These plots were generated from one- to six-dimensional MDS solutions, and generally show distinct “elbows” at two dimensions for sorts with MDS stress > 0.1 in two dimensions, indicating that though these solutions are “fair” stress, two-dimensional solutions are likely adequate. However, two English informants’ sorts and two Somali informants’ sorts showed both MDS stress > 0.1 in two dimensions and a scree “elbow” at ≥ 3 dimensions, indicating that these sorts were significantly better modeled by higher-dimensional solutions. See Figure 16 for all scree plots for English and Somali informants. Spearman rank tests showed no significant correlations between stress and $d$.
for English speakers ($\rho = -0.2143; p = 0.6445$) or Somali speakers ($\rho = 0.5091; p = 0.1097$).

**Figure 16. Exp. 3 Scree Plots.** In 1 to 6 dimensions. A. Scree plots for English informants. Purple and green plots (subjects 4 & 6) show “fair” stress in 2D, but elbow at second dimension, indicating 2D solutions are likely adequate. Blue plot (subject 5) shows “poor” stress in 2D and elbow at third dimension, indicating 3D solution is necessary. B. Scree plots for Somali informants. All plots except purple and gray (subjects 7 & 9) show either “good” stress in 2D or scree elbow at second dimension; both of these require a 3D solution.

**Unique Hue Selections.** Somali participants showed greater variability in selections of all unique hues, with ranges of 7, 4, 7, and 4 samples each for green, blue, red, and yellow unique hue selections. For English-speaking informants, these ranges were 3, 3, 2, and 3, respectively. Median selections were as follows: green = 2.5G 5/10 for both English and Somali speakers; blue = 7.5B 5/10 for English speakers and 10B 5/10 for English speakers; red = 7.5R 5/12 for both groups; and yellow = 7.5Y 8/10 for both groups. Mann-Whitney U tests showed no significant differences between medians for any unique hue. See Figure 17 for English and Somali informants’ modal hue selections in UV space, and Figure 18 for histograms of selection dispersion. Spearman rank-order analyses of the order-ranked unique hue lists, as performed in Experiment I,
showed no significant correlations between any unique hues, indicating that red, yellow, green, and blue unique hue selections were independent with respect to one another.

![Figure 18. Exp. 3 Unique Hue Selection Frequency. Y-axis indicates number of subjects who selected each sample within a range. Darker colors indicate English speakers’ unique hue selections; lighter colors indicate Somali speakers’ unique hue selections. Not all colors within the ranges were selected.](image)

For English-speaking informants, circular correlation analyses of two-dimensional MDS solutions’ and unique hues’ residual hue angles showed no significant correlations between MDS solutions and unique hue selections. No such analyses were performed for Somali informants, because the general disorganization of Somalis’ two-dimensional MDS solutions precluded drawing any meaningful conclusions about hue relationships from MDS.

**Discussion**

English speakers’ MDS solutions in Experiment III were not significantly different from the MDS solutions from Experiment II. However, from personal observations, subjects seemed to understand the instructions of the binary sort protocol better than the free sort protocol, and it was performed much more quickly.
When Somali informants’ sorts were aggregated, the MDS solution was low-stress and fit quite well to fiducial points in UV space. However, individual Somalis’ sorts showed high stress, and two-dimensional solutions fit generally poorly to CIE UV fiducial points, as shown by $d$ and by visual inspection of the individual solution maps. It is likely that the population-level results of the aggregate MDS fail to convey important information about the individual solutions, though the aggregate hue map does suggest a certain degree of partitioning into “warm” and “cool” colors, as has been described in the literature (Maffi, 1990).

The disorganization of Somali two-dimensional MDS solutions may be due to poor understanding of the task, fatigue, or reference to color dimensions other than hue in sorting—most likely, lightness, color warmth versus coolness, or some combination of those two factors. In order to better understand the dimensionality of MDS solutions, scree plots were created. These plots generally suggested that two dimensions gave adequate fit for both English and Somali sorts; however, it should be noted that interpreting scree plots is as much an art as a science. For several English and Somali informants, however, the scree plot, two-dimensional stress, and poor fit to UV all implied the necessity of solutions with three dimensions. It was thought, therefore, that these informants may have used both hue and lightness as difference criteria, and that three-dimensional solutions would give improved hue relationships, as they would disentangle the effects of hue and lightness. However, projection of three-dimensional solutions onto UV for these subjects (not shown) did not lead to any noticeable improvement in hue relationships, indicating that poor fit in two dimensions was not the result of simple hue-lightness interactions. Several one-dimensional Somali solutions
showed surprisingly low stress, suggesting that these informants might use only one dimension in sorting such as lightness or a warm-cool scale; yet on inspection these solutions did not appear to vary on either of those dimensions. Further experiments are needed in order to determine the color dimensions salient for Somali observers in dissimilarity rating.

Six of ten Somali informants (IDs 2, 5, 6, 7, 9, and 10) showed displacement of purple and/or yellow across the v* axis in their MDS solutions; however, only three showed the lexical purple-yellow/orange conflation described by Brown et al. (submitted). The YELLOW-OR-ORANGE terms jaale and yalow were used by Somali informants 5 and 10, respectively, to name the purple sample, and informant 6 used the ORANGE loanword oranji to name the lavender sample. It is interesting to note that the solutions showing displacement of yellow or purple showed generally good fit to two dimensions, with average stress = 0.105; though as expected all showed poor fit to the fiducial points plotted in CIE UV, with average $d = 0.471$. Thus, even though these participants did not appear to sort along the hue dimensions traditionally used, two dimensions were largely sufficient to model their sorts.

Though dissimilarity sorting was simplified for this experiment by the utilization of the binary sort protocol, it is possible that some informants simply did not understand the task, or that they used different criteria of dissimilarity for different stimulus pairs (for example, lightness dissimilarity for one color pair, hue dissimilarity for another pair, and light-cool dissimilarity for another). We opted against use of a magnitude estimation protocol in this Experiment because we were concerned that we would not have been able to accurately communicate the concept through a translator; however, as we learned,
communicating about relative dissimilarities between stimulus pairs may have been even more difficult. A pilot study should be undertaken in order to compare the efficacy and translatability of a magnitude estimation protocol to the binary sort. Future experiments with Somali informants could use isoluminant stimuli in order negate or isolate the effects of non-hue dimensions on color scaling, though it is possible that color dissimilarity rating and ranking across large regions of hue space is inherently difficult and that these changes would affect the results very little.

As in Experiment II, red, yellow, green, and blue unique hue selections were independently distributed for both English and Somali informants, indicating that while unique hues varied, their variance was not systematic. There were no significant differences between median hue selections for the groups; however, Somali informants’ unique hue selections varied more than English speakers’ selections did. It is possible that this is due to definitional differences of color categories, or that some Somali speakers referred to color prototypes rather than true unique hues. One Somali informant (ID 7) was a user of the GRUE color-naming motif; yet, this informant was able to understand the instructions for the selection of unique cagaar (green) and buluug (blue). This subject selected 10GY as unique green and 10B as unique blue. While 10GY was the most yellow unique hue chosen, it was also the modal choice, selected by 4 out of 9 informants; 10B was similarly standard, in that though it was the second-reddest unique blue chosen, only 4 out of 9 informants chose samples that were more greenish. Linguistic relativity predicts that this subject would choose a very bluish unique green due to the extension of the cagaar category into blue, and a very reddish unique blue due to the reddish characteristic of the lone sample named buluug. Instead, this informant
actually chose the most yellowish unique green (though, again, it was not an unusual choice), and an only moderately reddish unique blue, well within the bounds of normality. Of course, because these are the results of only one GRUE user, few conclusions can be drawn about the relationship between color naming and unique hue selections; nonetheless, the results that we do find seem not to support a Whorfian linguistic relativity hypothesis.

Finally, English speakers’ MDS solutions showed no significant correlations with unique hue solutions, as in Experiment II. While two-dimensional solutions showed generally good measures of stress and Procrustes $d$, color reversal was not uncommon in maps, and a more precise method of measuring and analyzing perceptual hue differences would likely be necessary to sensitively measure correlation between MDS solutions and unique hues. To this end, we decided to use a magnitude estimation protocol and classical MDS in Experiment IV, as these methods allow for more nuanced estimation of color differences than dissimilarity sorting, and the magnitude estimation protocol is conceptually simpler and thus less likely to be subject to participant misunderstanding.
Experiment IV

Introduction and Study Design

Experiment IV was designed to compare classical MDS (hereafter referred to as cMDS), unique hues, and color naming within English-speaking subjects, with several chief aims. First, we wished to test the hypothesis that individual differences in the sizes and locations of lexical red, yellow, green, and blue categories are correlated with differences in individual unique hue and MDS settings. Because color-naming sessions were lengthy (145 samples would be tested, rather than the 23 samples tested in Experiment III) and we had limited time for data collection with undergraduate subjects, MDS data were collected via the faster magnitude estimation protocol. Unlike dissimilarity rankings, magnitude estimations allow for “ties” between stimulus pairs; thus, we had to use the metric cMDS algorithm to extract perceptual mappings of the same stimulus set as in Experiment III. This metric algorithm attempts to recreate the original inter-point distances of the dissimilarity matrix, whereas the non-metric algorithm attempts only to recreate a monotonic transformation of dissimilarity ranks. Because informants performing magnitude estimation had to maintain a stable criterion for judging color differences throughout the session, but could not change prior responses, we used a “semi-bounded” difference rating scale, in which the upper bound of difference ratings could be increased if the informant deemed the existing ceiling insufficient for a given comparison. Before analysis, magnitude estimations were
standardized to a 0-to-10 scale to ensure consistency. Because classical MDS solutions are shown to generally have greater stress (Agarwhal et al., 2007), we also compared cMDS results from this experiment to non-metric MDS results from Experiment III.

To compare color naming with the cMDS and unique hue data, we use k-means clustering to extract the categorical structure from our informants’ color naming responses. The numbers of color categories revealed by our analysis are validated with the gap statistic analysis (Tibshirani, Walther and Hastie, 2003) introduced to the analysis of color naming by Lindsey and Brown (2006; 2009), whose calculation by two methods we compare here.

We also introduce a novel technique which we call cluster stability analysis (CSA) as an additional way of defining the numbers of color categories shared among our informants. The results of CSA are then compared to those obtained by gap statistic analysis, both in this study and as employed by Lindsey and Brown (2014) in their study of American English color naming.

Subjects

In total, 50 subjects participated in this study (33 female). Participants were recruited from The Ohio State University College of Optometry student pool \((n = 12)\) and the Ohio State University Psychology Introductory Psychology (PSYC 1100) student pool \((n = 38)\). Two optometry students (both female; IDs 49 and 50) were excluded because they did not complete the experimental protocol. Thus, analyses described below are based on 48 subjects. Optometry students were compensated with ten-dollar gift cards from Target® or Kroger® as desired; PSYC 1100 students were given one hour of credit in the Research Experience Program (REP), which contributed to their class grade. All
participants were required to pass the HRR® pseudoisochromatic color vision test (Richmond Products, Albuquerque, NM). Criteria for inclusion were the same as for English-speaking subjects in Experiment III.

**Apparatus and Stimuli**

Optometry students were tested at the lab-based light box described in Experiment I, while PSYC 1100 students were tested at the mobile light box detailed in Experiment III. Both apparatuses produced very similar stimulus illuminations.

Stimuli for the magnitude estimation cMDS component of Experiment IV were identical to those described in Experiment III. Unique hue stimuli were the same as those described in Experiment II; however, rather than being placed ad hoc on the light box surface in a hue circle configuration, the color samples were displayed in slots arranged in a circle on a stiff gray square background. The background was made of Color-aid Gray 4.5 sheets bonded to 0.5-cm thick black foam, and was 54 by 54 cm. The circular arrangement of color samples was 45 cm in diameter, with roughly 1.5 cm between adjacent samples. This arrangement was more likely to cause participants to be biased by the spatial characteristics of a color wheel than the ad hoc ovoid arrangement used in previous experiments; however, due to time constraints imposed by the REP setup, it was impractical to set up the unique hue arrangement for every participant. See Figure 19 for a picture of the unique hue stimuli in the mobile light box.
Figure 19. Exp. IV Modified Unique Hue Arrangement. The author at the mobile light box used in Experiment IV, performing the unique hue selection task with modified stimulus presentation.

For the color naming experiment, a 145-sample subset of the 330-sample WCS Munsell color chart was used. Seven value levels (3-8) of Munsell hues were used, along with black, white, and three intermediate values of gray. Selections were drawn from the original WCS color chart in a checkerboard fashion, and used the same alpha-numeric coding system as the WCS chart, with the exception that the value axis was reversed such that “A” indicated black and “J’ indicated white. See Figure 20 for a diagram of the color stimulus set and coding scheme, and Appendix B for colorimetric data of the stimulus set.
Methods

In all experimental sessions, color difference ratings (magnitude estimations) were obtained before unique hue selections to avoid rating bias from exposure to a “color wheel” as described earlier.

Participants were told that they would be judging the differences between pairs of colors on a scale from one to ten, with one being most similar and ten most different; a participant was allowed to adjust the “ceiling” of her scale as she saw fit in the event that a pair of colors appeared more different than a pair already rated as a “ten.” Each participant first rated five practice cards in order to ensure she grasped the concept. This set of practice cards was drawn from the discarded subset of the original eleven-color stimuli from Experiment II; thus, there were no duplicated comparisons between the
practice and experimental cards. See Appendix A for the set of practice cards. Following this practice, the participant rated all 28 color pair stimuli in fixed pseudorandom order.

In color naming sessions, participants were first instructed that they would be providing one-word color names for 145 colored samples. They were further instructed to provide color names which would be applicable to any object, and which they would employ in everyday conversation to describe the color of an object to another person (After Berlin and Kay, 1969; Kay et al., 2009; Lindsey and Brown, 2014). If a participant provided more than one color word for a sample (such as “blue-green” or “lime green”) or used relational modifiers (such as “greenish”), she was asked to rename the sample with a monolexemic color word. Participants were required to name each sample.

Apart from the change in stimulus presentation detailed above, the protocol for unique hue selection was unchanged from Experiment II.

Analysis

Color naming was analyzed on a group level using several exploratory statistics, such as per-sample across-subject consensus, color name usage frequency, and comparisons of color name usage between subject groups. Next, cluster analysis (preceded by gap statistic analysis to determine the optimal number of clusters $k_{opt}$) was performed to characterize lexical color categories; then, motif analysis was used to determine whether distinct color-naming schemes existed. Finally, the stability of clusters across 100 k-means trials was determined.

*Gap Statistic Analysis for Optimal Numbers of Lexical Categories.* In order to optimize k-means and motif analyses, it was necessary to first determine the appropriate number of clusters $k_{opt}$, defined as the statistically significant number of color naming
categories in the ensemble data. As k-means analysis always returns a solution with the user-input $k$ clusters, so long as $k \leq n_s$, it cannot by itself empirically determine the optimal number of clusters $k_{opt}$. In order to find $k_{opt}$ we performed gap statistic analysis. This is a technique used in unsupervised machine learning to estimate the “true” number of clusters, and it is based on the idea that the sum of pairwise differences among within-cluster elements (dispersion) is less for the clustered data than for a clustered reference set based on the data. The reference set is guaranteed to have an optimal $k = 1$, as it is randomized with respect to the order statistics germane to clustering, particularly the relative positions of individual informants’ patterns. There typically are $n$ reference data sets, and the estimated reference set dispersion is expressed as the mean dispersion ± 1 SD obtained for the $n$ reference sets. The optimal value of $k$ for the clustered data is the smallest $k$ for which a $k+1$ clustering solution does not significantly decrease the difference in dispersion between k-means solutions for experimental data and reference sets (Tibshirani et al., 2002).

First, k-means analyses for $k$s ranging from 1:30 were applied to 140-element binary vector bitmaps encoding each chromatic color term used by each informant. Following Lindsey & Brown (2006; 2009; 2014), all color terms applied to any of the five achromatic samples in the set of test colors were excluded from the analysis. The $i$th element of each binary vector was assigned a value of 1 if the term was used to label the $i$th chromatic color sample, and zero otherwise. Separately, k-means analyses with the same range of $k$s were applied to 50 reference sets of 20 reference files each of randomized experimental data, to give mean dispersion values for these reference solutions over the range of $k$ values. In order to randomize the color-naming patterns for
reference sets while retaining pertinent information about the size and shape of patterns, the vectors were transformed into 20-by-7 matrices and randomly rotated about columns and flipped by columns and/or rows (at 50% probability for each flip), then transformed back into vectors (see Lindsey & Brown, 2006 for further details).

