Vision Research 151 (2018) 99–116

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Modeling color preference using color space metrics

Karen B. Schloss [a,b,*], Laurent Lessard [b,c], Chris Racey [a,b], Anya C. Hurlbert [d]

[a] Department of Psychology, University of Wisconsin–Madison, 1202 West Johnson Street, Madison, WI 53706, USA
[b] Wisconsin Institute for Discovery, University of Wisconsin–Madison, and Wisconsin Institutes for Discovery, 330 N. Orchard St., Madison, WI 53715, USA
[c] Department of Electrical and Computer Engineering, University of Wisconsin–Madison, 1415 Engineering Dr., Madison, WI 53706, USA
[d] Institute of Neuroscience, Newcastle University, Newcastle upon Tyne NE2 4HH, UK

ARTICLE INFO

Article history:
Received 21 February 2017
Received in revised form 25 June 2017
Accepted 3 July 2017
Available online 28 July 2017

Number of Reviews = 2

ABSTRACT

Studying color preferences provides a means to discover how perceptual experiences map onto cognitive and affective judgments. A challenge is finding a parsimonious way to describe and predict patterns of color preferences, which are complex with rich individual differences. One approach has been to model color preferences using factors from metric color spaces to establish direct correspondences between dimensions of color and preference. Prior work established that substantial, but not all, variance in color preferences could be captured by weights on color space dimensions using multiple linear regression. The question we address here is whether model fits may be improved by using different color metric specifications. We therefore conducted a large-scale analysis of color space models, and focused in-depth analysis on models that differed in color space (cone-contrast vs. CIELAB), coordinate system within the color space (Cartesian vs. cylindrical), and factor degrees (1st degree only, or 1st and 2nd degree). We used k-fold cross validation to avoid over-fitting the data and to ensure fair comparisons across models. The best model was the 2nd-harmonic Lch model ("LabC Cyl2"). Specified in CIELAB space, it included 1st and 2nd harmonics of hue (capturing opponency in hue preferences and simultaneous liking/disliking of both hues on an opponent axis, respectively), lightness, and chroma. These modeling approaches can be used to characterize and compare patterns for group averages and individuals in future datasets on color preference, or other measures in which correspondences between color appearance and cognitive or affective judgments may exist.

1. Introduction

Central goals in the study of color cognition are to understand how the perceptual experience of color maps onto cognitive and emotional judgments about color, and to understand how these judgments influence people's beliefs and behaviors. The study of color preference provides a direct route to this goal. The results of previous studies suggest that there are systematic mappings between color appearance and preference (Guilford & Smith, 1959; Hurlbert & Ling, 2007; McManus, Jones, & Cottrell, 1981; Ou, Luo, Woodcock, & Wright, 2004; Palmer & Schloss, 2010). For example, hue preferences in industrialized cultures tend to vary along a blueness-yellowness axis, with a peak at blue and a trough around yellowish-green. Despite this robust pattern in average preference data, there are large individual differences (for reviews, see Hurlbert & Owen, 2015; Schloss & Palmer, 2015), the origins of which are not entirely understood and have yet to be captured fully through a predictive model.

The aim of this study is to determine the best quantitative model for describing and predicting color preference patterns based on color appearance alone. By "best", we mean the model having the smallest number of input factors that captures the largest amount of variance in preference across the individuals being studied. The input factors are related to the specification of the color stimulus only, and do not include any characteristics of the individual. Thus, the aim is not to probe the causal origins of individual differences in color preference (e.g., as in Schloss, Hawthorne-Madell, & Palmer, 2015), but instead to describe and predict color preference from a specification of color appearance alone. In this sense, the rationale follows earlier attempts to demonstrate that the "affective value" of a color may be predicted systematically from accurate "color-specification" (Guilford & Smith, 1959). Such a model also provides a parsimonious way to characterize preference patterns across populations, through

* Corresponding author at: University of Wisconsin–Madison, Department of Psychology and Wisconsin Institutes for Discovery, Brogden Hall, 1202 West Johnson Street, Madison, WI 53706, USA.
E-mail address: kschloss@wisc.edu (K.B. Schloss).

http://dx.doi.org/10.1016/j.visres.2017.07.001
0042-6989/

weights on the input factors (Hurlbert & Ling, 2007), provided the model fits the data well.

Given that multiple color spaces and color ordering systems exist for specifying color appearance, a first task in building a model is choosing the color space in which the dimensions of appearance are specified. These dimensions ultimately should emerge as the ones that map best to variations in preference. Although the causal origins of preferences are not addressed by the model, the results may provide a deeper understanding of how and at which stage of visual processing color preferences are embedded. The second task is to determine the exact form of the quantitative relationship between the color dimensions and preference, e.g. whether it is linear or nonlinear. We approached these tasks by conducting a large-scale analysis of multiple candidate color spaces and different coordinate representations within them, to determine which were most effective at capturing variations in color preferences. We evaluated the models using two previously obtained datasets from different countries. A practical outcome of this work is a set of tools for building models that compactly describe and predict variations in color preference patterns across large populations and individuals.

1.1. Considerations in constructing models of color preferences based on color appearance

For a model based on color appearance to be useful in describing and predicting color preferences, it should satisfy at least two criteria: (1) the dimensions of the color space used should capture (and allow parameterization of) all possible variations in color appearance and (2) variations in at least some of these dimensions should elicit variations in color preference. For example, suppose a model includes no dimension that captured variations in lightness. It might fit color sets dominated by variations in hue and saturation but not those that have variations in lightness. The first criterion would be violated because it fails to capture all variations in color appearance. Further, the second criterion would be violated if color preferences for some datasets varied only along the lightness dimension (not in hue or saturation) because the model has no dimensions to capture that variability. To these points, models based on hue only were successful at characterizing individual preferences for colors which varied only in hue (Hurlbert & Ling, 2007), but required augmentation by factors encoding saturation and lightness to characterize preferences for colors that varied in hue, saturation, and lightness (Ling & Hurlbert, 2007).

Models that meet the above criteria may nonetheless differ in their ability to account for variations in color preference. These differences arise from issues around the choice of color appearance metric and the derivation of the quantitative relationships between appearance dimensions and preference, as we describe in more detail below.

1.1.1. Color spaces

Color appearance may be specified in a variety of color spaces or color ordering systems, for example the CIE 1931 standard tristimulus space (CIE, 2004) or the Munsell notation system (Munsell, 1921). A key question is whether the choice of color space affects the ability of the model to fit the color preference data. It is beyond the scope of this paper to review systematically all existing color spaces and systems, but it is important to note that the differences in their origins, standardizations, and uses make some spaces more amenable to modeling preference quantitatively than others.

Generally speaking, we make a distinction between color spaces defined initially in terms of the CIE standard colorimetric observer's color-matching functions ("standardized") vs. those based on human cone photoreceptor activations ("physiological"). Both types of color space are derived from color discrimination or matching judgments made by human observers and specify color appearance in terms of three basic descriptors. Here, for standardized appearance spaces, we focus on the near perceptually-uniform color spaces CIELAB and CIELUV[1], which are based on data from perceptual color-difference measurements of standardized color samples and which explicitly define descriptors corresponding to the perceptual attributes of hue, chroma, and lightness (Kuehni & Schwarz, 2008; Wyszecki & Stiles, 1982). CIELAB space is widely used to specify color stimuli in visual psychophysics studies, and has specifically been used for relating preference to color appearance. Ou et al. (2004), for example, found that 70% of the variance in average preference across a set of 20 colors could be accounted for by a nonlinear function of the L^*, a^*, and b^* color coordinates.

"Physiological" color spaces based directly on cone photoreceptor activations are derived from neurobiological or psychophysical measurements of color discrimination, and their dimensions are more readily related to early stages of visual processing (Derrington, Krauskopf, & Lennie, 1984; Eskew, McLellan, & Giulianini, 1999). For example, cone-opponent contrast spaces (Derrington et al., 1984; Eskew et al., 1999), which we examine here, quantify the appearance of a test color in terms of its contrast against a uniform adapting background along three cardinal directions, defined by transformations of the L, M, and S cone photoreceptor activations, $[S-(L+M)]$, $[L-M]$ and $[L+M]$, sometimes referred to as "blue-yellow", "red-green", and "luminance" mechanisms. These dimensions are generally equated with the second-stage of color encoding in the human visual system and thought to be represented by neurons at early stages in the visual pathway (Lennie & Movshon, 2005). For brevity, we subsequently refer to $[S-(L+M)]$ as "$S-LM$", $[L-M]$ as "LM", and $[L+M]$ as "Lum".