These k-means analyses gave four sets of outputs. For experimental data, 50 sets each of mean and mean-squared within-cluster dispersion were generated from k-means analyses with \( k = 1:30 \). Then, \( B = 50 \) sets of \( n = 20 \) reference dispersion distributions were generated for mean and mean-squared dispersion from k-means analyses with the same range of \( k \). For both measures of dispersion, a Mathematica script performed 500 random samplings with replacement, each consisting of one set of experimental dispersions and one distribution of reference dispersions, then converted them to base-10 logarithms. Next, the algorithm calculated population standard deviation and standard error \( (s_k) \) for reference solutions at each \( k \) value, as shown below:

\[
S_k = \sqrt{\frac{1}{1+B} \times sd(k)}
\]

Eqn. 3

Here, \( B \) refers to the number of reference distributions and \( sd(k) \) to the standard deviation of dispersion from k-means analyses of these reference distributions. Then the algorithm calculated the mean element-wise difference between log-experimental and log-reference data (the “gap” for a given \( k \)):

\[
g_k = (1/B) \sum_n \log(W_{kb}^*) - \log(W_k)
\]

Eqn. 4

Here, \( W_{kb}^* \) indicates dispersion within a set of k-means analyses for the reference sets, and \( W_k \) indicates mean or mean-squared dispersion for a k-means analysis of the
experimental set. Finally, the gap statistic $G_k$ was calculated from $g_k$, $g_{k+1}$, and $s_{k+1}$ as follows:

$$G_k = g_{k+1} - g_k - s_{k+1}$$  \hspace{1cm} Eqn. 5

For each of the 500 gap statistic analyses, the largest value of $k$ that returned a gap statistic $G_k > 0$ was considered the optimal cluster number for that analysis; the mode of these optimal $k$s across trials was then considered to be the “true optimal” number of color name categories, $k_{opt}$.

**Stability Analysis for Optimal Numbers of Lexical Categories.** Despite the efficacy of $k$-means analysis and the apparent constraints on English color naming, even analyses with an appropriate $k$ (as determined by gap statistic analysis) and many iterations ($n > 1,000$) do not converge on an invariant solution. This suggests that at least some of these $k$ categories are ambiguous, and that $k_{opt}$ as defined by the gap statistic may be too liberal an estimate of the number of categories shared among English informants in this study. Previous studies by Lindsey and Brown (2014) and others (Berlin and Kay, 1969; Boynton and Olson, 1987) have characterized color categories in English-speaking subjects; however, little attention has been dedicated to the stability of these categories, which may give insight into the evolution of English color categorization. To investigate this stability, across-subject consensus maps of category usage were correlated between $k$-means trials, and the percentage of optimal high correlations was determined at each $k$.

Consensus maps were created by re-mapping cluster assignment indices from $k$-means analyses back onto their color-term bitmaps. Then, the bitmaps for each index were summed across subjects and normalized to create cluster consensus maps. These consensus maps were then correlated across trials using MATLAB’s corrcoef function.
a given lexical category’s consensus map were invariant between multiple k-means analyses at a certain $k$, then each consensus map in a given trial would correlate highly ($R > 0.95$ for the purposes of this analysis) with 1 map in each other trial, giving $n_t-1$ high correlations for each map (where $n_t$ is the number of trials). It follows that if all possible combinations of correlations were performed between multiple trials, the number of high correlations would be calculated as below, where $t$ is the number of trials:

$$R_{>0.95} = k \times \sum_{i=1}^{t-1} i$$  \hspace{1cm} \text{Eqn. 6}$$

The total number of correlations is:

$$R_t = k^2 \times \sum_{i=1}^{t-1} i$$  \hspace{1cm} \text{Eqn. 7}$$

Thus, the fraction of correlations giving $R>0.95$ is $1/k$ for a comparison of stable maps. For $k$’s of 4 to 21, the fraction of all correlations with $R>0.95$ was calculated and divided by this optimal fraction to give the ratio of actual high correlations to optimal high correlations, $r_c$. These ratios and the corresponding $k$s were then plotted as the dependent and independent variables on a Cartesian coordinate system. The ratio $r_c$ was used as an adjunct method for determining the optimal number of clusters. By this method, the optimal $k$ gave the highest $r_c$. The fraction of optimal high correlations for each map, $r_{c,i}$, was also calculated by the number of high-correlation flags for that map divided by $n_{\text{trials}}-1$. Maps with comparatively low $r_{c,i}$ at a given $k$ represent ill-defined lexical color categories.

Gap Statistic Analysis for Number of Motifs. A k-means solution of $k_{opt}$ clusters for all informants’ color-naming patterns was partitioned into lists of cluster indices for
each informant, such that an informant’s unique color term bitmaps were then filled with indices from 1 to $k_{opt}$. For each informant these and the achromatic bitmaps were combined into a single 145-element pattern; then, the fraction of the 145 samples assigned to each index was calculated. These fractions were used as data sets for gap statistic analysis, and reference feature vectors were generated as described previously, with fraction index as the randomized variable. K-means analyses and gap statistic analyses were performed as described previously, with $k_{motifs} = 1:7$, to give $k_{optMotifs}$. Then, a second k-means analysis was performed using the empirical $k_{optMotifs}$, giving informant motif indices, and for each sample in each motif the modal category and its consensus were calculated. 

*MDS Analysis.* To analyze participants’ magnitude estimations, they were first normalized to a 10-point scale, then analyzed in MATLAB using the classical MDS function cmdscale. The resulting individual solutions were then scaled and rotated via Procrustes transformation for best fits in the UV plane of CIE LUV space, as described in Experiment I. As in previous experiments, Procrustes $d$ was used as the criterion for goodness of fit of a transformed solution to fiducial UV coordinates. A similar analysis was performed on the group average magnitude estimations.

*Analysis of Unique Hues.* As in previous experiments, the modal unique hue selections for red, yellow, green, and blue were calculated, and rank-coded lists of unique hue selections were compared using Spearman rank-order correlations and permutation tests in order to determine whether unique hue selections co-varied.

*Comparisons of Color Naming, cMDS, and Unique Hue Solutions.* Color naming, cMDS solutions, and unique hue selections were all compared to determine if any correlations were present. The first hypothesis to test was that subjects with more color
names would show better-organized cMDS hue maps. To test this, subjects’ numbers of raw and glossed color names were compared to Procrustes d with Spearman rank-order tests.

Then, to test the hypothesis that the positions of lexical color categories are correlated with MDS solutions, the centroids of color categories were compared to cMDS solutions in CIE UV space. Category centroids were determined as follows: for each of the red, yellow, blue, and green binary category maps, “hits” were replaced by their samples’ UV coordinates, and category centroids were calculated by averaging U and V within a category. Category centroid and cMDS solution data sets were converted into hue angles within CIE LUV space, and for each data set all individual color centroid and cMDS solutions’ hue angles were subtracted from the across-subject mean. This created residual hue angles, which were tested, in all permutations, for correlation across subjects using the circ_corrcc circular correlation test as in Experiment II. Category centroids and cMDS solutions were each compared to unique hue selections in the same fashion, using residual hue angles and the circ_corrcc to test all comparison permutations.

Results

Color Naming. While some subjects showed variations in naming of gray samples, using terms such as “charcoal” and “silver” (including one subject who did not use the term “gray” at all), no subjects used terms germane to hue for gray samples, and thus we will consider these alternative terms to all gloss to gray. With respect to the remaining 140 chromatic test samples, subjects varied in the number of color terms they employed, from a minimum of 8 to a maximum of 39. The modal number of terms used was 16 (n=8), and the mean was 20.6, with a standard deviation of 6.45. A Mann-
Whitney U test showed that the median number of color terms used by men (20) was not significantly different from the median number of terms used by women (21). See Figure 21 for distribution of number of color terms.

![Figure 21: Numbers of Raw Color Terms](image)

**Figure 21. Numbers of Raw Color Terms.** Number of informants using each total number of color terms (chromatic and achromatic). * Indicates 11, Berlin and Kay’s number of BCTs as measured in 1969; ‡ Indicates 20, the number of English color categories measured by Lindsey and Brown (2014).

Numerous samples were named with high consensus. In total, 48 samples (including all 7 samples where the consensus terms were those normally applied to achromatic colors) were named with ≥80% consensus, 13 samples were named with ≥90% consensus, and 4 samples were named with ≥95% consensus. Only 3 samples were named with 100% consensus. All participants used the eight English BCTs described by Berlin and Kay (red, orange, brown, yellow, green, blue, purple, and pink) to name at least one sample; no other color terms were used by all participants. See Figure 22 for a map of high-consensus samples, and Figure 23 for color term usage frequency.
Figure 22. High-Consensus Samples. Map of samples named with high consensus (≥80%) across subjects. Lighter-colored samples indicate higher consensus. Samples with asterisks were named at 100% consensus. Black and gray are false-colored to orange and cyan, respectively.

Figure 23. Raw Color Term Usage. Histogram of number of informants using the 46 most common color names (all used by four or more informants).
Gap statistic analysis with 500 randomly sampled trials showed the modal optimal number of clusters to be $k_{opt} = 16$ using sums-of-squared-means intra-cluster distances, and $k_{opt} = 17$ using sums-of-means intra-cluster distances. Thus, secondary measures of optimal clustering were used to help determine the correct number of color categories. See Figure 24 for histograms of $k_{opt}$ distribution from gap statistic analysis.

![Figure 24. Optimal Number of Chromatic Clusters](image)

Cluster stability analysis was performed for $4 \leq k \leq 21$. Figure 25 shows the ratio of actual to optimal high-stability clusters ($r_c$) plotted as a function of $k$. The function fits a roughly parabolic curve, with $r_c \geq 95\%$ for $15 \leq k \leq 10$. The ratio $r_c$ is equal to 92.7\% at $k = 16$ and 86.9\% at $k = 17$, indicating decreasing stability of clusters beyond $k = 15$. In order to account for as much variation as possible in color naming while avoiding overfitting, $k = 15$ (the highest $k$ which showed $r_c > 95\%$) and $k = 16$ (the lowest $k_{opt}$ given by gap statistic analysis) were examined.
Figure 25. High-Consensus Clusters by $k$. Percentage of optimal number of high-correlation clusters, $r_{c,i}$ plotted as a function of number of clusters. $K=10:15$ shows $r_{c,i} \geq 95\%$ (red line). *Indicates $k=16$, the $k_{opt}$ derived by gap statistic analysis with a sums-of-means-squared metric. ‡Indicates $k=17$, $k_{opt}$ derived by gap statistic analysis with a sums-of-means metric, and calculated by Lindsey and Brown (2014).

When $k = 15$, the cluster for medium-value yellow-green samples (henceforth, MUSTARD) is the only cluster which shows high volatility ($r_{c,i} = 0.484$). Clusters for RED, PEACH, ORANGE, BROWN, YELLOW, BEIGE, GREEN, AQUA, TEAL, BLUE, LAVENDER, MAGENTA, PURPLE, and PINK were all extremely stable, with $r_{c,i} > 0.95$. In Figure 23 (p. 69) it can be seen that the raw term “mustard” is used by slightly less than half of informants ($n = 19$). “Gold” is also used by 19 informants, and depending on the subject, spans an area between light yellow-orange and medium yellow. Different clustering trials could, therefore, assign samples named gold to either the YELLOW or the MUSTARD cluster, which we would predict would lower the $r_{c,i}$ of both. In fact, the YELLOW cluster shows the lowest $r_{c,i}$ of any high-consensus cluster in
this analysis ($r_{c,i} = 0.9798$); additionally, different trials’ solutions for the MUSTARD cluster show varying degrees of infiltration into the light yellow-orange region of the map. See Figure 26 for an illustration of this effect, as well as of the multiple loci of “gold.”

![Figure 26](image)

**Figure 26. MUSTARD and “Gold” Variability.** Inset panels depict I) Raw color naming; II) Glossed color naming; III) Across-subject consensus map for usage of mustard category. Cyan arrows indicate subjects’ foci for usage of “gold”. A. “Gold” employed for medium-dark yellow to yellow-green samples. Consensus map for mustard shows low consensus for samples G09 and H08. B. “Gold” employed for medium-light yellow-orange samples. Consensus map for MUSTARD shows high consensus for samples G09 and H08. “Gold” can be assigned to either the YELLOW or MUSTARD cluster in different k-means trials. Both glossed maps based upon assignment of “gold” to MYSTARD, as in Trial 1.
For $k = 16$, clusters for dark yellow-green (henceforth OLIVE) ($r_{c,i} = 0.758$),
MUSTARD ($r_{c,i} = 0.484$), PEACH ($r_{c,i} = 0.879$), and light and dark peach ($r_{c,i} = 0.121$ for each) were volatile, with light and dark peach revealed in some solutions but not others. Meanwhile, clusters corresponding to RED, ORANGE, BROWN, YELLOW, BEIGE, GREEN, AQUA, TEAL, BLUE, LAVENDER, MAGENTA, PURPLE, and PINK all showed $r_{c,i} > 0.95$. While the olive cluster is relatively robust, the transient presence of light and dark peach indicate that $k = 16$ likely represents an over-clustering of our subjects’ color naming data. Thus, further analysis was performed for $k = 15$.

Motif analysis with $k = 15$ showed four color-categorization motifs. Motif 4 was the most traditional, with non-basic color categories modal for only 14 samples, and AQUA and TEAL categories both very lightness-restricted. LAVENDER was the most extensive non-basic category, used as the mode for four adjacent samples. Motifs 1 and 2 were both characterized by greater usage of non-basic color categories, with Motif 1 showing greater usage of MUSTARD and Motif 2 showing greater usage of LAVENDER. Motif 3 showed an expanded TEAL category and very limited usage of LAVENDER, such that it was not modal for any samples. All motifs showed an Abney effect-like shift in the category boundary of BLUE with increasing lightness, as evidenced by the greenest sample glossed to BLUE at each Munsell value level. See Figure 27 for visualization of modal glossed color terms for each motif.
Figure 27. Modal Sample Glosses for Color-Naming Motifs. A. Motif 1, showing expanded MUSTARD and PEACH, contracted YELLOW and ORANGE, and balanced TEAL and AQUA. B. Motif 2, showing more conservative usage of PEACH and MUSTARD, and expanded LAVENDER. C. Motif 3, showing expanded TEAL which largely overtakes AQUA, no LAVENDER, and minimal usage of MUSTARD and MAGENTA. D. Motif 4, showing conservative usage of all secondary color terms, with TEAL and AQUA restricted to very dark and light samples. Note that all motifs show a shift of the left category boundary for BLUE, with greener samples named BLUE at higher values.

MDS and Unique Hue Selections. Unique hue selections were distributed similarly to previous experiments, with the greatest variability seen in selections of unique blue and the least variability in selections of unique red. See Figure 28 for visualization of modal unique hues in CIE UV space and a histogram of unique hue selection frequency.
Figure 28. Exp. IV Unique Hue Selections. A. Modal unique hue selections in UV space. Circles: modal unique hue selections for red, yellow, green, and blue. Lines: connect u*v* coordinates of MDS stimuli. B. Histogram of unique hue selection frequency.

Classical MDS produced hue maps generally similar to those generated by non-metric MDS in Experiment III. However, some participants’ solutions showed focal displacement of colors, particularly yellow and purple, and others appeared generally disorganized. See Figure 29 for several representative hue maps. Despite the number of poor fits, the mean goodness-of-fit $d$ to CIE UV was not significantly different between the non-metric MDS protocol used in Experiment III ($\mu d = 0.134$) and the cMDS protocol used in Experiment IV ($\mu d = 0.187$), with a Mann-Whitney $U$ value of 0.221. Stress values were not compared between the two experiments, because stress is calculated differently for classical MDS than non-metric. Also worthy of note is the general displacement of red and orange away from one another in both individual and aggregate hue maps, congruent with the effect seen in previous MDS experiments with English-speaking informants. See Figure 30 for a comparison of aggregate hue maps from Experiments III and IV.
Figure 29. Exp. IV MDS Solutions. Representative English informants’ MDS solutions for Experiment IV. Triangles connected by lines represent stimulus fiducial coordinates; circles represent MDS solutions. A. Best-ordered solution (stress=0.238; \(d=0.066\)). B. Typical solution (stress=0.224, \(d=0.182\)) showing near-conflation of green-chartreuse and blue-cyan pairs, as well as stretching across the v* axis. C. Poorly ordered solution (stress=0.548, \(d=0.198\)) showing no reversals but near-conflation of red, orange, and yellow and a generally unusual shape. D. Worst-ordered solution (stress=0.300; \(d=0.704\)) showing general disorganization among chartreuse-to-blue colors, and displacement of purple across the v* axis.
Figure 30. Exp. IV vs. Exp. III Aggregate MDS Solutions. A. Aggregate non-metric MDS solution from Experiment III ($d=0.052$). B. Aggregate classical MDS solution from Experiment IV ($d=0.070$). These aggregate solutions show extremely similar shape, notably being stretched along the $v^*$ axis and showing greater-than-expected separation between red and orange.

While many participants’ sorts fit well to UV space, stress calculations showed poor fit in two dimensions for 38 subjects (with the remaining 10 showing only fair fit), and 9 of 48 subjects showed poor Procrustes goodness-of-fit $d (> 0.2)$. These trends suggested that classical MDS may not give meaningful information about hue relationships within this protocol. In particular, if these two effects were correlated such that worse fit of the two-dimensional solution to empirical ratings implied poorer goodness of fit to fiducial points, this would mean that only those participants with “standard” two-dimensional cMDS solutions (low $d$) used hue dimensions meaningfully, and that no meaningful inferences about hue relationships could be drawn for participants with “unusual” two-dimensional cMDS solutions (high $d$). In order to test this, the maximum relative errors of cMDS solutions (in two dimensions) were compared to Procrustes goodness-of-fit $d$ (also in two dimensions) using a Spearman rank-correlation test, which failed to reject the null hypothesis that the eigenvalues did not correlate.
significantly with Procrustes $d$ ($\rho=0.244; p=0.094$). This non-correlation is theoretically interesting because it shows that even when two-dimensional solutions are insufficient to model sorts, these sorts can still fit well to fiducial hue dimensions; conversely, solution projections which fit poorly to hue dimensions do not necessarily imply poorer fit of a two-dimensional solution.

Spearman rank-order tests showed non-significant negative correlations between Procrustes $d$ and the numbers of raw ($\rho = -0.279, p = 0.055$) and glossed ($\rho = -0.235, p = 0.108$) color terms employed by subjects.

When residual hue angles of lexical red, yellow, green, and blue category centroids were compared to residual hue angles of the corresponding colors’ cMDS solutions via circular correlation, only one correlation was within initial significance parameters for two-tailed testing (centroid blue versus MDS red: $\rho = 0.3232; p = 0.0106$); however, after correcting significance for multiple comparisons ($p$ required for significance: $0.025 / 32$ comparisons = 0.0008), this correlation was not considered significant. Similarly, residual hue angle comparisons of centroids against unique hues and cMDS solutions against unique hues did not show any significant correlations.

**Discussion**

Color naming in the present study, which employed a palette of 145 test colors, showed similar results to previous experiments with English speakers performed by Lindsey and Brown (2014; 330 color samples), with comparable modal and mean numbers of unique terms used. Also congruent with Lindsey & Brown (2014) was the consensus of usage of the 11 English BCTs across participants: only these terms were used by all participants, and only these terms were used with >80% consensus for any
given samples. No non-basic color term was used with >50% consensus for any of the samples. Contrary to numerous previous studies by numerous authors, including Lindsey and Brown (2014), Rich (1977), and Swaringen et al. (1978), we found no significant differences in the number of raw or glossed color terms employed by men and women; in previous studies, we also showed no significant difference in perceptual hue scaling via MDS between men and women, either in two-dimensional stress or in Procrustes goodness-of-fit d to fiducial CIE UV space.

The gap statistic, developed by Tibshirani et al. (2002), is an intuitive and powerful means of determining the optimal clustering of data sets; however, it is unclear whether using mean or mean-squared within-cluster differences is ideal. A previous study of color naming by Lindsey and Brown used mean within-cluster distances, which gave a solution of \( k_{opt} = 17 \). This study used both mean and mean-squared differences, which through 500 trials each gave modal \( k_{opt} = 17 \) and \( k_{opt} = 16 \), respectively; however, \( k_{opt} = 14-18 \) was frequent in both analyses. In order to determine what \( k_{opt} \) best described the clustering data, cluster stability across trials was used as a secondary measure of clustering optimality. Clustering solutions were very stable across trials for \( k = 10-15 \), but less stable at \( k = 16 \) \((r_c = 92.7\%\)\) and \( k = 17 \) \((r_c = 86.9\%\)\). At \( k = 16 \), cluster solutions for the BCTs plus beige, aqua, teal, magenta, and lavender were extremely stable, while cluster solutions for hunter, olive, and peach were volatile, with light and dark peach categories occasionally emerging. Apart from peach, all of these categories are very limited in scope, with only one to four samples modally binned to them. This suggests that categorical terms for light reds, oranges, and yellows and for dark yellows and very dark greens are used with poor consensus across informants, and that novel categories
within these regions of color space do not encroach highly onto BCT regions, but rather subdivide regions of color space occupied by secondary color terms.

As in Lindsey and Brown (2014), the BLUE category boundary in the present results is shifted in a greenish direction as Munsell value increases (from row C to row I). This effect is robust across informants, and appears despite the fact that our illuminant, rated at 5900 K, is significantly yellower than Standard Illuminant C (rated at 6770 K), for which the Munsell color system’s appearance has been calculated. If illuminant were to cause shifts in color appearance, it would be expected that the shift would be in the opposite direction—the greater proportion of long-wavelength light in the illuminant would cause light stimuli to appear more yellowish, rather than more bluish.

The results of analysis of stress and goodness-of-fit $d$ showed that cMDS solutions can be projected into CIE UV space and appear to retain much of the pertinent information about hue. This is significant because, unlike most previous MDS experiments, none of our protocols used isoluminant stimuli. Our results show that while color dimensions beyond hue do influence scaling for English speakers, these effects are largely independent from that of hue; and that color conflation in scaling may be mediated partially by hue confusion, rather than entirely by color dimensions beyond hue. Note however that stretching of the $v^*$ axis and wide separation of red and orange are still apparent in the results of this experiment, and thus the effects of luminance likely cannot be completely dissociated from those of hue. Additionally, as in previous experiments, numerous subjects’ sorts showed color reversals, which is troublesome in evaluation of perceptual hue spaces.
While color naming and MDS are both generally qualitative procedures, several statistics of each provide more quantitative measures. These measures, which are cluster centroid UV coordinates, number of color names, and number of color categories for color naming, and goodness-of-fit $d$ and Procrustes-scaled UV coordinates for classical MDS, provide opportunities to explore possible links between color naming and color perception. Linguistic relativity dictates that an individual’s color naming ought to non-trivially constrain perception of color differences. It is a reasonable hypothesis that, if linguistic relativity holds true, the number of color names or color categories an individual employs should predict performance on a color difference rating task. Kay and Kempton (1984) showed that, to a degree, category boundaries are salient when making judgments of color difference using an odd-one-out paradigm. The universalist explanation of this is that local maxima in JNDs (just-noticeable differences) are used as natural, universal points for lexical partitioning of color space; thus, color naming categories and color difference rating are both reflective of low-level processes (Witzel and Gegenfurtner, 2013). However, this does not account for one of the secondary results of Kay and Kempton’s study. When an indigenous Mexican language which has no blue-green division, Tarahumara, was tested, no boundary effects were seen in color difference rating. A strong Whorfian hypothesis for this effect is that lexical partitioning affects the perception of colors within and across category boundaries. A weaker version, which Kay and Kempton suggested, is that the universalist hypothesis largely explains the etiology of color categorization. However, due to the inherent difficulty of quantifying differences between colors, observers consciously or subconsciously refer to a “color wheel” of their own lexical categories when making these comparisons.
Our results showed no significant correlation between the number of color terms or color categories used by a participant and the goodness of that participant’s cMDS solution fit to UV space. The null hypotheses that there are no relationships between internal color space and number of lexical color terms or categories, therefore, cannot be rejected.

Kay and Kempton’s experiments showed that lexical category boundaries were employed in estimations of color difference across small regions in color space—specifically, on or near the boundary between more bluish and more greenish blue-green colors. One aim of Experiment IV was to determine if such effects were present at larger perceptual scales—that is, were lexical color category locations correlated with color scaling via cMDS? Shepard and Cooper (1992) showed that color-normal subjects showed nearly identical MDS solutions for sorts of color name pairs as for sorts of pairs of actual colors. This indicates that these English-speaking subjects had sufficient knowledge of color relationships to accurately rate dissimilarity, even when color percepts were not available for reference. We hypothesized that this knowledge was related to lexical color categorization, and thus that the centroids of lexical color categories would be correlated with the positions of cMDS solutions. For example, a participant with poorly separated red and yellow lexical categories might be expected to scale the colors close together in classical MDS; conversely, one with highly separated red and yellow categories might scale them further apart. A special case is that of blue-green discrimination. Not all participants employed categorical color terms for aqua or teal, and on gross inspection blue, teal, and green showed variability in ordering and spacing in cMDS solutions. It is possible that the lexical delineation of teal and/or aqua
categories might be correlated with greater perceptual dissimilarity between blue, green, and blue-green samples.

To determine the relationships between lexical category centroids and cMDS solutions, we performed circular correlation of residual centroid and cMDS solution hue angles. These correlations of red, yellow, green, and blue category centroids and cMDS coordinates showed no significant relationships, indicating that the color category centroids do not co-vary with MDS solutions. This suggests that, at the large scale on which the color dissimilarity rating was performed, participants did not use their lexical category centroids as landmarks to aid in color scaling. This is not an entirely surprising result; Regier et al. have shown that color prototypes, which are likely more salient in “anchoring” perceptual color space, are not the same as lexical category centroids, nor do they even necessarily co-vary; for example, in languages utilizing GRUE-type color naming schemes for blue and green, the prototype of GRUE is a focal blue or green, rather than blue-green. Additionally, color prototypes show very low variation, even across languages (Regier, Kay, and Cook, 2005; Rosch, 1975).

In examining the relationship between TEAL/AQUA usage and cMDS solutions, we searched for order reversals of blue, cyan, and green cMDS solutions in angular UV space, then compared the number of such reversals between TEAL/AQUA users and non-users. We also compared the mean distances between blue, cyan, and green solutions for users and non-users. Though most subjects used aqua and/or teal categories extensively and across nearly all lightness levels, subjects 8, 10, 11, 13, 23, 33, 39, and 41 did not employ either category between blue and green within the E and F levels of the WCS chart, which were similar luminance to the stimuli used for color difference
rating. (Subjects 8 and 11 did both use aqua within these levels, but only for one sample within the confines of their blue categories.) Participants 13, 33, and 41 all showed some degree of blue-cyan-green confusion in their cMDS plots. Thus, 3/8 of mid-lightness TEAL/AQUA non-users showed this confusion, as compared to 1/5 of users. While this statistic appeared promising, when t-statistics were calculated for differences in mean distances between blue, cyan, and green cMDS solutions, no significant differences were found between users and non-users. This indicates that, though subjects varied in their lexical partitioning of blue-to-green colors, these differences were not correlated with perceptual distance between blue, cyan, and green.

Another prediction of the linguistic relativity hypothesis is that unique hues selection will be biased in some way by an observer’s color language, particularly by his lexical partitioning of color space. Webster et al. have shown that unique hue selections (particularly for blue, green, and yellow) vary significantly among normal observers, and that the effect is not mediated by interobserver variations in sensitivity along the cone-opponent axes (Webster et al., 2009). However, circular correlations of our color naming and unique hue results showed no significant relationships between category centroids and unique hues, so we fail to reject the null hypothesis that unique hue and color category centroid locations are independently distributed.

Finally, as in Experiments II and III, we compared residual hue angles of MDS solutions and unique hue selections, in order to determine if they were co-distributed. As in the previous experiments, no significant relationships between MDS solution angles and unique hue angles in UV space were found.
General Discussion

Overview

The experiments detailed in the preceding sections were performed in order to investigate the relationships between color naming, MDS color scaling, and unique hue selections. Additionally, we hoped to better characterize color scaling in English versus Somali-speaking informants, and to investigate the number of consensus categories in English color naming. The major findings of these experiments follow.

First, our non-metric and classical metric MDS studies with English-speaking informants in Experiments I-IV replicate many of the features of Shepard and Cooper’s (1992) non-metric MDS experiment using heteroluminant stimuli, after which we modeled our paradigm. However, unlike Shepard and Cooper, we found what appears to be a significant contribution of luminance differences to dissimilarity judgments, as reflected by two-dimensional MDS solutions.

Second, comparison of MDS solutions with unique hue selections for English speakers in Experiments II-IV suggest no relationship between perception of color differences and the locations of unique hues in color space. This non-correlation was also true for Somali speakers’ MDS solutions and unique hues.

Third, individual Somali informants in Experiment III appear to make color dissimilarity judgments that are categorically different from those of English informants. Many Somalis’ MDS maps do not appear to vary predictably by CIE UV ordering, and
hue reversals are common. In particular, many Somali informants’ MDS maps show purple-yellow conflation, which we expected to be present based upon previous color naming data which showed that many Somalis use the word glossing to YELLOW-OR-ORANGE to name purple and lavender samples. However, most informants who exhibited this conflation in MDS lexically separated yellow in purple. Additionally, the differences between Somali and English color dissimilarity sorting are much less evident in MDS solutions based upon aggregated sorts.

Fourth, when we analyzed our English color naming data in Experiment IV with gap statistic and cluster stability analysis, they showed generally concordant results which well replicated those obtained by Lindsey and Brown (2014).

Fifth, comparison of red, green, blue, and yellow lexical category centroids with classical MDS solutions and unique hues in Experiment IV did not show any significant correlations. This suggests that lexical category locations do not govern perception of color differences or unique hues, as suggested by the linguistic relativity hypothesis.

Finally, English speakers’ color names for blue-to-green Munsell samples showed a pronounced Abney-like effect with decreasing sample saturation. This is surprising, given the Munsell system’s claim to correct for the effects of colorimetric purity on hue appearance.

Given these findings, this Discussion will focus on their significance, and the ways in which future studies of color naming and perception could correct for weaknesses within our protocols. Finally, we will summarize the theoretical implications of our findings with respect to the linguistic universalist and relativist hypotheses.
Luminance Differences Affect Two-Dimensional MDS

Our pilot study showed similar results to Shepard and Cooper’s 1992 non-metric MDS study with heteroluminant stimuli, with seven of our nine subjects’ MDS solutions showing “good” stress (< 0.1) and the remaining two showing “fair” stress (< 0.2). However, our MDS solutions showed variability not reported by Shepard and Cooper. A number of MDS maps showed focal reversal of colors, particularly for blue, green, violet, purple, and magenta samples. Additionally, several trends in the results indicated that two-dimensional solutions may have been sub-optimal. Orange and red-orange were widely separated in most solutions, and there appeared to be stretching along the $v^*$ axis between blue-purple and yellow-orange colors. Comparison of Procrustes-transformed MDS maps to stimulus fiducial points in UV space showed generally “fair” fits, with only one-half showing “good” fits to UV. Thus, even though two-dimensional solutions were internally consistent, they did not necessarily project well into the UV perceptual hue space. We hypothesized that this could be due to a confounding effect of lightness differences between samples, which subjects may have conflated with hue difference.

When we examined our stimuli, we discovered that stimulus $L^*$ co-varied with $v^*$ in a nearly monotonic fashion, as seen in Figure 31. Because lightness varied almost continuously along the $v^*$ axis, we determined that in a two-dimensional MDS analysis variation due to lightness could be projected onto the $v^*$ axis and cause it to stretch without greatly degrading the solution’s two-dimensional stress. Because we wished to continue using heteroluminant stimuli in our MDS experiments in order to preserve color appearance, we decided to modify our color sort stimuli in Experiment II.
Figure 31. Experiment I MDS Stimuli L* and v* Coordinates. Asterisks represent samples whose L* values deviated from a strictly monotonic function of v*. The slope of the regression line is 0.317.