In developing a method for describing individual differences in color preference, Hurlbert and Ling (2007) demonstrated the effectiveness of these cone-opponent dimensions in capturing preference variations. Specifically, they found that individuals' preferences for eight colors, which varied only in hue (iso-luminance, iso-saturation) could be decomposed into two principal components that matched the $[S-(L+M)]$ and $[L-M]$ cone-opponent contrast axes. A regression model captured 70% of the variation in individual color preference judgments with the two cone-contrast predictors. Therefore, preference patterns of individual participants could be parsimoniously characterized by personalized weights on these cone-contrast axes.

Hurlbert and Ling's (2007) results suggest that when colors vary primarily in hue, color appearance that is encoded by the physiological second-stage chromatic mechanisms is sufficient to account for variability in color preferences. Both model criteria are satisfied for this specific dataset: (1) the dimensions of the space adequately capture variations in color appearance and (2) variations in the dimensions elicit variations in color preferences. However, this two-component cone-contrast model should be insufficient for describing preference patterns for colors that vary in saturation and lightness in addition to hue because the first criterion will be violated. When colors vary in hue, saturation, and lightness, more dimensions are required to represent them.[2]

[1] These (approximately) perceptually uniform color spaces are based on initial specifications of colors in terms of the CIE XYZ tristimulus coordinate system, which is based on standardized measurements of light sources and color-matching functions of a standard observer. The tristimulus coordinates are thus related by linear transformations to receptor spectral sensitivities of the average trichromatic observer, but do not themselves form a uniform space or directly represent the perceptual attributes of color.

[2] Although two dimensions are in theory adequate to encode both hue and saturation (corresponding to angle and radius in the chromaticity plane), in practice, the two dimensions uncovered in the cone-opponent contrast model are indistinguishable from hue-angle-encoding dimensions, since for iso-saturation stimuli, the radii do not vary.

Ling and Hurlbert (2007) addressed this limitation by adding two components to the cone-contrast model: lightness (or luminance contrast, in cone-contrast space) and saturation (s_{uv} in CIELUV space). They tested this model on average and individual UK participants' preferences for different sets of colors that varied in hue, saturation, and lightness. Depending on the color set and task, the extended cone contrast model accounted for 47–74% of the variance in average color preferences and an average of 46–61% of the variance in individual participants' color preferences. The greater end of that range was for color sets that varied less in saturation and lightness. For preference datasets from US participants using color sets that had more extreme values of saturation, the model explained far less variance: 37% in average color preferences (Palmer & Schloss, 2010) and an average of 39% in individual participants' color preferences (Schloss et al., 2015).

This reduction in performance of the extended cone-contrast model for broader stimulus sets suggests that a different set of color appearance metrics might be more effective than second-stage encoding mechanisms in capturing variations in preference. To address this question, Palmer and Schloss (2010) tested how well preferences were accounted for by participants' subjective ratings of yellowness vs. blueness, redness vs. greenness, lightness, and saturation of each color. These subjective color appearance dimensions accounted for more variance in group average color preferences (60%) than did the extended cone-contrast model (37%) (Palmer & Schloss, 2010), and also performed better in describing individual preferences (Schloss et al., 2015). Further, Sorokowski, Sorokowska, and Witzel (2014) found that gender differences in color preference were better modeled by a "red-versus-blue" component (calculated from hue similarity with selected red and blue colors, ignoring variations in lightness) than by cone-opponent contrast axes, which suggests that higher-level categorical representations might better characterize color preferences than low-level dimensions.

It is also clear in some preference datasets that there are substantial interactions between hue, saturation and lightness. For example, there is a substantial effect of lightness for "warm" hues (dark oranges and dark yellows are especially disliked, whereas very light oranges and yellow are liked) (Palmer & Schloss, 2010; Schloss et al., 2015; Taylor & Franklin, 2012; Yokosawa, Schloss, Asano, & Palmer, 2016). The fact that this striking shift in preference occurs with a change in categorization of the color from brownish to pure yellow suggests that a later-stage neural encoding model, which better predicts categorical color appearance, may be necessary to capture variations in preference. Taken together, these results suggest that color space models that characterize colors using higher stages in visual processing might be more effective at accounting for color preferences.

1.1.2. Coordinate systems

A second consideration is the choice of coordinate system within the color space. A particular location in color space may be specified in Cartesian coordinates (e.g. distance along x and y axes) or cylindrical coordinates (e.g. radius and angle with respect to the origin). These coordinates may have different perceptual correlates. For example, in the CIELAB chromaticity plane, specifying a color's location by its a^* and b^* coordinates corresponds (roughly) to specifying its redness/greenness and blueness/yellowness, respectively, relative to the color at the origin (the "neutral" color). The cylindrical coordinates of radius ($\sqrt{a^2+b^2}$) and angle ($\mathrm{atan2}(b,a)$) are defined as the approximate correlates of the perceptual attributes of chroma[3] and hue[4]. CIELAB L^* is the correlate of lightness, the same in both the Cartesian and cylindrical coordinate systems.

A key difference between the two coordinate systems is that hue and saturation are coupled in the Cartesian system but decoupled in the cylindrical systems. In the cylindrical representation, variations in angle encode variations of hue independent of chroma, and variations of radius encode variations in chroma (or saturation, at constant lightness) independent of hue. Previous models of color preference have represented both hue and saturation by mixing Cartesian and cylindrical coordinates, e.g., with the addition of s_{uv} in the extended cone contrast model (Ling & Hurlbert, 2007) and the use of saturation ratings in the color appearance model (Palmer & Schloss, 2010). Adding a separate factor for saturation provides a way to capture variations of preference with saturation. However, in both examples, hue is encoded by two Cartesian coordinates corresponding roughly to redness-greenness and blueness-yellowness, which are not independent of saturation, unless their values are normalized to the unit circle.

Both sets of coordinates may be used in any color space. That is, for a color vector in any color space, its hue may be defined as the angle with respect to the origin, and its saturation (or lightness-normalized chroma) as the magnitude. The extent to which hue, saturation, and lightness coordinates defined in this way in any particular color space actually correspond to the perceptual attributes of hue, saturation and lightness, though, is not completely understood, and has been investigated by color appearance studies. For example, phenomena such as the Abney effect describe the deviation of perceived hue from the radial hue line in the CIE chromaticity plane (Wyszecki & Stiles, 1982), and a recent study systematically examines how well seven different measures of saturation predict perceived saturation in natural images (Schiller & Gegenfurtner, 2016).

It remains an open question as to where and how neuronal activity in the visual system encodes these perceptual attributes, and to what extent they interact in neuronal representations, although recent findings suggest that representations differ between low- and high-level stages of visual processing (Bohon, Hermann, Hansen, & Conway, 2016).

1.1.3. Factor degrees

Within a given color space and coordinate system, different functions of the coordinates may be inputs to a quantitative model. Linear functions use 1st degree factors only, whereas quadratic functions use 2nd degree factors in addition to 1st degree factors. In Cartesian coordinates, if the 1st degree factors are x and y, then the 2nd degree factors are x^2, y^2, and xy. In cylindrical coordinates, we defined the 1st degree factor as the 1st harmonic of the hue angle (with a period of 360°) and the 2nd degree as the 2nd harmonic (with a period of 180°).

The interpretation and effect of different degrees of factor depends on the color space and coordinates. For example, in a color space with color-opponent axes (such as CIELAB or cone-contrast space, as opposed to RGB space), a positive weighting on one axis means the observer likes hues at its positive pole and dislikes hues at its negative pole. Using 1st degree factors only, the model can capture greater preference for one pole than the other, but it cannot capture a simultaneous liking or disliking for both poles of the color-opponent axis. Earlier models using 1st degree factors only in color-opponent spaces (Hurlbert & Ling, 2007; Ling & Hurlbert, 2007; Ou et al., 2004; Palmer & Schloss, 2010; Sorokowski et al., 2014) effectively assumed that components underlying color preferences operate in a hue opponent nature, which is not necessarily true (Bimler, Brunt, Lanning, & Bonnardel, 2014; Schloss & Palmer, 2017). It is not necessary for individuals who like red to also dislike green; they can like (or dislike) both red and green. Further, the same model weight of zero on an opponent axis may result

[3] Chroma (C^*) is related to saturation (s) by the transformation $C^*/L^* = s$.

[4] In these equations and references to L^*, a^* and b^* axes in subsequent sections, we use L, a, and b, dropping the asterisk to avoid clutter when exponents are used.

from drastically different patterns of color preferences: strong liking of both endpoints of the dimension, strong disliking of both endpoints of the dimension, or total indifference to that dimension. With 2nd degree factors, however, a model is able to represent non-opponent preferences for opponent hues on one dimension. This observation is a key motivation for our use of 2nd degree factors in the models we test.