Thus, Experiment II used a white instead of gray background for the stimulus cards in the hopes of diminishing relative lightness effects, while retaining the color appearance of necessarily light colors such as yellow and orange. It also introduced a unique hue protocol, which we hoped to compare to MDS results to determine if unique hues and MDS maps co-varied.

To measure the effectiveness of the white stimulus cards at reducing v* axis stretching, we compared the average distances between orange and purple MDS solutions in Experiments I and Experiment II, with the alternate hypothesis that these distances would be greater in Experiment I. A t-test showed this to be the case, as the mean distance between purple and yellow in Experiment I ($\mu = 152.3$) was significantly higher than in Experiment II ($\mu = 132.0; p = 0.012$). Separation of orange and red was our other metric for lightness effects, and so we compared this across experiments as well. Red and
orange showed slightly greater separation in Experiment I ($\mu = 65.5$) than in Experiment II ($\mu = 58.2$), but the effect was not significant ($p = 0.387$). Thus, we concluded that the altered stimulus presentation appeared to reduce the $v^*$ axis stretching effect, but did not affect the judgment of dissimilarity between orange and red. Future MDS experiments could correct for these luminance effects by decreasing overall luminance contrast, or they could utilize stimuli of a uniform range of luminance and saturation values in order to separate their effects.

**MDS is Uncorrelated with Unique Hues**

Unique hue selections showed no correlations amongst themselves when analyzed by rank-order, nor with MDS solutions when analyzed by hue angle. While the stimuli for the two protocols were not equivalent (the Munsell stimuli for unique hue selection were, with the exception of blue, slightly less saturated than those for color dissimilarity sorting), one would nonetheless expect hue selections and MDS to co-vary if the unique hues were perceptually salient and our MDS protocol were sufficiently sensitive to capture variations in perceptual hue differences. However, as previously mentioned, MDS solutions in Experiment II often showed large-scale displacement of colors, including reversals of colors in hue space. While this may have been due to perceptual differences, it was more likely due to task difficulty. Thus, our MDS protocol in this Experiment may have obscured any perceptual differences that may have been present. In proceeding experiments, we hoped to improve the sensitivity of the MDS protocol in order to better capture individual differences in perception of hue dissimilarity.
Somali versus English Dissimilarity Sorts

In Experiment III, we sought to compare Somali to English-speaking informants’ perceptual hue maps obtained with MDS. Though we were only able to collect data from a limited number of Somali informants, we observed several interesting effects in their color dissimilarity ranking habits.

Somali MDS maps, unlike those obtained from our English informants, generally showed numerous deviations from fiducial CIE UV hue ordering; most notably, six of ten maps featured displacement of purple and/or yellow across the v* axis. This effect is consistent with lexical categorization of purple and yellow (Lindsey and Brown, submitted), as well as non-lexical behavior in free-sort tasks (Lindsey and Brown, 2013). However, four out of the six informants who showed this behavior used separate terms for purple/lavender and yellow (though one, ID 6, did use the “orange” loanword *oranji* to name the lavender sample). This indicates that the displacement of yellow and purple in MDS is likely not due entirely to a Whorfian effect. It may be that the two colors have a common association in Somali culture even though these subjects used different names for the two colors, or it could be an effect of the general disorganization of Somali perceptual hue maps derived from our MDS protocol.

If the conflation of purple with yellow were the only way in which Somali subjects’ sorts deviated significantly from objective hue relationships, we would expect that sorts which omit purple would show far better hue ordering and lower Procrustes goodness-of-fit d to UV space. To test this hypothesis, all color-pair comparisons involving purple were removed from sorts for the six Somali informants with unusual purple-yellow solutions, and two-dimensional MDS and Procrustes comparison to the
remaining seven fiducial points were calculated. Student’s t-tests showed that mean goodness-of-fit $d$ to fiducial points did not improve significantly, with mean $d = 0.471$ derived from solutions of the full set of 28 stimulus pairs and mean $d = 0.316$ from the reduced set of 21 stimulus pairs ($p = 0.270$). However, one subject (ID 5) did display a marked improvement in goodness-of-fit, showing $d = 0.428$ with the full set of comparisons and $d = 0.058$ with the reduced set. See Figure 32 for an illustration of these two maps. While other analyses may be needed to characterize the effect of unusual purple-scaling behavior on MDS solutions, this analysis shows that for most Somali informants, the effect of purple is not the only factor contributing to poor fit to CIE UV.

Figure 32. Somali S5 cMDS Solutions with Full and Reduced Stimulus Sets. A. cMDS map generated with full set of 28 comparisons of 8 colors ($d=0.428$). Note displacement of purple across the $v^*$ axis. B. cMDS map generated with reduced set of 21 comparisons of 7 colors ($d=0.058$). Note the generally good fit to fiducial coordinates, though green-chartreuse, blue-cyan, and red-orange solution pairs are all nearly conflated. This suggests the subject’s hue dissimilarity judgments were largely informed by standard hue relationships, though with disproportionately low perceived dissimilarity for adjacent samples, and purple perceived as very similar to yellow.
As discussed previously, MDS solutions showing purple-yellow displacement did not necessarily exhibit high stress in two dimensions; thus, it is possible that these participants sorted based upon one or two internally consistent color dimensions that are not related to known hue dimensions. Maffi (1990) suggested that warm-versus-cool is a very salient hue dimension to Somali speakers, as it is to people from many cultures in “early” stages of color lexicon evolution (Berlin and Kay, 1969). Thus, we calculated one-dimensional solutions for the six Somali subjects whose MDS solutions showed purple-yellow displacement, in order to determine if these subjects performed their sorts along a warm-cool dimension. Two of these six sorts (IDs 2 and 5) showed a distinctive pattern when MDS was performed in one dimension, with blue, green, cyan, and chartreuse clustered together into a “cool” grouping; red, orange, and yellow in a “warm” grouping; and purple in between. However, one-dimensional MDS solutions with this arrangement did not show significantly lower stress than those with other arrangements which did not conform to any known dimension of color. Thus, we find it unlikely that these informants were reliably using a warm-cool dimension as the criterion for their sorts. Further research on the salient color dimensions in Somali language and daily visual experience is needed before any definite conclusions can be drawn about dimensionality of MDS solutions.

The Number and Nature of English Color Naming Categories

In Experiment IV we used a color naming task to better characterize the nature of English color categories, with attention focused to the number of shared categories among informants and their relative stability. We also used color naming data, along with a classical MDS paradigm and unique hue selections, to investigate the relationships
between color categories, judgments of color differences based on MDS, and unique hues. While we found little evidence for any linguistic relativity effects between color naming and MDS or unique hues, our color-naming results corroborated those found by Lindsey and Brown (2014), showing 15 shared chromatic categories and a strong Abney-like effect.

Gap statistic analysis of color naming showed the optimal number of clusters to be 17 and 16 when calculated with sums of means and sums of means-squared intra-cluster dispersion, respectively, so we turned to cluster stability analysis with 100 k-means analyses at a range of k values to better characterize the number of shared color categories. This analysis showed that the percentage of actual to optimal high-consensus clusters, $r_c$, was > 95% for $k = 10:15$ chromatic clusters, indicating that gap statistic analysis was likely too liberal. We thus determined $k = 15$ to be the number of high consensus categories shared by informants. Of these categories, all except MUSTARD showed very high consensus across trials ($r_{c,i} > 95\%$). Notably, MUSTARD was the least-used category, used by 35 of the 48 participants. Additionally, one of the terms which frequently glossed to it, “gold,” showed a bimodal distribution, being used in the region “typical” for MUSTARD focused at WCS position F09, and in the light yellow-orange region focused on WCS position H07. Thus, “gold,” similar to “tan” as described by Lindsey and Brown (2014), is a color term that, though used by numerous subjects ($n = 19$), shows sub-optimal mutual information. Future experiments using the Color Communication Game (Lindsey and Brown, 2015) should be utilized to directly test the group mutual information conveyed by “gold.” See Figure 33 for a histogram of glossed color term usage frequency and a bubble chart of “gold” usage locations and frequency. It
should be noted that the usage of 15 chromatic clusters (being the largest $k$ giving $r_c > 95\%$) was chosen partially because we wished to account for as much variation as possible; a more conservative way to determine $k_{opt}$ via CSA would be to fit a second-order polynomial to the function of $r_c$ and $k$ and use the polynomial maximum as $k_{opt}$.

Future studies will explore this possibility.

![Figure 33. Color Category and “Gold” Usage.](image)

**Figure 33. Color Category and “Gold” Usage.** A. Histogram of frequency of usage of the 15 color categories revealed by cluster stability analysis. *Indicates that MUSTARD was the category used by the fewest informants (n=35). Note that only BCT categories were used by all informants. B. Bubble chart of usage of the raw color term “gold.” Circle centers indicate WCS samples named with “gold” by any subjects, and circle diameters indicate number of subjects who used “gold” to name given samples. Red arrows indicate the two disparate foci, at F09 and H07. For some informants, “gold” indicates a specific hue of light yellow-orange (focus H07), while for others, it refers to a general category of mid-value yellow-greens (focus F09).

**English Color Naming Is Not Correlated with MDS or Unique Hue Selection**

The addition of the color naming task to our model allowed us to directly test for Whorfian effects on cMDS and unique hue selections. However, numbers of color names and categories did not correlate significantly with cMDS maps’ goodness-of-fit to UV color space; nor did lexical color category centroids for BLUE, GREEN, YELLOW, and RED correlate with either cMDS solutions or unique hue selections. A total of eight subjects did not use any terms which glossed to TEAL or AQUA to name colors between blue and green of WCS value E or F (the approximate values of our cMDS stimuli). We hypothesized a Whorfian effect in which usage of either of these intermediary categories
would be correlated with greater perceptual spacing of our blue, cyan, and green cMDS stimuli; however, t-tests showed no differences between aqua/teal users and non-users in mean Euclidean distance between blue and cyan, cyan and green, or blue and green.

While it is tempting to conclude from these results that no Whorfian effects between color naming, color difference perception, and unique hues exist, it is possible that such effects could be present, but that our protocols did not measure perceptual differences precisely enough to detect them (Type II error). First, with respect to comparisons of color naming and unique hue selections, it is important to recognize that both used stimuli drawn from a relatively coarse and discrete sampling of color space. This may have decreased sensitivity to individual differences in color category centroids, boundaries and unique hue coordinates. None of our protocols asked informants to select or name best examples of hues, and according to prototype categorization theory these relatively invariant best examples, rather than unique hues or category centroids, should be the most salient (Regier, et al., 2005; Rosch, 1973; 1975). Future experiments should utilize more tightly controlled stimuli for unique hue selection and color naming protocols, presented in continuous rather than discrete format on a computer monitor, and with a protocol for selection of category prototypes in order to better measure hue landmarks which could be salient to perceptual relativity effects.

Finally, our dissimilarity ranking and magnitude estimation protocols and MDS analyses may not have been sensitive enough to accurately capture differences in color perception across subjects—or, possibly, that they did not even represent a true perceptual mapping. We were not able to calculate the statistical power of the MDS analysis (though future work will attempt to do so using a modified permutation test of
MDS solutions); regardless, several effects seen in our results hint that MDS was insufficiently sensitive. While reducing the number of stimuli from 11 to 8 in Experiment III made the protocol faster, and reducing the number of comparisons ought to by definition decrease stress in a given number of dimensions (Agarwhal, 2007), reducing the number of comparisons of similar colors (for example, by eliminating stimulus cards for gold and yellow-chartreuse) increased the weight given by the algorithm to each individual comparison. This likely magnified the effect of “outlier” dissimilarity judgments that conformed poorly to objective hue relationships, causing poorer fit to CIE UV and making the positions of certain hue solutions uninterpretable. Additionally, certain anomalies of color mapping lead us to suspect that our paradigm did not accurately measure color difference perception. For example, the hue map for subject 12 (see Figure 29, p. 76) shows cyan mapped to nearly 0,0 in UV coordinates. Certainly, we do not interpret this to mean that this informant perceives cyan to be virtually indistinguishable from gray; nor do we interpret instances in which cyan and blue are flipped in angular UV space to mean that informants perceive blue to be perceptually closer to green than cyan. In fact, no subjects rated cyan and blue as more dissimilar than green and blue; thus, these effects in MDS solutions must be due to the algorithm’s attempt to fit a solution to disparate magnitude estimations for colors separated more widely in color space. From this, we can infer that, most likely, estimations of dissimilarity for colors widely separated in color space are perceptually difficult, and may require prior knowledge of hue relationships. This could explain the scattershot nature of many of the Somali MDS maps from Experiment III—beyond a certain angle of separation in hue space, estimations of hue dissimilarity are no longer the domain of
perceptual judgment, but are based at least partially on a working concept of a “color wheel” or on non-hue color associations.

This hypothesis is supported by Shepard and Cooper’s lexical color MDS experiment, in which they had color-normal English speakers sort pairs of color names, rather than pairs of colors, and found strikingly similar results to the perceptual MDS experiment performed with color pairs (1992). In spite of this evidence that dissimilarity judgments across large regions of color space are as much cognitive as perceptual tasks, some studies have shown success improving the hue relationships in MDS experiments by using isoluminant stimuli. One recent such study by Boehm, MacLeod, and Bosten (2014) shows generally excellent preservation of fiducial hue ordering; yet, even in this experiment, several color-normal English-speaking informants show numerous hue reversals. Other methodologies for MDS data collection do exist, such as free-association tasks which measure probability of affiliation to generate proximity matrices, and which likely are not as sensitive to outlier effects as traditional color dissimilarity rating or ranking paradigms (Bimler and Uuskula, 2014; Chollet, Valentin, and Abdi, 2014; Lindsey and Brown, 2013).

Additionally, comparisons of both suprathreshold (Kay and Kempton, 1984) and just-noticeable color differences (Witzel and Gegenfurtner, 2013) have been used to compare perception of color differences along categorical boundaries, with their results showing mixed support for universalist and relativist positions. Of course, these studies are not without their problems. The interpretation of Kay and Kempton’s study depends somewhat on the assumption that the color samples are equally perceptually spaced, while Witzel and Gegenfurtner’s study is a methodological tour de force that would be
extremely difficult to replicate in the field with experimentally naïve subjects. Future experiments testing the link between color naming, unique hue selection, and color perception should find a way to improve and simplify these more sensitive measures of color dissimilarity estimation, along with more continuous modes of color naming and unique hue selection, in order to improve sensitivity to potential perceptual relativity effects and decrease the probability of false negative errors in data analysis.

**English Color Naming of Munsell Samples Shows an Abney-like Effect**

Naming of blue-to-green samples showed evidence of a perceptual shift similar to the Abney effect, in which (among other perceptual hue shifts) desaturation of bluish-green colors causes them to appear more blue, due to nonlinearities in the behavior of the opponent channels (Abney, 1909; Burns, Elsner, Pokorny, and Smith, 1984). This effect was originally demonstrated for spectral lights desaturated by white light, but has also been demonstrated with surface colors, including Munsell samples, desaturated under a variety of conditions (Pridmore, 2007). Due to physical constraints on printing, blue-to-green Munsell samples are necessarily desaturated at higher lightness. However, the Munsell color system is designed to correct for this, with samples of constant Munsell hue and varying Munsell chroma and value “contoured” through CIE space to appear isochromatic (Newhall, 1940). In our color-naming experiments, informants routinely named blue-to-green samples more frequently with “blue” and less frequently with “green” as lightness increased. In order to quantify this apparent shift, we determined the greenest samples modally named with the “blue” gloss for every Munsell value level, then plotted their CIE xy coordinates into the CIE xy color solid. This plot shows a marked curvature, similar to that shown by Burns, Elsner, Pokorny, and Smith (1984) for
constant hue appearance, and Lindsey and Brown (2014) with similar color-naming data, in spite of the Munsell system’s correction for the Abney effect. It is possible that the Munsell samples have not been sufficiently calibrated to compensate for the Abney effect, or that the perceptual-cognitive criterion for blue-green categorization is independent from that for hue constancy. See Figure 34 for an illustration of this effect.