Other evidence supports this reasoning. Using Principal Components Analysis, Bimler et al. (2014) found there were four components to hue preference, two of which were hue opponent (one peak and one trough), but two of which were not hue opponent (two peaks and two troughs). As Bimler et al. (2014) noted, double-peaked components can account for patterns in which people have the same preference valence for both poles of a color-opponent axis (i.e. are both liked or disliked). This result is consistent with early evidence that hue preferences can be fit by a weighted combination of multiple harmonics, in which the first harmonic (one peak/one trough) and second harmonic (two peaks/two troughs) accounted for 75% of the variation (Stamm, 1955). Therefore, models that contain 1st and 2nd degree factors will likely be more effective than models that contain 1st degree factors alone.

1.2. On the distinction between describing, predicting, and explaining color preferences

When modeling color preferences it is important to consider what kinds of conclusions can be drawn from the model results. The results of the same kind of statistical model, such as multiple linear regression (MLR) used here, can support different kinds of conclusions depending on the types of factors that are used. We make a distinction between three core goals: describing, predicting, and explaining patterns of color preference data. The terms "describe", "predict", and "explain" are often used interchangeably in the literature when discussing how well a model fits a color preference dataset. One might say that a regression model *describes x%*, *explains x%*, or *predicts x%* of the variance in color preferences. It is reasonable to use these terms interchangeably when discussing statistics, but it can obfuscate important theoretical distinctions about the kinds of conclusions that can be drawn from the results. Here we clarify those distinctions.

Describing color preferences involves characterizing a pattern in observed data (e.g., on average, people like blues more than yellows). Descriptions are useful for communicating patterns in color preference data and determining what patterns need to be explained by theories. However, descriptions are not theories in and of themselves. For example, the description that Japanese participants like lighter colors more than US participants do (Yokosawa et al., 2016) highlights a difference but it does not explain why that difference exists.

Predicting color preferences involves using information about how much people like colors observed by the model to predict preferences for colors unobserved by the model. The goal is to accurately anticipate future judgments given knowledge about prior judgments. The desire is to be as accurate as possible, even if the dimensions are not interpretable. Predicting and describing color preferences can have a symbiotic relationship. For example, one might use a complex associative model with hidden layers for the purpose of predicting or interpolating preferences, but use a simpler, qualitative model for characterizing the nature of the predictions.

Explanations of color preferences define possible causal accounts that answer *how* and *why* questions such as: How are color preferences formed? Why do they exist? Why do they differ between individuals? Why do they change over time? These *how* and *why* questions are fundamentally different from the *what* questions involved in describing and predicting color preferences (e.g., *what* colors do people like?). Explanations are often tested with the same kinds of regression models used to describe color preferences, but the difference lies in the nature of the factors that go into the model. For example, Palmer and Schloss (2010) used linear regression to account for color preferences using different kinds of models. Their Weighted Affective Valence Estimate (WAVE) model tested the hypothesis that color preferences could be *explained* by how much people like objects that are associated with the colors (Ecological Valence Theory; EVT)[5]. Their color appearance model evaluated how well color preferences could be described by dimensions of color appearance (yellow/blue, red/green, light/dark, saturation). The WAVE model provides support for a theoretical account that explains how color preferences are formed, whereas the color appearance model provides a useful description of the data without explaining why color preferences map onto color appearance axes in the way they do.

1.3. Current approach: describing and predicting but not explaining

The goals of this study are to understand which color space representations are most effective at describing observed patterns of color preferences and predicting preferences for colors previously unseen by the model. We conducted new analyses of data from previous preference studies (Ling & Hurlbert, 2007; Schloss et al., 2015) to determine which color space models are most effective, with the fewest number of factors. We compare models that characterize colors in different color spaces likely to correspond to different stages of visual processing. We also introduce the use of cylindrical coordinates and periodic regression to better capture the perceptual dimensions of hue, chroma, and lightness. By doing so, we are able to systematically evaluate how color space, coordinate system, and factor degree influence the ability of color space models to fit color preference data.

For each model, we use multiple linear regressions to determine weights on each of the factors for a given group or individual for a given set of colors. The accuracy of the model in *describing* the observed data is quantified as the fit (R^2) between the weighted combination of factors and the original data. The accuracy of the model in *predicting* new data is quantified as R^2 between the model predictions for untrained colors (using the weights from the regression equation calculated from the trained colors) and preferences for those untrained colors.

Our scope excludes reviewing and evaluating evidence for factors that influence individual variations, including culture (Choungourian, 1968; Reddy & Bennett, 1985; Saito, 1981, 1996; Yokosawa et al., 2016), sex (Bimler et al., 2014; Eysenck, 1941; Helson & Lansford, 1970; McManus et al., 1981), age (Adams, 1987; Dittmar, 2011; Pereverzeva & Teller, 2004), and ecological associations (Schloss, Strauss, & Palmer, 2013; Schloss et al., 2015), (for reviews, see Bimler et al., 2014; Hurlbert & Owen, 2015; Schloss & Palmer, 2015). Instead, we focus on determining which models are more/less effective for capturing such individual variations, whatever their cause.

Our approach precludes using color preference models that may have strong explanatory value but are methodologically cumbersome to implement. For example, the WAVE model helps explain how color preferences are formed, why they differ between individuals, and why they change over time (Palmer & Schloss, 2010; Schloss & Palmer, 2017). However, using the

[5] Palmer and Schloss (2010) argued how causal directions might be inferred from this correlational analysis but the evidence provided was still correlational. Further evidence from experimental manipulations support the causal claim that object preferences cause color preferences (Strauss, Schloss, & Palmer, 2013).

WAVE model to predict preferences for new, untested colors is methodologically expensive[6] (Schloss et al., 2015). It would be faster and easier to ask people to simply rate their preference for many new colors than to collect the data to form the model predictions. Although the models we studied do not provide explanatory accounts, they are powerful tools for describing and predicting color preferences. Moreover, they are straightforward to implement because they are derived from factors that are predefined in standard color spaces.

Our approach, like that of some other studies (e.g. Hurlbert & Ling, 2007), goes beyond earlier attempts to derive quantitative relationships between color appearance coordinates and color preference (e.g. Ou et al., 2004) in doing so not only for average data from a single population, but also for individual data within different populations. We go beyond earlier analyses of individual variations in color preference (e.g. Hurlbert & Ling, 2007; Ling & Hurlbert, 2007; Schloss et al., 2015; Bimler et al., 2014) by explicitly testing models built from data on subsets of colors, using the technique of k-fold cross-validation to predict individuals' preferences for previously unseen colors. By using cross-validation, we are able to evaluate the performance of models with more parameters than were previously used without concern about over-fitting the data.

2. Methods

In this section, we first describe the color preference datasets that were used for testing the models (Section 2.1). We then describe the models we evaluated closely in this study (Section 2.2), which are a representative subset of 40 color space-based models described in the Supplementary Material. Finally, we describe our approach to evaluating the models using k-fold cross-validation (Section 2.3).

2.1. Description of color preference datasets

We evaluated preference models using two independent datasets, Schloss et al.'s (2015) dataset from the Berkeley Color Project (32 colors; "BCP-32") and Ling and Hurlbert's (2007) Newcastle University dataset (126 colors; "NCL-126"). We chose these two datasets because they have been previously used to study individual differences in color preferences and contain data from different populations (US vs. UK) on different sets of colors, as described below.

2.1.1. BCP-32 color preference dataset

The methods for obtaining the color preference data for the BCP-32 colors used here are described in Schloss et al. (2015), and summarized as follows.

2.1.1.1. Colors. The BCP-32 colors contain eight hues (red, orange, yellow, chartreuse, green, cyan, blue, and purple) sampled at four saturation/lightness levels (saturated, light, muted, and dark) (see Table S8 in the Supplementary Material for coordinates in CIE 1931 xyY space and in Munsell space). The saturated set contained the most saturated color of each hue that could be produced on Palmer and Schloss's (2010) display. The Munsell value and chroma of the light, muted and dark colors were scaled with respect to the value and chroma of the saturated colors of each hue. The light, muted, and dark colors within each Munsell hue differed in value but had equal Munsell chroma. The colors were presented on a neutral gray background that approximated CIE Illuminant C (CIE x = 0.312, y = 0.318, Y = 19.26). To convert to cone-contrast coordinates here, we use the background color as the neutral point. To convert to CIELAB and CIELUV, we use a white point with the same chromaticity as the background but with a luminance of Y = 116 cd/m^2.[7]

2.1.1.2. Data collection. As described in Schloss et al. (2015) there were 48 participants (24 females) with normal trichromatic color vision. During the experiment they were presented with each color one at time. They rated how much they liked each color on a line-mark slider scale from "not at all" to "very much" by sliding the cursor along the scale and clicking to record their response. Participants judged each color four times divided over two testing sessions (four blocks) in a blocked randomized design. The colors were displayed as a small square (100 × 100 pixel) centered on the screen of a 20" iMac LCD monitor (1680 × 1050 pixel resolution at 60 Hz resolution). Trials lasted until participants made their response, and the next trial began 500 ms later. The response data were scaled to range from −100 to +100. Before beginning the experiment participants completed an anchoring task so they knew what liking "not at all" and "very much" meant for them in the context of these colors (see Schloss et al., 2015 for details).

2.1.2. NCL-126 color preference dataset

The methods for obtaining the color preference data for the NCL-126 colors used here are described in Ling and Hurlbert (2007), and summarized below.

2.1.2.1. Colors. The NCL-126 chromatic colors contain three subsets of colors: the Munsell set, the NCS set and the CIELUV set (called Group 1, Group 2, and Group 3 in Ling and Hurlbert (2007), respectively). See Table S9 in the Supplementary Material for coordinates in CIE 1931 xyY. The 85 colors in the Munsell set were chosen to include 10 Munsell hues (R, YR, Y, GY, G, BG, B, PB, P, RP), sampled at three value levels (3, 5, and 7), and three chroma levels (2, 6, and 8), excluding those out of gamut for the display. For each of the ten hues, the set also included the highest value/chroma combination displayable. The 17 colors of the NCS set were the chromatic colors from Ou et al. (2004). The 24 colors in the CIELUV set included 8 hues at different saturation and lightness levels in CIELUV coordinates, including eight hues at the same saturation and lightness as in the stimulus set of Hurlbert and Ling (2007).[8] The NCL-126 colors were presented on a neutral gray background (CIE x = 0.3127, y = 0.329, Y = 50 cd/m^2). To convert to cone-contrast coordinates here, we use the background color as the neutral point. To convert to CIELAB and CIELUV we use a white point with the same chromaticity as the background, and a luminance of Y = 110 cd/m^2.

2.1.2.2. Data collection. The NCL-126 color preference dataset modeled here is a subset of Ling and Hurlbert's (2007) dataset, including all of the "Task 1" ratings data for 126 chromatic colors for 40 participants, 20 males (mean age 20.8 yrs.) and 20 females (mean age 19.2 yrs.). All participants had normal trichromatic

[6] The WAVE for a color represents the mean valence of all objects associated with the target color, weighted by how well the color of the object matches the target color. Calculating the WAVE for a color for a given individual or group of participants involves obtaining a representative set of objects associated with that color, the valence of each object for that individual/group, and a metric of how well the color of each object matches the target color.

[7] Participants also rated their preferences for 5 achromatic colors, including black, white, and three intermediate grays. The luminance of the white was 116 cd/m^2.

[8] The full stimulus set included 8 achromatic colors from the Munsell and NCS systems, which we do not analyze here because they are at the origin of the chromaticity plane and do not have hue angles to input into cylindrical models.

vision. Participants were presented with rectangular color patches ($2° \times 3°$ in size) one at a time in a randomized order. They were asked to rate each color in terms of "liked-ness" by using a horizontal slider bar scaled from "Dislike" on its left end to "Like" on its right end.[9] Participants rated each color twice in separate sessions and responses were averaged over testing sessions.

2.2. Description of the models

We began with an exploratory examination of 40 different MLR models to evaluate which combinations of color metric factors were most effective at describing and predicting color preferences. We varied the color space, the coordinate system used to represent colors within that color space, and the number of degrees in each factor. A model with three factors (f_1, f_2, and f_3), for example, would use a regression equation with weights on each factor (w_1, w_2, and w_3) of the form:

$$\text{preference} = k + w_1 f_1 + w_2 f_2 + w_3 f_3 \quad (1)$$

The factors used in the various models were computed from color dimensions in standard color spaces. Details about the full set of 40 models can be found in the Supplementary Material. The models are based on five representative color spaces, including standardized and physiological spaces, with varying color dimensions, coordinate systems, and factor degrees. Table S1 lists the color dimensions used and their spaces of origin. Tables S2–S7 describe the factors within each model, which can be broken down into the following categories: basic models, augmented models, quadratic models, unipolar models, categorical models, and cylindrical models.

For each of the 40 models in the exploratory analyses, we calculated its ability to fit existing preference data and predict untrained data, using the techniques described in detail below (Section 2.3). Based on the pattern of model fits (Figs. S1–S4 in the Supplementary Material), we eliminated some models from further in-depth study, as described in Section 2.2.1. For example, a model corresponding to higher-order categorical processing of color was removed from further study because it required more factors to perform as well as the other selected models. Therefore, to systematically investigate the effects of key model features, we focused on a subset of eight models that fit into the following orthogonal design: 2 Color Spaces × 2 Coordinate Systems × 2 Factor Degrees (Table 1). In particular, this allowed us to compare representatives of standardized and physiological color spaces.

2.2.1. Color spaces and dimensions

In the preliminary analyses (see Supplementary Material), we observed that models built in the two standardized perceptually uniform spaces, CIELUV and CIELAB, performed similarly to each other and better than those in CIEXYZ space, for similar levels of complexity. Therefore, we selected only one perceptually uniform space as representative of standardized spaces (CIELAB), which is the most widely used standardized perceptually near-uniform color space. Several versions of cone-opponent contrast spaces exist; ("DKL" space, Derrington et al., 1984 and "ESK" space, Eskew et al., 1999); we selected the latter as the representative of a physiological space for further investigation. In each of the two representative color spaces, we defined four dimensions: 2 hue, 1 lightness, and 1 saturation/chroma.

The cone-contrast "CCLS" models thus include the cone-opponent and luminance factors from cone-contrast space (LM, $S - LM$, and $L + M$), together with saturation (s_{uv}) from CIELUV space. The CCLS model in Cartesian coordinates with 1st degree factors corresponds to Ling and Hurlbert's (2007) extended cone-contrast model, which has been studied extensively in the literature. Therefore, we chose this color space for closer examination in order to enable direct comparisons to prior research. The "LabC" models include L, a, b and chroma (C_{ab}) from CIELAB space.

The "LabC" dimensions also correlate well with the dimensions in Palmer and Schloss's (2010) subjective color appearance model: for the BCP-32 colors, light-dark ratings correlate with L ($r = 0.96$, $p < 0.001$), red-green ratings with a ($r = 0.89$, $p < 0.001$), yellow-blue ratings with b ($r = 0.88$, $p < 0.001$), and saturation ratings with C_{ab} ($r = 0.76$, $p < 0.001$). Thus, the LabC factors may provide an approximation to the subjective appearance model but are methodologically less expensive to use because they do not require individual participant judgments for each color.

2.2.2. Coordinate systems

The coordinate systems we tested were Cartesian and cylindrical systems, which are different ways of specifying the same colors within cone-contrast and CIELAB spaces. The dimensions coding for lightness and saturation/chroma were the same in the models for both systems but the dimensions coding hue differed. In the Cartesian system, hue was represented in terms of x and y axes, which were LM and $S - LM$ in CCLS models and a and b in LabC models. As mentioned in the introduction, hue is confounded with saturation/chroma in this representation, which is why previous studies using Cartesian representations of hue had separate factors coding for saturation/chroma (Ling & Hurlbert, 2007; Palmer & Schloss, 2010). Although this additional information is redundant in the sense that it can be derived from existing basic factors (e.g. chroma in CIELAB coordinates satisfies $C_{ab} = \sqrt{a^2 + b^2}$), it might nevertheless improve the performance of MLR because the dependence between dimensions is nonlinear.

In the cylindrical system, hue was represented in terms of $\cos(h)$ and $\sin(h)$, where h is hue angle in cone-contrast space or CIELAB space. In cone-contrast space, we calculate h as $h_{cc} = \text{atan2}(S - LM, LM)$. In CIELAB space h is $h_{ab} = \text{atan2}(b, a)$ (i.e., h in CIE Lch space). We used the convention $0° \leqslant h < 360°$. Given that hue is a periodic quantity, hue angle cannot be used directly as a factor in MLR because assigning it a nonzero weight would lead to different predictions for 0° and 360°, even though those hues are equivalent. Therefore, we used periodic factors in our regression analyses, which is a standard way of addressing this issue of periodicity in linear modeling.