Figure 34. Abney Effect-like Shift in English Color Naming. Blue line plots CIE xy coordinates of greenest samples modally named “blue” by subjects in this experiment; purple line plots similar data from Lindsey and Brown (2014). Black and red lines indicate constant hue judgments taken from Burns, Elsner, Pokorny, and Smith (1984). Green line indicates CIE xy coordinates for all values of WCS hue 19 (Munsell hue 7.5G). In this study, I19 and C27 were the greenest and most violet samples modally named “blue.”

Conclusions

While others have shown excellent hue relationships in MDS experiments with heteroluminant stimuli, our color dissimilarity tasks using heteroluminant colors showed evidence of luminance-dependent effects, even after partial correction via use of white
background for our stimuli. Additionally, we believe that dissimilarity magnitude estimations or rankings across large angles in color space may require knowledge of color relationships; or, that if these relationships are not known, subjects default to comparisons along dimensions of color that are currently cryptic to us.

Though numerous Somali informants showed dissimilarity sort behavior consistent with previously collected color naming patterns for purple and yellow, we do not take this to be definitive evidence of a Whorfian effect. Only 1/3 of informants whose MDS solutions revealed this behavior used the YELLOW-OR-ORANGE category term to name the purple sample; additionally, with one notable exception, removing comparisons of purple from these informants’ dissimilarity sorts did not significantly improve hue relationships or fit to CIE UV space, indicating that purple-yellow conflation is not the only deviation from CIE UV in an otherwise well-ordered perceptual hue space. Finally, population-level analysis of Somalis’ sorts revealed hue relationships very similar to those from English-speaking subjects. The picture of the relationship between Somali color naming and perception for purple and yellow is thus unclear: while many Somalis use the same word for purple and yellow, not all show perceptual sequela of this as revealed by MDS; and some Somalis who categorically separate purple and yellow do conflate them on MDS tasks.

Likely, there is some difference from CIE UV in the color dimensions salient to many Somali speakers which causes both of these effects, and while this difference may well be a Whorfian effect, no analyses within the scope of our protocols were able to isolate or characterize it. However, we do report the novel findings of both categorical and perceptual purple-yellow affiliation. Characterizations of color categorization, even
by linguistic relativists, have always demanded that individuals’ color categories be continuous in color space (Roberson et al., 2000); yet this experiment and previous color naming experiments from our lab (Lindsey and Brown, 2012; submitted) show evidence that for many Somali informants, the YELLOW-OR-ORANGE category is not.

English color naming data shows evidence of emerging near-basic categories such as peach, beige, aqua, teal, lavender, and magenta, as well as mustard, a bimodal category with low group mutual information. Additionally, English color naming of Munsell samples shows a pronounced Abney effect-like shift in categorization of blue-to-green samples, despite the system’s correction for this effect.

In summary, we see that while color sorts with heteroluminant stimuli show good hue ordering in most English-speaking subjects, sorts can still be biased by a number of factors beyond hue dimensions. Additionally, based on the results of most Somali and some English speaking informants’ sorts, we extend this finding to conclude that MDS is likely limited as a paradigm for quantification of individual differences in hue space perception in naïve subjects. We also show some evidence of a Whorfian-like effect on purple and yellow for some Somali speakers, though the imprecise nature of the relationship leaves a linguistic relativity interpretation open for debate. Finally, we describe a Somali YELLOW-OR-ORANGE category which defies conventional criteria for color grouping, along with continued evolution of the English color lexicon, including the current coalescence of a categorical term for mid-value yellows and yellow-greens. While these results do not emphatically overturn Berlin and Kay’s linguistic universalism of color cognition, they offer a strong challenge to their depiction of the foundations of categorization, and to their maximum number of universally shared color categories.
Bibliography


Newton, I. (1671). A letter of Mr. Isaac Newton, Professor of the Mathematicks in the University of Cambridge; containing his new theory about light and colors: Sent by the Author to the Publisher from Cambridge, Febr. 6 1671/72; in order to be communicated to the R. Society. Philosophical Transactions of the Royal Society, (6), 3075-3087.


Appendix A: Study Materials
English Volunteer Consent Form

This is a copy of the consent form for volunteer subjects, used for English speaking informants in Experiments I and II, and for PSYC 1110 students in Experiment IV.

The Ohio State University Consent to Participate in Research

Study Title: Color Perception in Somali, African-American, and Caucasian Adults

Principle Investigator: Dr. Angela M. Brown

Sponsor: National Science Foundation

- This is a consent form for research participation. It contains important information about this study and what to expect if you decide to participate. Please consider the information carefully. Feel free to discuss the study with your friends and family and to ask questions before making your decision whether or not to participate.

- Your participation is voluntary. You may refuse to participate in this study. If you decide to take part in the study, you may leave the study at any time. No matter what decision you make, there will be no penalty to you and you will not lose any of your usual benefits. Your decision will not affect your future relationship with The Ohio State University. If you are a student or employee at Ohio State, your decision will not affect your grades or employment status.

- You may or may not benefit as a result of participating in this study. Also, as explained below, your participation may result in unintended or harmful effects for you that may be minor or may be serious depending on the nature of the research.

- You will be provided with any new information that develops during the study that may affect your decision whether or not to continue to participate. If you decide to participate, you will be asked to sign this form and will receive a copy of the form. You are being asked to consider participating in this study for the reasons explained below.

1. Why is this study being done?
This is a study of color perception in Somali, African-American, and Caucasian adults. We know that certain parts of the eye change as we get older. For example, the eye’s lens becomes yellowish and the eye sensitivity to certain colored lights is less in older people than in younger people. The effects may also depend on where a person grew up, whether near or far from the equator. We are studying the possible effects of these changes on color vision.

2. How many people will take part in this study?
300
3. What will happen if I take part in this study?
We may test your color vision using tests printed on paper or tests shown by computer. These tests will require you to tell us what you see, or to arrange colors in order.

We may measure the color of your eye’s lens using colored lights. You will tell us whether you see certain lights or not, whether a light is flickering or not, or whether a pattern moves to the right or to the left.

We may find out about your perception of color by showing you colors by computer or using papers. You will tell us what you would call each color, or which colors look the same and which colors look different.

4. How long will I be in the study?
We would like to test you for one to three testing sessions. Each session will be from one-half to two hours long.

5. Can I stop being in the study?
You may leave the study at any time. If you decide to stop participating in the study, there will be no penalty to you, and you will not lose any benefits to which you are otherwise entitled. Your decision will not affect your future relationship with The Ohio State University.

6. What risks, side effects, or discomforts can I expect from being in the study?
There are no known risks to any of these procedures, beyond the risks that might be involved in any trip outside your home.

7. What benefits can I expect from being in the study?
There are no direct benefits to you. We will learn more about color vision.

8. What other choices do I have if I do not take part in the study?
You may choose not to participate without penalty or loss of benefits to which you are otherwise entitled.

9. Will my study-related information be kept confidential?
Efforts will be made to keep your study-related information confidential. However, there may be circumstances where this information must be released. For example, personal information regarding your participation in this study may be disclosed if required by state law. Also, your records may be reviewed by the following groups (as applicable to the research):

- Office for Human Research Protections or other federal, state, or international regulatory agencies;
- U.S. Food and Drug Administration;
- The Ohio State University Institutional Review Board or Office of Responsible Research Practices;
- The sponsor supporting the study, their agents or study monitors; and
• Your insurance company (if charges are billed to insurance).

If the study involves the use of your protected health information, you may also be asked to sign a separate Health Insurance Portability and Accountability Act (HIPAA) research authorization form.

10. What are the costs of taking part in the study?
There are no costs of taking part in this study.

11. Will I be paid for taking part in this study?
You will not receive payment or any other consideration for participating in this research.

12. What happens if I am injured because I took part in this study?
If you suffer an injury from participating in this study, you should notify the researcher or study doctor immediately, who will determine if you should obtain medical treatment at The Ohio State University Medical Center.

The cost for this treatment will be billed to you or your medical or hospital insurance. The Ohio State University has no funds set aside for the payment of health care expenses for this study.

13. What are my rights if I take part in this study?
If you choose to participate in the study, you may discontinue participation at any time without penalty or loss of benefits. By signing this form, you do not give up any personal legal rights you may have as a participant in this study.

You will be provided with any new information that develops during the course of the research that may affect your decision whether or not to continue participation in the study.

You may refuse to participate in this study without penalty or loss of benefits to which you are otherwise entitled.

An Institutional Review Board responsible for human subjects research at The Ohio State University reviewed this research project and found it to be acceptable, according to applicable state and federal regulations and University policies designed to protect the rights and welfare of participants in research.
14. Who can answer my questions about the study?

For questions, concerns, or complaints about the study you may contact Dr. Angela M. Brown at The Ohio State University College of Optometry (614) 292-4423.

For questions about your rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-6231.

If you are injured as a result of participating in this study or for questions about a study-related injury, you may contact Dr. Angela M. Brown at The Ohio State University College of Optometry (614) 292-4423.
CONSENT

IRB Protocol Number: 2008H0154
IRB Approval Date: Version: VOLUNTEER
Biomedical/Cancer

Signing the consent form

I have read (or someone has read to me) this form and I am aware that I am being asked to participate in a research study. I have had the opportunity to ask questions and have had them answered to my satisfaction. I voluntarily agree to participate in this study.

I am not giving up any legal rights by signing this form. I will be given a copy of this form.

- Printed name of subject ____________________  Signature of subject ____________________
  Date and time: __________ AM/PM

- Printed name of person authorized to consent for subject (when applicable) ____________________  Signature of person authorized to consent for subject (when applicable) ____________________
  Relationship to the subject: ____________________  Date and time: __________ AM/PM

Investigator/Research Staff

I have explained the research to the participant or his/her representative before requesting the signature(s) above. There are no blanks in this document. A copy of this form has been given to the participant or his/her representative.

- Printed name of person obtaining consent ____________________  Signature of person obtaining consent ____________________
  Date and time: __________ AM/PM

Witness(es) – May be left blank if not required by the IRB

- Printed name of witness ____________________  Signature of witness ____________________
  Date and time: __________ AM/PM

- Printed name of witness ____________________  Signature of witness ____________________
  Date and time: __________ AM/PM

Doc date: 07/12/13  Page 5 of 5  Form date: 10/07/08
Print date: 07/12/13
English Paid Consent Form

This is a copy of the consent form for paid subjects, used for English-speaking informants in Experiment III and Optometry students in Experiment IV.
We may also measure the color of your eye's lens using colored lights.

4. How long will I be in the study?
One to three testing sessions, which will last less than one and a half hours each.

5. Can I stop being in the study?
You may leave the study at any time. If you decide to stop participating in the study, there will be no penalty to you, and you will not lose any benefits to which you are otherwise entitled. Your decision will not affect your future relationship with The Ohio State University.

6. What risks, side effects or discomforts can I expect from being in the study?
There are no known risks to any of these procedures, beyond the risks that might be involved in any trip outside your home.

7. What benefits can I expect from being in the study?
There are no direct benefits to you. We will tell you anything we learn about your color vision.

8. What other choices do I have if I do not take part in the study?
You may choose not to participate without penalty or loss of benefits to which you are otherwise entitled.

9. Will my study-related information be kept confidential?
We will try very hard to keep your study-related information confidential. However, there may be circumstances where this information must be released. For example, personal information regarding your participation in this study may be disclosed if required by state law. Also, your records may be reviewed by the following groups:
   - Office for Human Research Protections or other federal, state, or international regulatory agencies;
   - U.S. Food and Drug Administration;
   - The Ohio State University Institutional Review Board or Office of Responsible Research Practices;
   - The sponsor supporting the study, their agents or study monitors; and
   - Your insurance company (if charges are billed to insurance).

If the study involves the use of your protected health information, you may also be asked to sign a separate Health Insurance Portability and Accountability Act (HIPAA) research authorization form.

10. What are the costs of taking part in this study?
There are no costs of taking part in this study.
11. Will I be paid for taking part in this study?
You will receive a gift card valued at $10.00 redeemable for merchandise from a local
business. Payment will be given to you at the end of each visit but notice, by law, payments to
subjects are considered taxable income.

12. What happens if I am injured because I took part in this study?
If you suffer an injury from participating in this study, you should notify the researcher or
study doctor immediately, who will determine if you should obtain medical treatment at The
Ohio State University Medical Center. If other problems arise as a result of the study, contact
Dr. Angela M. Brown at the Ohio State University College of Optometry if (614-292-4423).
The cost for this treatment will be billed to you or your medical or hospital insurance. The
Ohio State University has no funds set aside for the payment of health care expenses for this
study.

13. What are my rights if I take part in this study?
By signing this form, you do not give up any personal legal rights you may have as a
participant in this study.
You will be provided with any new information that develops during the course of the
research that may affect your decision whether or not to continue participation in the study.
You may refuse to participate in this study without penalty or loss of benefits to which you
are otherwise entitled.

An Institutional Review Board responsible for human subjects research at The Ohio State
University reviewed this research project and found it to be acceptable, according to
applicable state and federal regulations and University policies designed to protect the rights
and welfare of participants in research.

14. Who can answer my questions about the study?
For questions, concerns, or complaints about the study you may contact Dr. Angela M. Brown
at the Ohio State University College of Optometry (614) 292-4423.
For questions about your rights as a participant in this study or to discuss other study-related
concerns or complaints with someone who is not part of the research team, you may contact
Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-0251.

Signing the consent form
I have read (or someone has read to me) this form and I am aware that I am being asked to
participate in a research study. I have had the opportunity to ask questions and have had them
answered to my satisfaction. I voluntarily agree to participate in this study.
I am not giving up any legal rights by signing this form. I will be given a copy of this form.

Printed name of subject: ___________________________ Signature of subject: ___________________________ AM/PM

Date and time: ___________________________

Printed name of person authorized to consent for subject (when applicable): ___________________________

Signature of person authorized to consent for subject (when applicable): ___________________________ AM/PM

Relationship to the subject: ___________________________

Date and time: ___________________________

Investigator/Research Staff

I have explained the research to the participant or his/her representative before requesting the signature(s) above. There are no blanks in this document. A copy of this form has been given to the participant or his/her representative.

Printed name of person obtaining consent: ___________________________

Signature of person obtaining consent: ___________________________ AM/PM

Date and time: ___________________________

Witness(es) - May be left blank if not required by the IRB

Printed name of witness: ___________________________

Signature of witness: ___________________________ AM/PM

Date and time: ___________________________

Printed name of witness: ___________________________

Signature of witness: ___________________________ AM/PM

Date and time: ___________________________
English HIPAA Form

This is a copy of the HIPAA privacy form used for English-speaking subjects, which participants signed to authorize us to use PHI such as name, age, and gender.

THE OHIO STATE UNIVERSITY
AUTHORIZATION TO USE
PERSONAL HEALTH INFORMATION IN RESEARCH

Title of the Study: Color perception in Somali, African-American, and Caucasian adults
Protocol Number: 2008H0154
Principal Investigator: Angela M Brown

Subject Name

Before researchers use or share any health information about you as part of this study, The Ohio State University is required to obtain your authorization. This helps explain to you how this information will be used or shared with others involved in the study.

- The Ohio State University and its hospitals, clinics, health-care providers, and researchers are required to protect the privacy of your health information.
- You should have received a Notice of Privacy Practices when you received health care services here. If not, let us know and a copy will be given to you. Please carefully review this information. Ask if you have any questions or do not understand any parts of this notice.
- If you agree to take part in this study your health information will be used and shared with others involved in this study. Also, any new health information about you that comes from tests or other parts of this study will be shared with those involved in this study.
- Health information about you that will be used or shared with others involved in this study may include your research record and any health care records at The Ohio State University. For example, this may include your medical records, x-rays, or laboratory results. Psychotherapy notes in your health record (if any) will not, however, be shared or used. Use of these notes requires a separate, signed authorization.

Please read the information carefully before signing this form. Please ask if you have any questions about this authorization, the university’s Notice of Privacy Practices or the study before signing this form.

Initials/Date: ______________________

Page 1 of 3 03/16/12
• Those who oversee the study will have access to your information, including the following:
   - Members and staff of The Ohio State University’s Institutional Review Board, including the Western Institutional Review Board
   - The Ohio State University Office of Responsible Research Practices
   - University data safety monitoring committees
   - The Ohio State University Office of Research

• Your health information may also be shared with federal and state agencies that have oversight of the study or to whom access is required under the law. These may include the following:
   - Food and Drug Administration
   - Office for Human Research Protections
   - National Institutes of Health
   - Ohio Department of Job and Family Services.