Using the sine and cosine of hue as separate factors is standard in Fourier decomposition. The additional degree of freedom afforded by using both sine and cosine yields a model that is rotation-invariant; the same fit is produced regardless of how the reference hue angle of zero is assigned. The weights on the sine and cosine factors (w_1, w_2) can be interpreted as weights on two principal hue-opponent mechanisms in the chosen color space, assuming the space has opponent axes as in cone-contrast and CIELAB spaces. The *dominant hue*, or the hue angle at which color preference is maximum, can be calculated directly from the weighted combination of the two mechanisms. Linear combinations of sine and cosine functions yield a single sinusoid curve with a phase shift, according to the following formula:

$$w_1 \cos(h) + w_2 \sin(h) = A \cos(h - \delta) \quad (2)$$

where $A = \sqrt{w_1^2 + w_2^2}$ and $\delta = \text{atan2}(w_2, w_1)$. A is the amplitude of the resulting sinusoid, and δ is the hue angle where the curve reaches its peak. Preference falls off as the hue shifts away from δ,

[9] Ling and Hurlbert's (2007) participants also completed a two-alternative forced choice (2AFC) task for a subset of the colors. The data did not differ significantly from the ratings data and the weights obtained from the models were highly positively correlated between all cross-comparison pairs of tasks and color sets. Therefore the model results from this paper should generalize to preferences obtained from 2AFC tasks.

Table 1
Models constructed by choosing one of two color spaces (cone-contrast or CIELAB), one of two coordinate systems (Cartesian or cylindrical), and either first alone or first- and second-degree factors. The names we use to describe the model and corresponding factors are also listed.

Color Space	Coordinate System	Factor Degree	Model Name	Factors
Cone-Contrast + CIELUV s_{uv}	Cart	1st	CCLS Cart	LM, $S-LM$, Lum, s_{uv}
		2nd	CCLS Cart2	LM, $S-LM$, $(LM)^2$, $(S-LM)^2$, $(LM)(S-LM)$, Lum, s_{uv}
	Cyl	1st	CCLS Cyl	$\cos(h_{cc})$, $\sin(h_{cc})$, Lum, s_{uv}
		2nd	CCLS Cyl2	$\cos(h_{cc})$, $\sin(h_{cc})$, $\cos(2h_{cc})$, $\sin(2h_{cc})$, Lum, s_{uv}
CIELAB	Cart	1st	LabC Cart	L, a, b, C_{ab}
		2nd	LabC Cart2	L, a, b, a^2, b^2, ab, C_{ab}
	Cyl	1st	LabC Cyl	L, $\cos(h_{ab})$, $\sin(h_{ab})$, C_{ab}
		2nd	LabC Cyl2	L, $\cos(h_{ab})$, $\sin(h_{ab})$, $\cos(2h_{ab})$, $\sin(2h_{ab})$, C_{ab}

and it reaches a minimum at the hue opposite δ. The constant factor A, which is always nonnegative, indicates the strength of the effect.

This decomposition is illustrated in Fig. 1 for four computationally generated patterns of color preferences. Fig. 1A illustrates a hue preference profile for a hypothetical individual who strongly prefers reds with a peak around $\delta = 40°$ in CIELAB space. The first two rows sum to the third, illustrating the decomposition: $0.766\cos(h) + 0.643\sin(h) = \cos(h - 40°)$. The preference peaks are also visualized in polar plots on the right. Similar decompositions are shown in Fig. 1B–D, which illustrate hue preference profiles for yellow, green, and blue, respectively.

2.2.3. Factor degrees

Within each coordinate system, we analyzed models that had 1st degree factors only and models that had 1st and 2nd degree factors. In the Cartesian system 1st degree factors were LM and $S-LM$ in CCLS models and a and b in LabC models. In the Cartesian system 2nd degree factors were the quadratic factors $(LM)^2$, $(S-LM)^2$, $(LM)(S-LM)$ in CCLS models and a^2, b^2, ab in LabC models. This is the two-dimensional equivalent of polynomial regression. Adding quadratic terms allows the model to capture more intricate dependencies among factors, such as having a similar preference for opposing hues. Including the quadratic cross-terms such as ab makes model fits invariant under linear transformations of the variables. For example, replacing a and b with $(a-b)$ and $(a+b)$ respectively will not change the model predictions. In particular, this also makes the model invariant under rotations in the (a, b) plane.

In the cylindrical system, the 1st degree factors were $\cos(h)$ and $\sin(h)$ as described in Section 2.3.2 and Fig. 1. The 2nd degree factors were the second harmonics; $\cos(2h)$ and $\sin(2h)$. As with the basic cylindrical models, coefficients corresponding to the 2nd harmonic can also be decomposed in terms of a dominant hue angle and amplitude:

$$w_3\cos(2h) + w_4\sin(2h) = A'\cos(2(h-\delta')) \quad (3)$$

where $A' = \sqrt{w_3^2 + w_4^2}$ and $\delta' = \frac{1}{2}\text{atan2}(w_4, w_3)$. For the 2nd harmonic, the dominant hue angle satisfies $0° \leqslant \delta' < 180°$ and indicates a preference axis; maximum preference occurs for opposite hues aligned with the axis and minimum preference occurs at opposite hues orthogonal to the axis.

Fig. 2 illustrates how a complex hue preference profile is decomposed into its 1st and 2nd harmonics. The preference profile in this example shows the hue preference predictions of our 2nd-harmonic Lch model ("LabC Cyl2") evaluated on the average preferences from the BCP-32 dataset. The details of the model predictions are in Section 3.3, but for now it can be considered as a typical hue preference profile. The process of using MLR amounts to searching for the coefficients that should multiply each of the four simpler sinusoids (two 1st harmonics and two 2nd harmonics) so they combine into the best fit for the data. The first harmonic shows a blue dominant hue, with a peak around blue and a trough around yellow. The second harmonic shows an axis from orangish-red to bluish-green with peaks around orangish-red and bluish-green and troughs around yellowish-green and blueish-purple. Summing the 1st and 2nd harmonics results in the characteristic hue preference profile that has a broad peak around blue/bluish-green, a trough around yellow, and a moderate preference for red (e.g., Guilford & Smith, 1959; Palmer & Schloss, 2010; Taylor & Franklin, 2012).

2.3. Approach to evaluating the models

We evaluated the models using the BCP-32 and NCL-126 datasets. For each dataset, we computed the variance accounted for by each of the MLR models in describing the (1) color preference judgments averaged over participants and (2) color preference judgments of each individual participant. The average models assign a single set of weights to the factors, which characterizes the average pattern of color preferences. The individual models can assign a different set of weights to the factors for each participant to best characterize that individual's pattern of color preferences.

Models with more factors are expected to perform better but they run the risk of overfitting the data. To mitigate this concern, we used k-fold cross-validation (Friedman, Hastie, & Tibshirani, 2001), which estimates the residual error that might occur if the model were used to predict outcomes that were not part of the original dataset.

We applied k-fold cross validation across colors as follows:

1. Randomly partition the colors into k sets of equal size. These are called *folds*.
2. Assign one fold to be the *test set*, and the remaining $k-1$ folds to be the *training set*.
3. Find the model weights associated with the training set, and compute the residual error when this model is used to predict the test set.
4. Repeat steps 2–3 using the k different folds as training sets and compute the average residual error.
5. Repeat steps 1–4 several times, choosing different random partitions every time, and average all the residual errors obtained.

Typical choices for k are five or ten. For convenience, we chose values of k that were divisible by the total number of colors in the dataset: $k = 8$ for BCP-32 colors and $k = 9$ for NCL-126 colors. For step 5, we repeated the entire process ten times.

Unlike the residual error computed on the entire dataset, which continues to shrink as we add more factors to a model, the residual error estimate from k-fold cross-validation will grow if the model begins to over-fit to the data. In general, the error estimate from k-fold cross-validation is always larger than the residual error found by using the entire dataset. Consequently, the variance explained is lower when using k-fold cross-validation. Although the models we

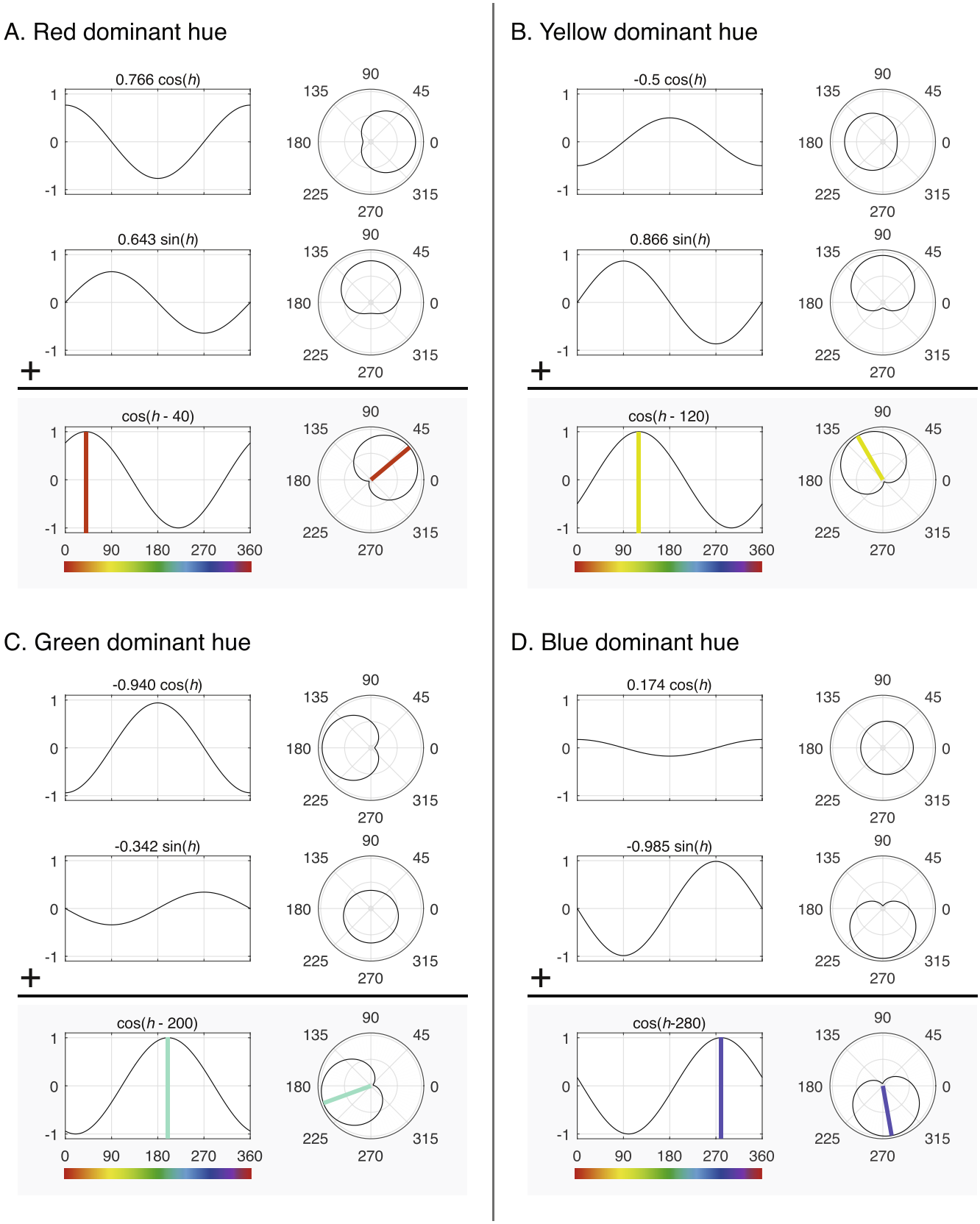

Fig. 1. Hue preference profiles that were computationally generated to illustrate how the cylindrical model represents color preferences of hypothetical individuals who most prefer red (A) yellow, (B), green (C), or blue (D). Each panel shows Fourier decomposition of sinusoidal hue preference profiles for the dominant hue. In each panel, the top plot shows the cos(h) component, the second from the top shows the sin(h) component, and the third shows the sum. Each graph may also be represented in polar coordinates (shown on the right). The polar plots are shifted so that the origin corresponds to a y-axis value of −1. The dominant hue is indicated by colored vertical lines in the rectangular plots and radial lines in the polar plots.

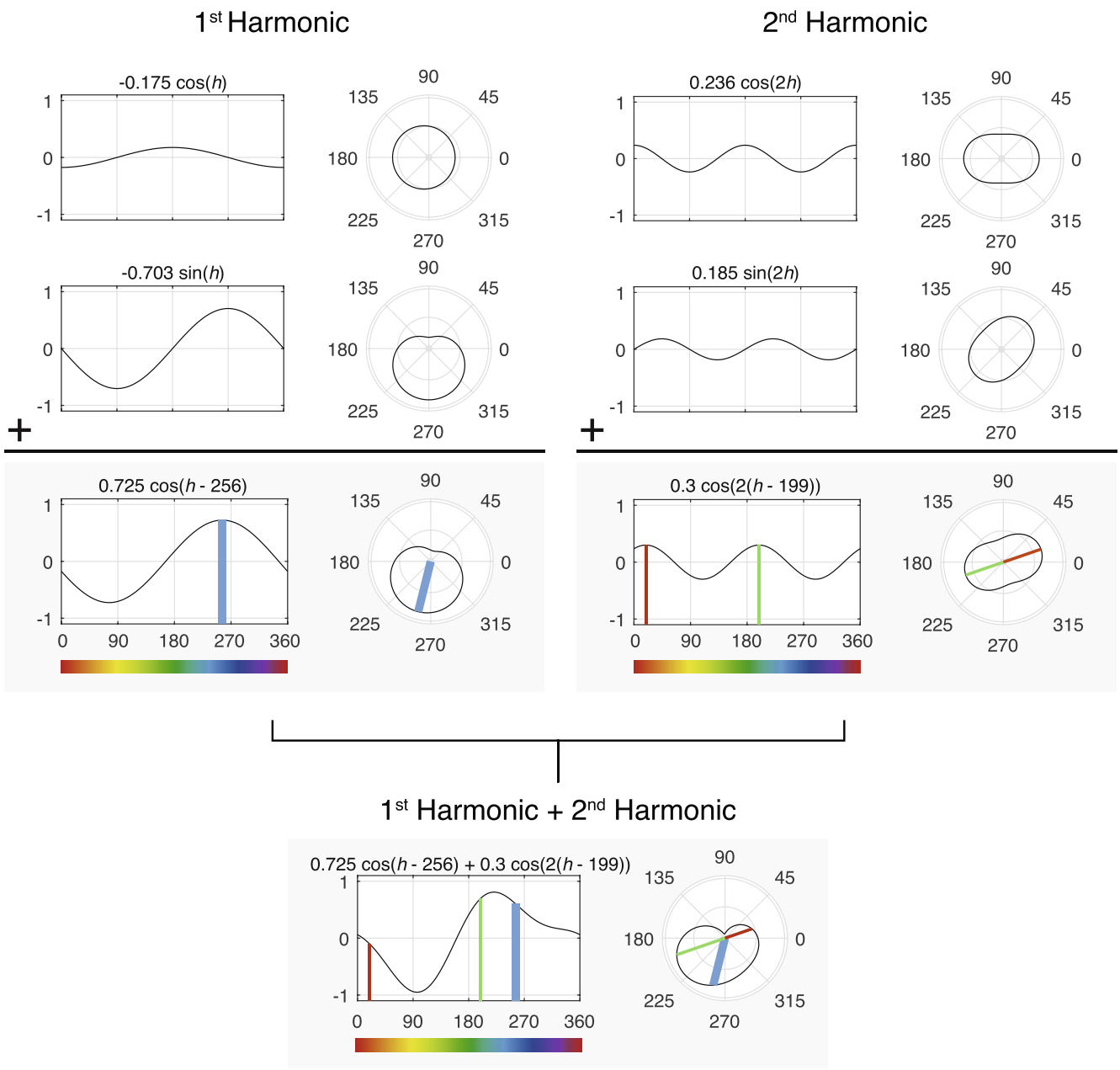

Fig. 2. Fourier decomposition of the average color preferences from the BCP-32 dataset. The left column shows the decomposition of the 1st harmonic into its fundamental components cos(*h*) and sin(*h*). The dominant 1st harmonic is blue, shown as a thick vertical line (in the rectangular plots) or radial line (in the polar plots). The right column shows the decomposition of the 2nd harmonic into its fundamental components cos(2*h*), and sin(2*h*). The dominant 2nd harmonic is along an orangish-red to bluish-green axis, shown in the thin colored lines in the rectangular and polar plots. The bottom plot shows the combination of all four components (the model prediction), with the thick lines representing the dominant hue angle (1st harmonic) and the thin lines representing the dominant axis (2nd harmonic).

evaluated differed in the number of parameters fitted, the R^2 values obtained from the cross-validation procedure provide a goodness of fit measure that allows us to make direct comparisons across models.

3. Results and discussion

Fig. 3 shows the proportion of variance (R^2) that each of the eight models accounted for in the color preference data (exact values reported in Table 2). In the following sections we compare the model fits on the average and individual data for each dataset using summary R^2 values. These R^2 values were computed in slightly different ways for the average vs. individual datasets and for the "All Data" vs. cross-validated datasets, which we describe in each section before discussing the results. We conducted the main statistical analyses on the individual cross-validated data because we could use analyses of variance (ANOVAs) to compare the means across individuals without concern about overfitting providing an unfair advantage to larger numbers of factors (Section 3.4).