The information that is shared with those listed above may no longer be protected by federal privacy rules.

Initials/Date__________________
Authorization Period

This authorization will not expire unless you change your mind and revoke it in writing. There is no set date at which your information will be destroyed or no longer used. This is because the information used and created during the study may be analyzed for many years, and it is not possible to know when this will be complete.

Signing the Authorization

• You have the right to refuse to sign this authorization. Your health care outside of the study, payment for your health care, and your health care benefits will not be affected if you choose not to sign this form.

• You will not be able to take part in this study and will not receive any study treatments if you do not sign this form.

• If you sign this authorization, you may change your mind at any time. Researchers may continue to use information collected up until the time that you formally changed your mind. If you change your mind, your authorization must be revoked in writing. To revoke your authorization, please write to:

  Dr. Angela M Brown, College of Optometry, 338 W 10th Ave, Columbus, OH 43210

  Or the HIPAA Privacy Officer, The OSU College of Optometry, 338 W 10th Ave, Columbus, OH 43210, tel. 614-292-3020

• Signing this authorization also means that you will not be able to see or copy your study-related information until the study is completed. This includes any portion of your medical records that describes your study treatment.

Contacts for Questions

• If you have any questions relating to your privacy rights, please contact: Mr. Shawn Currier, College of Optometry, 338 W 10th Ave, Columbus, OH 43210 or call (614)292-0841.

• If you have any questions relating to the research, please contact: Dr. Angela M Brown, College of Optometry, 338 W 10th Ave, Columbus, OH 43210 or call (614) 292-4423.

Signature

I have read (or someone has read to me) this form and have been able to ask questions. All of my questions about this form have been answered to my satisfaction. By signing below, I permit [Insert name of Principal Investigator] and the others listed on this form to use and share my personal health information for this study. I will be given a copy of this signed form.

Signature

(Signed or Legally Authorized Representative)

Print Name __________________________ Date ________ Time ________ AM/PM

(If legal representative, also print relationship to subject)
Somali Paid Consent Form

This is a copy of the consent form for paid subjects, used for Somali speaking informants in Experiment III.

Warqada ogalaanshaha ee ka qayb galka cilmibaarista ee jaamacada gobalka Ohio.

Cinwaanka la dersayoz Aragtid midabka ee Soomalida, African American, iyo dadka Cadaanka ah ee waaweyn.

Maamulaha Baaristaanka: Dr. Delwin T. Lindsey

Kaalif: National Science Foundation

- Tani waa warqada ogalaanshaha cilmi baarista. Waxay xanbaaranshada xog muhim ah oo ku saabsan daraasadani iyo waxa la fiilayo haddii aad goostid ka qayb galka. Fadlan, taqgal yihiin si gaar ah. La faan ugu saaxiibadaada iyo gudankaaga ka saabsan qayb galka cilmi barista.

- Ka qayb qaadashadaada waa mid tabaruc ah. Laga yaabce in aad didicid ka qayb qaadashada daraasadani. Haddii aad goostid inaad ka qayb qaadatadi, markaad doonto waxa iska dhaxaasi kartaa. Ge'aan kasta oo aad qaadar, wax dhibta ah oo ka soo gaaraya ma jirto. Faanido kaaga muqaalidaan majirto Go'aankaaga mu saameynayo mustaqbalka xirka adiga iyo jaamacadda Ohio ku dhaxeeyo. Haddii aad arday ama shaqade jaamacadda taahay, ge'aankaagu masaa meynayo derajoodaada iyo qabo kalamaadaada.

- Waxaa lagu yaabaa inaad wax ka faa 'iddid ama aadan ka faa' idiin natiijada soo. Baxda ka qayb galka daraasadan. Saan hoos ugu sharaxnay, ka qaybqalkaagu waxaa lagu yaabaa in natiijadiisii noqoto mid raad waxayee loo kubaa, taas oo noqon karta mid yar iyo mid aad u culus oo ku xiran dabeecada cilmi barista.

- Waxaa lagu soo cod siin doonaa xog kasa oo cusub oo ka soo if baxda inta daraasadu socoto taasoo lagu yaabaa inay saameyn ku yeelada go'aankaaga si wadista isu jecelinta. Haddii aad goostid inaad ka qayb gashay, waxa lagu weyn diin doonaa inaad saxiixdaa warqadan, waxaanad heli doonaa muqalka wey diin doonaa inaad saxiixdaa warqadan, waxaanad heli doonaa muqalka, warqadan. Waxa lagu weyn diin doonaa tixgelinta ka qayb galka daraasadan sababalka hoos.

date: 11/13/2015
print date: 11/12/2015
lagu sharaxay dartood.

1) Maxaa daraasadan loo qabtay?
Tani waa daraasad ku saabsan aragtida midabka Somaliga, Afican American, iyo dadka Cadaanka ah. Waxaynu ognahay in qaybo isha ka mid ah ay isbadasho kadib markii laweynadaaba, saameyntaasna waxa laga yaaba inay ku xiran tahay meeshaa uu qofku ku soo korey. Waxaynu dersaynayaa saameynuta suuragalka ah ee bedasha aragtida midabka.

2) Immisa dad ah baa ka qayb qaadan doona deraasad? 700


4) Immisa ayaan ku jiru doona deraasad? Hal ila sadex kulayn, middiina waxay dhamaanisa saacad iyo bar wuxi kuyar.

5) Ma joogin karaa ka mid noqoshada deraasad?
baad ka tegi kartaa. Hadaad goosatid inaad joogisid ka qayb galka deraasadan, wax dhibato ah oo kaal soo gaaraya majirto, faaide kaaga nuqsaanaysanaa majirto. Go'aankagu ma'saameenay5 mustaqbalka xiriirkada adiga iyo jaamacadda ka dhaxceeya.

6) Inta aan kujiirro deraasad, ma firi karaa inay I soo gaarto khatar, aaf dhinaceed ama raaxo dar? Majirto wax khatar ah oo lagaranayo oo saameyn ku leh habraaca deraasad lawado, taasoo ka baaxsan khatar kasta oo kugu dhiici karta safarada kabaxsan guurigaaga.
7) Waa maxay faa’iidoyinka aan ka filan karo ka mid noqoshada deraasadan? Wax faa’iido ah oo toos ah majirto. Waxaan kaliya oo ku sheegaynaa wixii aan kabaranay aragida midabkaaga.

8) Maxaa doorashooyin kale ii furan haddii aanan ka qayb qaadan deraasadan? Waxaad dooran kartaa inaad ka qayb galin ciqaab ja’an, ama faa’iido daro oo aad xaq u lahayd.

9) Deraasadan ku saabsan xogtu ma noqon doontaa mid qarsoodi ah? Waxaa aad ugu caddaalayaan in xogtaas noqto mid qarsoodi ah. Si Kastaba ha ahaatee, waxaa laga yaabaa inay daruur noqon doonto in uun mar xogtaas la sideway doono. Tusaale Xogta ku saabsan qofkii ka qayb galay deraasadan, waxaa laga yaabaa in lagu lifaad hoodo haddii xecrka gobalka u baahdo.
- Waxaa kaloo laga yaabaa in xogtaada ay kooxaha 5oo socda dib u eegan.
- Xafiiska ilaalinta bini aadnimada, ama kuwo dowlada, gobalka iyo kuwa Caalamiga ah.
- Maamulka cuntada iyo mukhaadiradka.
- Xafiiska mas’uulka ah arimaha cilmii baarista ee jamacada gobalka Ohio.
- Hay’ada mas’uulka ka ah cilmii barista, hay’adaheeda, ama kuwa duraadu hagaasiya iyo caamiskaaga.
Haddii baaristu la xiriirto isticmaalka xogta dhowrsan ee caafimaadkaga, waxaa kale oo lagaa condsan karaa inaad saxiibid warqad kale oo ku saabsan cilmii baaristaada.

10) Waa maxay qiimaha ay ku kici karto ka qaybqaadashada deraasadan? Ma jiraan wax qiiimo ah oo ay ku kacayso

11) Aniga wax ma la iigu siinayaa ka qayb qaadashada deraasadan?
Booqasho kasta waxaad heli doontaa kaar $25.00 ah oo aad kaga adeegan kartid meelaha ganacsiga. Sida shariga ahna waxaa saaran canshuur.

Qimaha wax kaqabashada caafimaadkaaga waxaa lagu soo dalici doona, adiga, karaanka caafimaadka, ama cisisaltaaga. Jaamacada Ohio uma hayso wax kharash ah oo lagu bixiyo wax kaqabashada caafimaadkaaga.


Waxaa lagu socdesiin doonaa xog kasta oo cusub oo soo iibadha intha cilmii baarista socoto. Tsasoo laga yaabo inay saameyn ku yeelato gdaankaaga ahaa inaad ku jirto ama ka joogto ka qayb qaadashada cilmii baarista. Waad diidi kartaa ka qayb qaadashada danbi la'aan, faa' idana kaaga muqsaami maysa. Hay'ada waakilka ka ah cilmii baarista ku saabsan arimaha ban'aadananimada ee jaamacada Gobalka Ohio oo dib u baartay mashruucan cilmii baarista ayaa sheegay mid la aqbal karo inay tahay iyo dhammaan inta ilaaliisa xuquuqda la aqbal karo inay tahay iyo dhammaan inta ilaaliisa xuquuqda aadaniga.

14) Kumaa su'aalahaayga ku saabsan cilmii baarista ka jawaabi kara? Wixii su'aal iyo cabashooyin ku saabsan, waxaad kale xiriiri karta Dr Angela M Brown, tel# 614 292-4423 ee ku liiyada cabiraada indhaha jaamacada.
Saxiix Warqada Ogalaanashaa:
Waxaan akhriyey warqadan waxaana dareensanahay in la weydiindoono ka qayb qaadashada Deraasadan cilmii baarista. Waxaan fursad u haystaa in aan waydiyey su'aaloo la iigana jawaabto amigoo raali ah. Waxaan si tabarucnaad ah u aqbalay in aan ka qaybqato deraasadan.
Marka aan saxiixto warqadan, wax xuquq ah oo iga Iimaysaa ma jirto. Waxaa la I siinidona warqadan muqulkeeda.

(Print Name) Daabaca Magaza Qofka (Signature) Saxiixa Qofka

(Date/Time) Tariikh/Wakhti (Am/Pm)/Subax/Galab

I have explained the research to the participant or his/her representative before requesting the signature(s) above. There are no blanks in this document. A copy of this form has been given to the participant or his/her representative.

Printed name of person obtaining consent  Signature of person obtaining consent

Date and time AM/PM

Doc date: 11/13/2015  Print date: 11/13/2015 5

126
Somali HIPAA Form

This is a copy of the HIPAA privacy form used for Somali speaking subjects, which participants signed to authorize us to use PHI such as name, age, and gender.

THE OHIO STATE UNIVERSITY
OGALAANSHAHA IN WAR CAAMFAAD EE SHAKHSIGA LAGU
ISTICMAALO CILMI BAADHIS
(THE OHIO STATE UNIVERSITY AUTHORIZATION TO USE PERSONAL HEALTH
INFORMATION IN RESEARCH)

Magaca Darasaadeh: Color Perception In Somali, African-American, And Caucasian Adults
Lambarka Mammanka OSU (OSU Protocol Number): 2008H0154
Baadhaaha Kooban: Dr. Delina L. Lindsey

Maqaca Lawba-Nsho __________________________

Ku hor nista aad cilmi-baadheysayn intifadaha ama wadaada wuxuu war caamfaad ah oo aad ku mugagto waalintaase ahaan, waa Ohio State University (Janacood Cibroodinta Oo ahaan) in laga rabah ugu horey ee ku hadliyso. 

• Ohio State University iyo cilmi-baadheysa, kuwaasheeda iyo daryeel caamfaadderinta oo cilmi-baadheysa waa sida rabah ugu horey fiican, astaanaadkeeda waqti guud ku hee daabso.

Waray shayd inaad hadashay Ooqaynta Dhaqanka Antumarka warbaan dad noo fahmo daryeel caamfaad haddii nido dama wixii kala u gaar ah. Haddii nido dama wixii kala u gaar ah, madax caawimaad iyo daryeel caamfaad, ku heli kara jirin. Fadlan si waaye 3 hooyka si aan jirin, kuwaasheeda ijyada adiga pipirka caamfaadka.

• Waray shayd inaad hadashay Ooqaynta Dhaqanka Antumarka waxaa ahaan. Haddii nido dama wixii kala u gaar ah, madax caawimaad iyo daryeel caamfaad, ku heli kara jirin. Fadlan si waaye 3 hooyka si aan jirin, kuwaasheeda ijyada adiga pipirka caamfaadka.

Haddii nido dama wuxu noqon kara jirin, warbixinta caawimaadka waxaa loo yaqaan ilaa daryeel caamfaad, leh daabso xukuumad. Sidoo kale, waxaa aan xanuunka, kuwaasheeda iyo daryeel caamfaad, haddii nido dama wuxu noqon kara jirin, warbixinta caawimaadka waxaa loo yaqaan ilaa daryeel caamfaad. Sidoo kale, waxaa aan xanuunka, kuwaasheeda iyo daryeel caamfaad, haddii nido dama wuxu noqon kara jirin, warbixinta caawimaadka waxaa loo yaqaan ilaa daryeel caamfaad. Sidoo kale, waxaa aan xanuunka, kuwaasheeda iyo daryeel caamfaad, haddii nido dama wuxu noqon kara jirin, warbixinta caawimaadka waxaa loo yaqaan ilaa daryeel caamfaad.

Fadlan si waaye 3 hooyka si aan jirin, kuwaasheeda ijyada adiga pipirka caamfaadka. Fadlan si waaye 3 hooyka si aan jirin, kuwaasheeda ijyada adiga pipirka caamfaadka. Fadlan si waaye 3 hooyka si aan jirin, kuwaasheeda ijyada adiga pipirka caamfaadka. Fadlan si waaye 3 hooyka si aan jirin, kuwaasheeda ijyada adiga pipirka caamfaadka.

Koobi: Magacyro Tairirka: ________________
Indhiis: Date

Page 1 of 4
(1 of 4)
Kuwa laga yaabo inay isticmaalisan, Wadaagaan ee Helo Karaan Warkaaga Isaga oo Qayb ka Ah Darasaaddan

- Cilmiga-baahdaya axisha Ohio State University ayaq darasaadda cilmiga-baahdhiisa ah, xaqab awood waakant iyo baahdo warka caafimaadkaaga. Shaqaalaha kale ee Ohio State University ee aan iyo xaqab darasaadda laakiin laga yaabo inay hawlgu qeexaay xayeelkaaga sababtoo oo ah awso la xidhiidha darasaadda ayaa sidoo kale helitaan waa xoolaha ayeen doona warkaaga.

Daakka kormeerka darasaadda ayaa helitaan waa xoolaha ayeen doona warkaaga, waxaa ku jira:

- Xanuufka axisha shaqaalaha Gedabada Dib-ta-fooostaa Hay'adda (Institutional Review Board) ee Ohio State University, oo ku jira Western Institutional Review Board
- Office for Responsible Research Practices (Xaafinta Dhaqaalada Cilmiga-baahdis ee Leh Mas'umiyad Wacan)
- Guudiyada kormeerka amaanka xogta ee Jasanadda
- Ohio State University Research Foundation (Hay'adda Samaafalka ah ee Cilmiga-baahdis ee Jasanadda Gobolka Ohio)

Waxa sidoo kale warka caafimaadkaaga lala wadaag karaa hay'adda federaalka iyo gobolka ee kormeerka darasaadda ama ciddii kale ee loo baahdo iyadoo sharigis la tix raacyo. Waxa kuwa ku jira kara:

- The Food and Drug Administration (Maxmiinta Cintada iyo Dawooyinka)
- The Office for Human Research Protections (Xaafinta Badbaadinta Bini'addanka la Cilmiga-baahdhiyo)
- The National Institutes of Health (Machayada Caafimaadka ee Qaran)
- The Ohio Department of Human Services (Wasaaradda Aseeyada Bini'addanka ee Ohio)

Cilmiga-baahdayaas, shirkadda axisha hay'adda ku baxaan Ohio State University ayaa sidoo kale laga yaabo in ay isticmaalisan, wadaagaan ama helaan warka caafimaadkaaga isaga oo la xidhiidha darasaaddan:

- Goobaha daryeel caafimaad, goobaha cilmiga-baahdis, taaxiyeysa daryeel caafimaad, ama kormeerayaad daramadaad ee looga qeyb doonaayn: None
- Sharyabshada gaarka ah iyada dalka iyada hay'adda kale ee dhinca warka caafimaadkaaga isaga oo la xidhiidha darasaaddan: None
- Qabaqabayxia cilmiga-baahdisa iyaha shirkadda u leeyahay ama xidhiidha la aha qabaqabayxia: The National Science Foundation
- Hay'adda Cilmiga-baahdisa ee Qanduraas Leh: None
- Guudiyada kormeerka xogta axisha amaanka iyada cidda kale ee kormeerka amarkaas cilmiga-baahdisa: The Ohio State University Institutional Review Board
- Ciid kale: None

Waxa laga yaabaa in warka lala wadaago kuwa kor loogu qoray xamay dib u badbaad kara qeybarka astumaanta ee federaalka.

Muqadda Ogalaanuubaha

Kooxiiis Maamucyo/Taaskiib: ____________________________
Initials/Date: ____________________________

Bogga 2 ka 4 (Page 2 of 4)
Ogalanashahan wakhtiga su ma dhib doono haddii aamad beddeinta uu yigaaga oo aamad si qoraal ah dib ugu noqon. Ma jurtu taamik uu an oo warkaaga la baabi in doono ama aan la isteemal doono. Taa saabsan keemooyay waxa weeye iyada oo laga yaabo in muu badan u hadda warka la isteemaalay ee la abuuray intii ay daraasadda socotay, sun Rugalmi aamay shayn in la ogado goorta taas la dhameyn doono.

**Saxeexidda Ogalanashaha**

- Waa aad no qab u leedahay inaad diid saxeexidda ogalamanashahan. Haddii aad doortid inaad saxeesin warqaddan wax sameeyn ah ma soo gaddii doontay daryeelka caasimadkaaga ee ka baxsan daraasadda, buuxda kharshku daryeelka caasimadkaaga, yoo fashooyinka daryeelka caasimadkaaga.

- Haddii aamad saxeesin warqaddan uu awoodi doortid inaad ka quybaashid daraasadda ama hehi doontay wax ah daweyn la yidhiida daraasadda.

- Haddii aad saxeesid ogalamanashahan, wolkl kasta ayaa beddeli kartaa go'aankaaga. Wuxuu ciimmahaddhayad u sii wadi karaan isteemalka warqo la ursi dig ilaa yeq wakhtiga aad si raami ah u beddehay go'aankaaga. Haddii aad beddeeshid go'aankaaga, waa inaad si qoraal ah dib ugu noqot ogalamanashahaaga. Si aad dib ugu noqot ogalamanashahaaga, fadlan warqad u soo dir: Dr. DeWitt T. Lindsey, PhD; lindsey.45@osu.edu; (419)-555-4359), Dr. Angela M. Brown, PhD; Brown.112@osu.edu; 614-292-4423 and Dr Jacqueline Davis, OD, MPH; JDavis3@optometry.osu.edu; 614-293-2920 College of Optometry, 338 W. 10th Ave. Columbus, OH 43210

Maka aad saxeesa ogalamanashahan waxa macnabrow waayay inaad awoodi karaan inaad xoolid aad aad koox ka sameeynayt warkaaga la yidhiidha dawaynti warqadda ilaa la dhameystiray warqadda waxa ugu jira tagtka oo ka tirsan diwanaad caasimadkaaga oo warqadda dawaynti warqadda la yidhiidha.

**Cidda Su'aalisaha Lagala Xidhiidhayo**

- Haddii aad qabtidi wax su'aal oo la xidhiidha waqooyiga aasumanta, fadlan la xidhiidhi: Insert contact information for the appropriate HIPAA Privacy Officer.

- Haddii aad qabtidi wax su'aal oo la xidhiidha cilmii badbaadaha, fadlan la xidhiidhi: Dr. Angela M. Brown, PhD; Brown.112@osu.edu; 614-292-4423 and Dr Jacqueline Davis, OD, MPH; JDavis3@optometry.osu.edu; 614-293-2920

**Saxeeex**

Waa ahriyay (ana qof ayaa i.ahriyay) warqaddan waxaa aam awooday inaan weydiyo su'aal. Dhammaan su'aalaha ku saabsan waxaa warqaddan si i gamsanay ayaa la iiga jawabaya. Maka aam saxeeso halka hoose, waxa aan Dr. Angela M. Brown, PhD yoo cidda kale ee ku qora warqaddan u ogalamanaya waa warqadda caasimadkaaga isteemaalan oo wadaagana daraasaddan daraadeed. Waxa la i sin doonaa nuqal ama koobi ah warqaddan saxeesan.

**Saxeeex**

(Lo-baadaha ama Waqtii Sheeri Ah) / Signature of Subject or Legally Authorized Representative

1. **Kooxix Maagaay/Taanik:** ____________________
2. **Initsila/Digaad:** ____________________

Boggaa 3 ka 4
(Page 2 of 4)

129
In Experiment IV, we used magnitude estimation to collect data for classical MDS. This magnitude estimation protocol required that subject maintain stable criteria for color difference rating, so we had subjects rate five practice stimuli before performing the main task in order to calibrate their dissimilarity scales. The five practice stimuli, comprised of yellow/magenta, magenta/cyan, gold/purple, green/yellow-chartreuse, and red/gold pairwise comparisons, are pictured below.
Appendix B: Stimulus Specification
Experiment I MDS Stimuli

This table shows the Color-aid stimuli used for the color dissimilarity sort protocol in Experiment I, along with our names for them, their Munsell equiavelents (if applicable), and their CIE LUV coordinates.

<table>
<thead>
<tr>
<th>Color-Aid ID</th>
<th>Color Name</th>
<th>Munsell Equivalent</th>
<th>L*</th>
<th>u*</th>
<th>v*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gw-HUE</td>
<td>GREEN</td>
<td>5G 5/8</td>
<td>58.562</td>
<td>-59.361</td>
<td>21.976</td>
</tr>
<tr>
<td>BG-HUE</td>
<td>CYAN</td>
<td>10BG 4/8</td>
<td>50.446</td>
<td>-44.607</td>
<td>-18.780</td>
</tr>
<tr>
<td>Bw-T1</td>
<td>BLUE</td>
<td>6.25 PB 4/12</td>
<td>47.380</td>
<td>-22.862</td>
<td>-59.669</td>
</tr>
<tr>
<td>BV-T1</td>
<td>VIOLET</td>
<td>8.75PB 4/12</td>
<td>47.259</td>
<td>-2.485</td>
<td>-59.940</td>
</tr>
<tr>
<td>V-T1</td>
<td>PURPLE</td>
<td>2.5P 4/12</td>
<td>45.925</td>
<td>14.298</td>
<td>-48.700</td>
</tr>
<tr>
<td>M-HUE</td>
<td>MAGENTA</td>
<td>2.5RP 4/12</td>
<td>43.527</td>
<td>65.744</td>
<td>-24.920</td>
</tr>
<tr>
<td>R-EX</td>
<td>RED</td>
<td>6.25R 4/14</td>
<td>51.770</td>
<td>118.673</td>
<td>8.133</td>
</tr>
<tr>
<td>RO-EX</td>
<td>RED-ORANGE</td>
<td>6.25YR 5/14</td>
<td>58.620</td>
<td>138.133</td>
<td>16.194</td>
</tr>
<tr>
<td>O-HUE</td>
<td>ORANGE</td>
<td>2.5YR 6/14</td>
<td>73.175</td>
<td>113.534</td>
<td>35.546</td>
</tr>
<tr>
<td>Y-HUE</td>
<td>YELLOW</td>
<td>5Y 8.5/12</td>
<td>94.105</td>
<td>29.997</td>
<td>66.003</td>
</tr>
<tr>
<td>YG-HUE</td>
<td>CHARTREUSE</td>
<td>N/A</td>
<td>74.632</td>
<td>-44.480</td>
<td>47.371</td>
</tr>
</tbody>
</table>
Experiment II MDS Stimuli

This table shows the Color-aid stimuli used for the color dissimilarity sort protocol in Experiment II, along with our names for them, their Munsell equivalents (if applicable), and their CIE LUV coordinates.

<table>
<thead>
<tr>
<th>Color-Aid ID</th>
<th>Color Name</th>
<th>Munsell Equivalent</th>
<th>L*</th>
<th>u*</th>
<th>v*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gw-EX</td>
<td>GREEN</td>
<td>2.5G 5/12</td>
<td>63.997</td>
<td>-62.478</td>
<td>35.868</td>
</tr>
<tr>
<td>BG-T1</td>
<td>CYAN</td>
<td>10BG 4/8</td>
<td>60.218</td>
<td>-50.972</td>
<td>-13.834</td>
</tr>
<tr>
<td>Bc-HUE</td>
<td>BLUE</td>
<td>3.75 PB 4/12</td>
<td>48.505</td>
<td>-37.297</td>
<td>-52.077</td>
</tr>
<tr>
<td>V-T2</td>
<td>PURPLE</td>
<td>8.75PB 4/12</td>
<td>53.589</td>
<td>12.837</td>
<td>-42.377</td>
</tr>
<tr>
<td>M-T1</td>
<td>MAGENTA</td>
<td>2.5RP 4/12</td>
<td>56.772</td>
<td>69.371</td>
<td>-21.482</td>
</tr>
<tr>
<td>R-EX</td>
<td>RED</td>
<td>6.25R 4/14</td>
<td>53.086</td>
<td>118.369</td>
<td>12.294</td>
</tr>
<tr>
<td>O-HUE</td>
<td>ORANGE</td>
<td>2.5YR 6/14</td>
<td>70.331</td>
<td>122.738</td>
<td>39.231</td>
</tr>
<tr>
<td>YO-EX</td>
<td>GOLD</td>
<td>N/A</td>
<td>84.086</td>
<td>76.023</td>
<td>59.746</td>
</tr>
<tr>
<td>Y-HUE</td>
<td>YELLOW</td>
<td>SY 8.5/12</td>
<td>92.254</td>
<td>31.756</td>
<td>72.576</td>
</tr>
<tr>
<td>YGw-HUE</td>
<td>YELLOW-CHARTREUSE</td>
<td>N/A</td>
<td>88.820</td>
<td>-18.283</td>
<td>68.349</td>
</tr>
<tr>
<td>YG-HUE</td>
<td>CHARTREUSE</td>
<td>N/A</td>
<td>78.847</td>
<td>-49.257</td>
<td>57.744</td>
</tr>
</tbody>
</table>

134
**Experiment III/IV MDS Stimuli**

This table shows the Color-aid stimuli used for the binary sort protocol in Experiment III and for the magnitude estimation protocol in Experiment IV, along with our names for them, their Munsell equivalents (if applicable), and their CIE LUV coordinates.

<table>
<thead>
<tr>
<th>Color-Aid ID</th>
<th>Color Name</th>
<th>Munsell Equivalent</th>
<th>L*</th>
<th>u*</th>
<th>v*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gw-EX</td>
<td>GREEN</td>
<td>2.5G 5/12</td>
<td>63.997</td>
<td>-62.478</td>
<td>35.868</td>
</tr>
<tr>
<td>BG-T1</td>
<td>CYAN</td>
<td>10BG 4/8</td>
<td>60.218</td>
<td>-50.972</td>
<td>-13.834</td>
</tr>
<tr>
<td>Bc-HUE</td>
<td>BLUE</td>
<td>3.75 PB 4/12</td>
<td>48.505</td>
<td>-37.297</td>
<td>-52.077</td>
</tr>
<tr>
<td>V-T2</td>
<td>PURPLE</td>
<td>8.75PB 4/12</td>
<td>53.589</td>
<td>12.837</td>
<td>-42.377</td>
</tr>
<tr>
<td>R-EX</td>
<td>RED</td>
<td>6.25R 4/14</td>
<td>53.086</td>
<td>118.369</td>
<td>12.294</td>
</tr>
<tr>
<td>O-HUE</td>
<td>ORANGE</td>
<td>2.5YR 6/14</td>
<td>70.331</td>
<td>122.738</td>
<td>39.231</td>
</tr>
<tr>
<td>Y-HUE</td>
<td>YELLOW</td>
<td>5Y 8.5/12</td>
<td>92.254</td>
<td>31.756</td>
<td>72.576</td>
</tr>
<tr>
<td>YG-HUE</td>
<td>CHARTREUSE</td>
<td>N/A</td>
<td>78.847</td>
<td>-49.257</td>
<td>57.744</td>
</tr>
</tbody>
</table>
Unique Hue Stimuli

This table shows the Munsell stimuli used for the unique hue selection protocols in Experiments II-IV, along with their CIE LUV coordinates.

<table>
<thead>
<tr>
<th>Munsell Sample</th>
<th>L*</th>
<th>u*</th>
<th>v*</th>
</tr>
</thead>
<tbody>
<tr>
<td>10GY 5/12</td>
<td>61.228</td>
<td>-48.154</td>
<td>45.730</td>
</tr>
<tr>
<td>2.5G 5/10</td>
<td>59.406</td>
<td>-47.568</td>
<td>31.028</td>
</tr>
<tr>
<td>5G 5/10</td>
<td>59.897</td>
<td>-50.446</td>
<td>22.961</td>
</tr>
<tr>
<td>7.5G 5/12</td>
<td>60.381</td>
<td>-52.399</td>
<td>18.450</td>
</tr>
<tr>
<td>10G 5/10</td>
<td>61.540</td>
<td>-54.130</td>
<td>13.005</td>
</tr>
<tr>
<td>2.5BG 5/10</td>
<td>59.969</td>
<td>-54.473</td>
<td>7.374</td>
</tr>
<tr>
<td>5BG 5/10</td>
<td>59.806</td>
<td>-56.697</td>
<td>-1.887</td>
</tr>
<tr>
<td>7.5BG 5/10</td>
<td>59.951</td>
<td>-55.752</td>
<td>-8.892</td>
</tr>
<tr>
<td>10BG 5/10</td>
<td>59.479</td>
<td>-56.045</td>
<td>-17.414</td>
</tr>
<tr>
<td>2.5B 5/10</td>
<td>59.038</td>
<td>-53.420</td>
<td>-24.750</td>
</tr>
<tr>
<td>5B 5/10</td>
<td>60.005</td>
<td>-51.232</td>
<td>-34.854</td>
</tr>
<tr>
<td>7.5B 5/10</td>
<td>58.348</td>
<td>-44.998</td>
<td>-38.136</td>
</tr>
<tr>
<td>10B 5/10</td>
<td>60.559</td>
<td>-41.617</td>
<td>-42.737</td>
</tr>
<tr>
<td>2.5PB 5/10</td>
<td>59.351</td>
<td>-32.975</td>
<td>-44.290</td>
</tr>
<tr>
<td>5PB 5/10</td>
<td>59.588</td>
<td>-24.395</td>
<td>-44.062</td>
</tr>
<tr>
<td>7.5PB 5/10</td>
<td>59.149</td>
<td>-12.457</td>
<td>-45.732</td>
</tr>
<tr>
<td>10PB 5/10</td>
<td>59.915</td>
<td>0.135</td>
<td>-43.338</td>
</tr>
<tr>
<td>2.5P 5/10</td>
<td>60.452</td>
<td>10.046</td>
<td>-40.953</td>
</tr>
<tr>
<td>5P 5/10</td>
<td>59.897</td>
<td>18.366</td>
<td>-34.280</td>
</tr>
<tr>
<td>7.5P 5/10</td>
<td>59.933</td>
<td>32.287</td>
<td>-30.992</td>
</tr>
</tbody>
</table>
## Unique Hue Stimuli