3.1. Describing average color preferences for all the colors

3.1.1. Calculating summary R^2

The summary R^2 values for the fits to the average of all the color preference data within a dataset ("All Data") are simply the squared correlations between average data and model predictions.

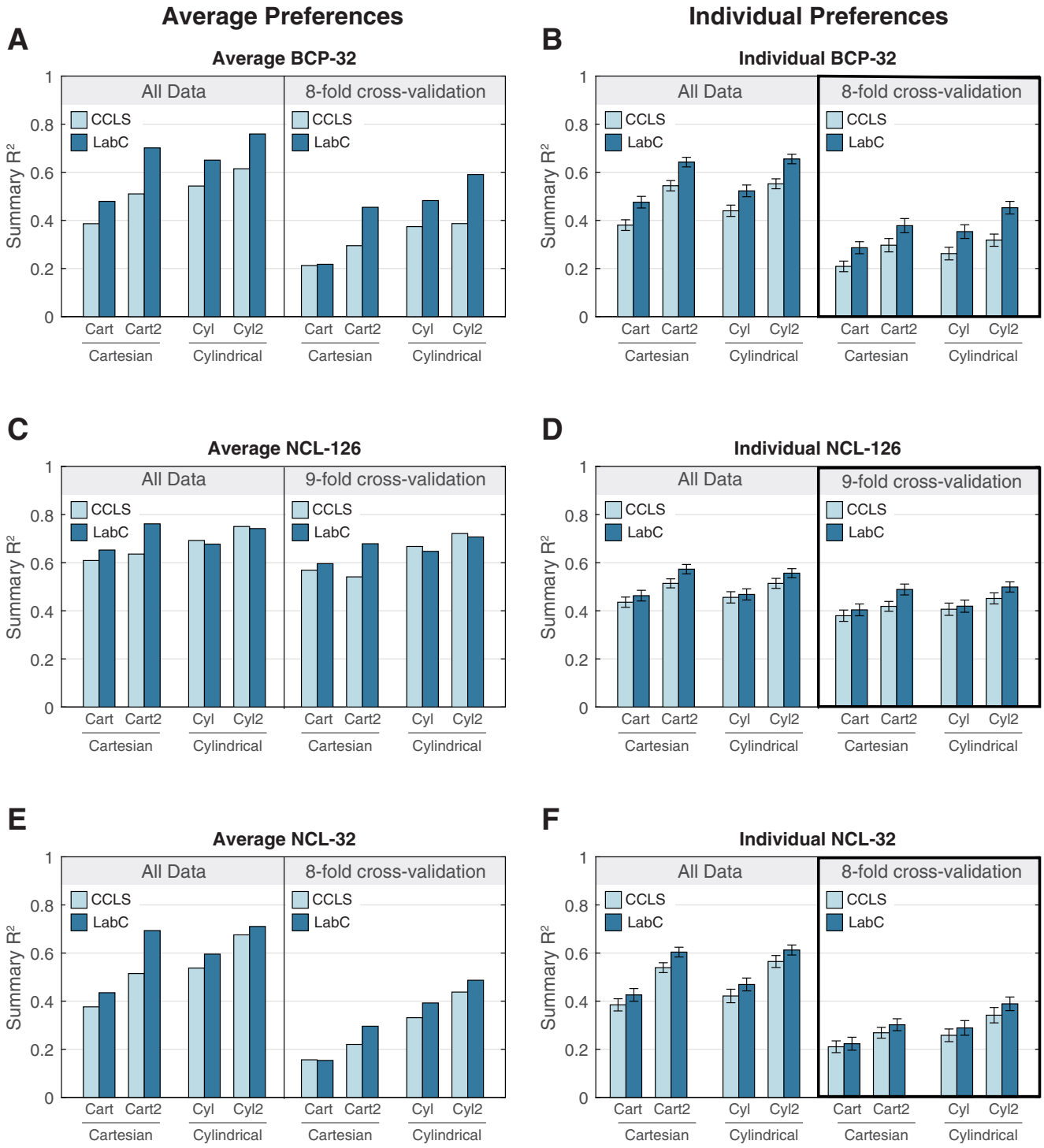

Fig. 3. Proportion of variance (summary R^2) accounted for by each model using all the colors ("All Data") and k-fold cross-validation for the BCP-32 dataset (A) averaged over participants and (B) for individual participants; the NCL-126 dataset (C) averaged over participants and (D) for individual participants; and the NCL-32 colors (subset of NCL-126 matched to BCP-32 colors, see text in Section 3.4 for details) (E) averaged over participants and (F) for individual participants. See the text for descriptions of how summary R^2 values were calculated. The error bars in B, D, and F represent the standard errors of the means of the model fits across individual participants. The thick outline around the cross-validation model fits in B, D, and F indicate the data on which we conducted the main statistical tests to compare models (Section 3.4). Table 2 presents the summary R^2 values in this figure.

This is a standard way of modeling average color preferences (e.g., Ling & Hurlbert, 2007; Palmer & Schloss, 2010).

3.1.2. Model results

The All Data plots in Fig. 3A and C show how well each model describes the average pattern of color preferences in the BCP-32 and NCL-126 datasets, respectively. Recall that the CCLS Cart model is Ling and Hurlbert's (2007) extended-cone contrast model. CCLS Cart accounted for 39% of the variance in the average BCP-32 dataset and 61% in the NCL-126 dataset, comparable to prior reports (Ling & Hurlbert, 2007; Palmer & Schloss, 2010).

Compared with the other models, the CCLS Cart model was the weakest for both the BCP-32 data set (Fig. 3A) and the NCL-126 dataset (Fig. 3C). For the BCP-32 dataset, fits were greater for LabC

Table 2
Summary R^2 values for each model corresponding to Fig. 3 (see details in text).

Dataset	Level	Model Type	CCLS Cart	LabC Cart	CCLS Cart2	LabC Cart2	CCLS Cyl	LabC Cyl	CCLS Cyl2	LabC Cyl2
BCP-32	Avg.	All Data	0.39	0.48	0.51	0.70	0.54	0.65	0.61	0.76
		8-fold cv	0.21	0.22	0.30	0.45	0.37	0.48	0.39	0.59
	Indiv.	All Data	0.38	0.48	0.54	0.64	0.44	0.52	0.55	0.66
		8-fold cv	0.21	0.29	0.30	0.38	0.26	0.35	0.32	0.45
NCL-126	Avg.	All Data	0.61	0.65	0.64	0.76	0.69	0.68	0.75	0.74
		9-fold cv	0.57	0.60	0.54	0.68	0.67	0.65	0.72	0.71
	Indiv.	All Data	0.44	0.46	0.51	0.57	0.46	0.47	0.51	0.56
		9-fold cv	0.38	0.40	0.42	0.49	0.41	0.42	0.45	0.50
NCL-32	Avg.	All Data	0.38	0.44	0.51	0.69	0.54	0.60	0.68	0.71
		8-fold cv	0.16	0.15	0.22	0.30	0.33	0.39	0.44	0.49
	Indiv.	All Data	0.38	0.43	0.54	0.60	0.42	0.47	0.56	0.61
		8-fold cv	0.21	0.22	0.27	0.30	0.26	0.29	0.34	0.39

models than CCLS models, for cylindrical models than Cartesian models, and for models that included 1st and 2nd degree factors than for models that only included 1st degree factors. The best model was the 2nd-harmonic Lch model ("LabC Cyl2"), which accounted for 76% of the variance. For the NCL-126 dataset there was less differentiation among the models, but LabC Cyl2 was among the best and it accounted for similar variance (74%) as it did for the BCP-32 dataset. Differences between the BCP-32 and NCL-126 datasets are discussed in Section 3.4.

3.2. Predicting average color preferences with cross-validation

3.2.1. Calculating summary R^2

The R^2 values for the fits to the average data using k-fold cross-validated data (Fig. 3A and C) were obtained as follows. As described in Section 2.3, k-fold cross-validation involves running the model k times for each of k folds, which produces model predictions for the untrained colors within each fold. If $k = 8$, that results in eight R^2 values. This procedure was repeated 10 times for each model resulting in 80 R^2 values for each model. The R^2 for cross-validation plots in Fig. 3A and C depict the average of these values for each model.