<table>
<thead>
<tr>
<th>Munsell Sample</th>
<th>L*</th>
<th>u*</th>
<th>v*</th>
</tr>
</thead>
<tbody>
<tr>
<td>10P 5/10</td>
<td>60.095</td>
<td>44.466</td>
<td>-24.959</td>
</tr>
<tr>
<td>2.5RP 5/12</td>
<td>59.824</td>
<td>60.414</td>
<td>-24.002</td>
</tr>
<tr>
<td>5RP 5/12</td>
<td>59.388</td>
<td>72.346</td>
<td>-13.392</td>
</tr>
<tr>
<td>7.5RP 5/12</td>
<td>60.470</td>
<td>80.150</td>
<td>-6.265</td>
</tr>
<tr>
<td>10RP 5/12</td>
<td>60.220</td>
<td>85.332</td>
<td>0.104</td>
</tr>
<tr>
<td>2.5R 5/12</td>
<td>60.131</td>
<td>90.200</td>
<td>5.643</td>
</tr>
<tr>
<td>5R 5/12</td>
<td>60.435</td>
<td>94.557</td>
<td>12.076</td>
</tr>
<tr>
<td>7.5R 5/12</td>
<td>60.16</td>
<td>91.711</td>
<td>17.468</td>
</tr>
<tr>
<td>10R 5/12</td>
<td>60.095</td>
<td>91.303</td>
<td>25.082</td>
</tr>
<tr>
<td>1.25YR 7/12</td>
<td>69.000</td>
<td>90.875</td>
<td>31.414</td>
</tr>
<tr>
<td>5YR 7/12</td>
<td>79.006</td>
<td>79.988</td>
<td>44.305</td>
</tr>
<tr>
<td>7.5YR 7/12</td>
<td>79.167</td>
<td>68.745</td>
<td>48.908</td>
</tr>
<tr>
<td>10YR 7/12</td>
<td>79.873</td>
<td>55.984</td>
<td>54.845</td>
</tr>
<tr>
<td>2.5Y 8/12</td>
<td>89.219</td>
<td>43.550</td>
<td>64.580</td>
</tr>
<tr>
<td>5Y 8/12</td>
<td>88.785</td>
<td>30.298</td>
<td>67.575</td>
</tr>
<tr>
<td>7.5Y 8/12</td>
<td>88.776</td>
<td>19.530</td>
<td>68.938</td>
</tr>
<tr>
<td>10Y 8.5/12</td>
<td>93.358</td>
<td>10.955</td>
<td>72.766</td>
</tr>
<tr>
<td>2.5GY 8/12</td>
<td>89.501</td>
<td>-3.204</td>
<td>72.720</td>
</tr>
<tr>
<td>5GY 7/12</td>
<td>80.122</td>
<td>-20.611</td>
<td>65.335</td>
</tr>
<tr>
<td>7.5GY 6/12</td>
<td>71.611</td>
<td>-37.693</td>
<td>56.921</td>
</tr>
</tbody>
</table>
### Experiment IV English Color Naming Stimuli

This table (continued through p. 141) shows WCS codes and CIE LUV coordinates for samples used in the English color naming protocol in Experiment IV.

<table>
<thead>
<tr>
<th>WCS</th>
<th>L*</th>
<th>u*</th>
<th>v*</th>
<th>WCS</th>
<th>L*</th>
<th>u*</th>
<th>v*</th>
</tr>
</thead>
<tbody>
<tr>
<td>C01</td>
<td>26.503</td>
<td>69.482</td>
<td>2.466</td>
<td>D04</td>
<td>39.726</td>
<td>94.246</td>
<td>20.935</td>
</tr>
<tr>
<td>C03</td>
<td>26.503</td>
<td>84.347</td>
<td>9.701</td>
<td>D06</td>
<td>39.726</td>
<td>62.265</td>
<td>24.196</td>
</tr>
<tr>
<td>C05</td>
<td>26.503</td>
<td>52.060</td>
<td>14.951</td>
<td>D08</td>
<td>39.726</td>
<td>34.668</td>
<td>25.455</td>
</tr>
<tr>
<td>C07</td>
<td>26.503</td>
<td>32.522</td>
<td>17.297</td>
<td>D10</td>
<td>39.726</td>
<td>22.081</td>
<td>29.502</td>
</tr>
<tr>
<td>C13</td>
<td>26.503</td>
<td>-1.040</td>
<td>17.634</td>
<td>D16</td>
<td>39.726</td>
<td>-30.784</td>
<td>29.952</td>
</tr>
<tr>
<td>C19</td>
<td>26.503</td>
<td>-33.496</td>
<td>7.455</td>
<td>D22</td>
<td>39.726</td>
<td>-49.280</td>
<td>-5.401</td>
</tr>
<tr>
<td>C21</td>
<td>26.503</td>
<td>-29.416</td>
<td>0.710</td>
<td>D24</td>
<td>39.726</td>
<td>-36.367</td>
<td>-12.525</td>
</tr>
<tr>
<td>C33</td>
<td>26.503</td>
<td>8.691</td>
<td>-38.750</td>
<td>D36</td>
<td>39.726</td>
<td>41.212</td>
<td>-27.172</td>
</tr>
<tr>
<td>C37</td>
<td>26.503</td>
<td>43.699</td>
<td>-19.128</td>
<td>D40</td>
<td>39.726</td>
<td>103.895</td>
<td>-1.658</td>
</tr>
<tr>
<td>C39</td>
<td>26.503</td>
<td>59.509</td>
<td>-6.624</td>
<td>E01</td>
<td>51.526</td>
<td>119.087</td>
<td>7.703</td>
</tr>
<tr>
<td>WCS</td>
<td>L*</td>
<td>u*</td>
<td>v*</td>
<td>WCS</td>
<td>L*</td>
<td>u*</td>
<td>v*</td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>E05</td>
<td>51.526</td>
<td>104.904</td>
<td>33.694</td>
<td>F10</td>
<td>62.370</td>
<td>31.123</td>
<td>46.162</td>
</tr>
<tr>
<td>E07</td>
<td>51.526</td>
<td>66.121</td>
<td>35.602</td>
<td>F12</td>
<td>62.370</td>
<td>11.361</td>
<td>56.266</td>
</tr>
<tr>
<td>E09</td>
<td>51.526</td>
<td>39.135</td>
<td>38.971</td>
<td>F14</td>
<td>62.370</td>
<td>-17.104</td>
<td>58.175</td>
</tr>
<tr>
<td>E11</td>
<td>51.526</td>
<td>19.084</td>
<td>43.781</td>
<td>F16</td>
<td>62.370</td>
<td>-53.893</td>
<td>53.348</td>
</tr>
<tr>
<td>E13</td>
<td>51.526</td>
<td>-0.690</td>
<td>45.158</td>
<td>F18</td>
<td>62.370</td>
<td>-64.000</td>
<td>24.151</td>
</tr>
<tr>
<td>E15</td>
<td>51.526</td>
<td>-33.598</td>
<td>48.268</td>
<td>F20</td>
<td>62.370</td>
<td>-65.614</td>
<td>11.630</td>
</tr>
<tr>
<td>E17</td>
<td>51.526</td>
<td>-64.766</td>
<td>34.889</td>
<td>F22</td>
<td>62.370</td>
<td>-66.727</td>
<td>-3.881</td>
</tr>
<tr>
<td>E23</td>
<td>51.526</td>
<td>-55.093</td>
<td>-10.403</td>
<td>F28</td>
<td>62.370</td>
<td>-50.078</td>
<td>-45.496</td>
</tr>
<tr>
<td>E33</td>
<td>51.526</td>
<td>8.887</td>
<td>-44.479</td>
<td>F38</td>
<td>62.370</td>
<td>67.679</td>
<td>-10.310</td>
</tr>
<tr>
<td>E35</td>
<td>51.526</td>
<td>31.876</td>
<td>-33.165</td>
<td>F40</td>
<td>62.370</td>
<td>99.111</td>
<td>3.022</td>
</tr>
<tr>
<td>E37</td>
<td>51.526</td>
<td>66.887</td>
<td>-23.799</td>
<td>G01</td>
<td>72.508</td>
<td>70.185</td>
<td>8.158</td>
</tr>
<tr>
<td>E39</td>
<td>51.526</td>
<td>101.105</td>
<td>-6.983</td>
<td>G03</td>
<td>72.508</td>
<td>97.354</td>
<td>21.796</td>
</tr>
<tr>
<td>F02</td>
<td>62.370</td>
<td>109.213</td>
<td>15.617</td>
<td>G05</td>
<td>72.508</td>
<td>92.400</td>
<td>33.985</td>
</tr>
<tr>
<td>F04</td>
<td>62.370</td>
<td>130.698</td>
<td>32.281</td>
<td>G07</td>
<td>72.508</td>
<td>93.084</td>
<td>53.828</td>
</tr>
<tr>
<td>F06</td>
<td>62.370</td>
<td>94.721</td>
<td>40.495</td>
<td>G09</td>
<td>72.508</td>
<td>53.267</td>
<td>58.520</td>
</tr>
<tr>
<td>F08</td>
<td>62.370</td>
<td>65.415</td>
<td>49.309</td>
<td>G11</td>
<td>72.508</td>
<td>24.215</td>
<td>64.899</td>
</tr>
<tr>
<td>WCS</td>
<td>L*</td>
<td>u*</td>
<td>v*</td>
<td>WCS</td>
<td>L*</td>
<td>u*</td>
<td>v*</td>
</tr>
<tr>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>G13</td>
<td>72.508</td>
<td>-1.749</td>
<td>65.735</td>
<td>H18</td>
<td>82.095</td>
<td>-43.937</td>
<td>19.722</td>
</tr>
<tr>
<td>G15</td>
<td>72.508</td>
<td>-32.167</td>
<td>54.668</td>
<td>H20</td>
<td>82.095</td>
<td>-45.173</td>
<td>10.331</td>
</tr>
<tr>
<td>G17</td>
<td>72.508</td>
<td>-59.854</td>
<td>35.915</td>
<td>H22</td>
<td>82.095</td>
<td>-34.057</td>
<td>0.875</td>
</tr>
<tr>
<td>G19</td>
<td>72.508</td>
<td>-54.197</td>
<td>15.977</td>
<td>H24</td>
<td>82.095</td>
<td>-34.206</td>
<td>5.928</td>
</tr>
<tr>
<td>G21</td>
<td>72.508</td>
<td>-58.401</td>
<td>6.810</td>
<td>H26</td>
<td>82.095</td>
<td>-30.787</td>
<td>-12.396</td>
</tr>
<tr>
<td>G23</td>
<td>72.508</td>
<td>-46.517</td>
<td>-5.458</td>
<td>H28</td>
<td>82.095</td>
<td>-22.024</td>
<td>-16.277</td>
</tr>
<tr>
<td>G29</td>
<td>72.508</td>
<td>-33.002</td>
<td>-37.365</td>
<td>H34</td>
<td>82.095</td>
<td>7.483</td>
<td>-14.065</td>
</tr>
<tr>
<td>G31</td>
<td>72.508</td>
<td>-10.716</td>
<td>-28.228</td>
<td>H36</td>
<td>82.095</td>
<td>29.538</td>
<td>-13.909</td>
</tr>
<tr>
<td>G33</td>
<td>72.508</td>
<td>4.836</td>
<td>-25.925</td>
<td>H38</td>
<td>82.095</td>
<td>39.275</td>
<td>-3.953</td>
</tr>
<tr>
<td>G35</td>
<td>72.508</td>
<td>30.319</td>
<td>-24.418</td>
<td>H40</td>
<td>82.095</td>
<td>47.785</td>
<td>3.425</td>
</tr>
<tr>
<td>G37</td>
<td>72.508</td>
<td>56.462</td>
<td>-17.341</td>
<td>I01</td>
<td>91.235</td>
<td>12.043</td>
<td>2.174</td>
</tr>
<tr>
<td>G39</td>
<td>72.508</td>
<td>61.666</td>
<td>-0.527</td>
<td>I03</td>
<td>91.235</td>
<td>15.604</td>
<td>5.203</td>
</tr>
<tr>
<td>H02</td>
<td>82.095</td>
<td>55.509</td>
<td>10.600</td>
<td>I05</td>
<td>91.235</td>
<td>15.897</td>
<td>7.582</td>
</tr>
<tr>
<td>H04</td>
<td>82.095</td>
<td>60.969</td>
<td>20.099</td>
<td>I07</td>
<td>91.235</td>
<td>17.961</td>
<td>13.639</td>
</tr>
<tr>
<td>H06</td>
<td>82.095</td>
<td>72.919</td>
<td>37.164</td>
<td>I09</td>
<td>91.235</td>
<td>22.258</td>
<td>28.854</td>
</tr>
<tr>
<td>H08</td>
<td>82.095</td>
<td>75.754</td>
<td>63.260</td>
<td>I11</td>
<td>91.235</td>
<td>17.754</td>
<td>47.775</td>
</tr>
<tr>
<td>H10</td>
<td>82.095</td>
<td>39.620</td>
<td>66.928</td>
<td>I13</td>
<td>91.235</td>
<td>0.265</td>
<td>48.863</td>
</tr>
<tr>
<td>H12</td>
<td>82.095</td>
<td>16.318</td>
<td>72.538</td>
<td>I15</td>
<td>91.235</td>
<td>-5.275</td>
<td>18.334</td>
</tr>
<tr>
<td>H14</td>
<td>82.095</td>
<td>-10.332</td>
<td>58.964</td>
<td>I17</td>
<td>91.235</td>
<td>-13.301</td>
<td>12.306</td>
</tr>
<tr>
<td>H16</td>
<td>82.095</td>
<td>-39.359</td>
<td>41.151</td>
<td>I19</td>
<td>91.235</td>
<td>-16.369</td>
<td>7.300</td>
</tr>
</tbody>
</table>
**Experiment IV English Color Naming Stimuli (ctd.)**

<table>
<thead>
<tr>
<th>WCS</th>
<th>L*</th>
<th>u*</th>
<th>v*</th>
</tr>
</thead>
<tbody>
<tr>
<td>I21</td>
<td>91.235</td>
<td>-19.277</td>
<td>3.626</td>
</tr>
<tr>
<td>I23</td>
<td>91.235</td>
<td>-18.668</td>
<td>-0.068</td>
</tr>
<tr>
<td>I25</td>
<td>91.235</td>
<td>-16.185</td>
<td>-2.805</td>
</tr>
<tr>
<td>I27</td>
<td>91.235</td>
<td>-9.745</td>
<td>-4.631</td>
</tr>
<tr>
<td>I29</td>
<td>91.235</td>
<td>-5.534</td>
<td>-5.301</td>
</tr>
<tr>
<td>I31</td>
<td>91.235</td>
<td>-2.109</td>
<td>-7.314</td>
</tr>
<tr>
<td>I33</td>
<td>91.235</td>
<td>3.980</td>
<td>-7.261</td>
</tr>
<tr>
<td>I35</td>
<td>91.235</td>
<td>6.867</td>
<td>-3.617</td>
</tr>
<tr>
<td>I37</td>
<td>91.235</td>
<td>9.203</td>
<td>-1.789</td>
</tr>
<tr>
<td>I39</td>
<td>91.235</td>
<td>13.417</td>
<td>1.282</td>
</tr>
<tr>
<td>A00</td>
<td>-12.607</td>
<td>0.935</td>
<td>0.524</td>
</tr>
<tr>
<td>C00</td>
<td>26.503</td>
<td>-0.491</td>
<td>-0.281</td>
</tr>
<tr>
<td>E00</td>
<td>51.526</td>
<td>-0.470</td>
<td>0.063</td>
</tr>
<tr>
<td>G00</td>
<td>72.508</td>
<td>-0.232</td>
<td>-0.679</td>
</tr>
<tr>
<td>J00</td>
<td>100.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*Note: L* values calculated, not measured.*
Appendix C: Details of Analysis
Scree Plot Generation and Interpretation

A scree plot is a visualization of MDS stress versus dimensionality, commonly used to determine the minimum number of dimensions needed to account for most of the variability in a non-metric MDS solution. In order to generate a scree plot for a given dissimilarity data set, MDS solutions in one to $n$ dimensions are calculated, and these solutions’ respective stress values are plotted along the y-axis, with number of dimensions plotted along the x-axis. Though stress by definition decreases monotonically with number of dimensions, a solution which is well-modeled in $m$ dimensions will improve noticeably less from $m:m + 1$ dimensions than from $m – 1:m$ dimensions. This effect is visible as an “elbow” in the plotted data (Wickelmaier, 2003). However, caution must be exercised when interpreting scree plots—oftentimes, there is no obvious elbow in the plot, or there are multiple elbows. Additionally, some solutions show “good” fit (stress < 0.1) to two dimensions, but an elbow at the third dimension or higher. In this study, good fit is prioritized over the scree elbow, and scree plots are used more for the interpretation of solutions which fit poorly in two dimensions.