3.2.2. Model results

The results of the cross-validated models illustrate the models' ability to predict preferences for colors that were not in the training set used to set the parameters for the models. Although here we had measurements of people's preferences for the untrained colors that we could use to assess how well the model predictions fit the data, this method can be used to predict preferences for untrained colors where no color preference measurements have been made. The pattern of model fits using cross-validation was strongly correlated with the pattern of model fits for all data within the BCP-32 dataset ($r(6) = 0.96$, $p < 0.001$) and the NCL-126 dataset ($r(6) = 0.92$, $p < 0.01$). The model fits using cross-validation were generally lower than the model fits for All Data (BCP-32: $t(7) = 15.52$, $p < 0.001$, $d = 5.51$; NCL-126: $t(7) = 4.77$, $p = 0.002$, $d = 1.69$), which is to be expected from the cross-validation procedure (see Section 2.3). For the BCP-32 dataset, the LabC Cyl2 model had the strongest predictive ability, accounting for 59% of the variance. For the NCL-126 dataset there was less differentiation among the models, but LabC Cyl2 was among the best, accounting for 71% of the variance.

3.3. Describing individual preferences for all the colors

3.3.1. Calculating summary R^2

We computed the R^2 values for individual fits for all the data (Fig. 3B and D) by first computing the R^2 for each individual and then averaging the R^2 values across individuals. This is a standard procedure for evaluating color preference model fits for individuals (Ling & Hurlbert, 2007; Schloss et al., 2015).

3.3.2. Model results

Ling and Hurlbert's (2007) CCLS Cart model accounted for an average of 38% of the variance in the BCP-32 dataset and 44% in the NCL-126 dataset. Similar to the average data, LabC models generally accounted for more variance than CCLS models, and models with 1st and 2nd degree factors accounted for more variance than models with only 1st degree factors. LabC Cyl2 was the model that accounted for the most amount of variance in individuals' preferences in BCP-32 dataset. There was less differentiation among the models for the NCL-126 dataset, but LabC Cyl2 was among the best, accounting for 56% of the variance. We conduct more thorough analyses of the individual models when cross-validation was used (Section 3.4) because cross-validation makes it possible to directly compare models with different numbers of factors without giving unfair advantage to models with more factors.

3.4. Evaluating model performance for cross-validated individual model fits

3.4.1. Calculating summary R^2

We obtained the R^2 values for the fits of cross-validated models for individual subjects by conducting the procedure described for the average data (Section 3.2) but for each individual, and then averaging the R^2 values across individuals (Fig. 3B and D).

3.4.2. Model results

These R^2 values represent the degree to which the model can predict individuals' preferences for colors that were not used to train the models. CCLS Cart was the model that accounted for the least amount of variance in the BCP-32 (21% variance) and NCL-126 datasets (38% variance). LabC Cyl2 was the model that accounted for the most amount of variance in the BCP-32 dataset (45% variance) and the NCL-126 dataset (50% variance). Below we conducted a systematic analysis to understand which features make models more/less effective at accounting for variance in color preferences.

We evaluated the performance of the models for the BCP-32 and NCL-126 datasets by conducting a mixed design ANOVA on the amount of variance accounted for in the individual participants' color preferences. There were three within-subject factors: 2 Color Spaces (CCLS, LabC) × 2 Coordinate Systems (Cartesian, cylindrical) × 2 Factor Degrees (1st degree, 2nd degree). There was one between-subjects factor: dataset (BCP-32 vs. NCL-126).

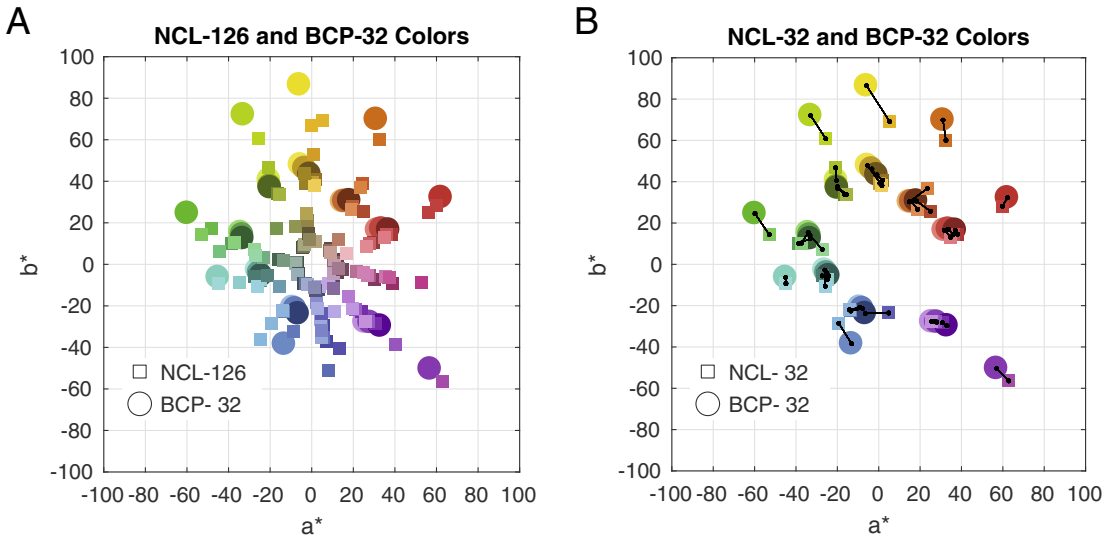

Fig. 4. The CIELAB (a^*, b^*) coordinates for (A) the BCP-32[10] colors (disks) and the full set of NCL-126 colors (squares) and (B) the NCL-32 colors that were matched to the BCP colors (black lines denote matched color pairs; see text for details on matching procedure). Note that the chromaticity coordinates of all colors are projected onto a single L^* plane, despite variations in L^*.

The ANOVA indicated that all four factors influenced the model fits. A main effect of Color Space indicated that LabC models (constructed from factors in higher-level CIELAB color space) fit the color preference data better than CCLS models (constructed from factors in lower-level cone-contrast space) ($F(1,86) = 75.62$, $p < 0.001$, $\eta_p^2 = 0.47$). A main effect of Coordinate System indicated that models using cylindrical coordinate systems fit the data better than models using Cartesian coordinates ($F(1,86) = 17.82$, $p < 0.001$, $\eta_p^2 = 0.17$). A main effect of Factor Degree ($F(1,86) = 71.66$, $p < 0.001$, $\eta_p^2 = 0.46$) indicated that models including 2nd degree factors in addition to 1st degree factors were better than models including 1st order factors alone. The better fit of the 2nd order models indicate that hue preference components are not only opponent—people can like (and dislike) both colors at opposite ends of hue-opponent axis dimensions. An interaction between Color Space and Degree ($F(1,86) = 9.34$, $p = 0.003$, $\eta_p^2 = 0.10$) indicated that the extent to which the 2nd degree models were better than the 1st degree models was greater within LabC models than within CCLS models.

Evaluating the between subjects factor of dataset, the overall fit was greater for the NCL-126 dataset than the BCP-32 dataset ($F(1,86) = 12.82$, $p < 0.001$, $\eta_p^2 = 0.13$). The pattern of model fits was also slightly different across the two datasets, as supported by a 2-way interaction between Color Space and Dataset ($F(1,86) = 13.79$, $p < 0.001$, $\eta_p^2 = 0.14$) and a 3-way interaction between Color Space, Coordinate System and Dataset ($F(1,86) = 8.31$, $p < 0.01$, $\eta_p^2 = 0.09$).

There were several differences between the datasets that could have caused these results, including participant sample (UK vs. US) number of colors (126 vs. 32), and range of colors. Although both sets of color sample the hue circle similarly, apart from larger gaps in the BCP-32 set in the bluish-purple and magenta hue range, the NCL-126 set includes colors at lower lightness and chroma levels not in the BCP-32 set and the BCP-32 set includes colors at high chroma not in the NCL-126 set (see Fig. 4A).

To understand the effects of these differences between datasets we sub-sampled the NCL-126 dataset to isolate the colors that were most similar to the BCP-32 colors, resulting in more equivalent and comparable data sets. We first calculated the distance (ΔE) between each BCP-32 color and each NCL-126 color in CIELAB space. We then determined which NCL-126 color best matched each of the BCP-32 colors (minimizing ΔE). Three of the 32 colors resulted in duplicate picks, so we chose the second-closest match those cases. We separately verified that this matching procedure resulted in the lowest possible overall ΔE over the entire set of matches, using an optimization routine that examined all possible matches. We refer to the new sub-sampled set as the NCL-32 colors (see Fig. 4B and Table S10 in the Supplementary Material).

The model results for the NCL-32 dataset are shown in Fig. 3E (fits with average data) and Fig. 3F (fits with individual data). To compare the model fits to those for BCP-32 dataset we conducted the same mixed design ANOVA as above, but this time including the NCL-32 dataset instead of the NCL-126 dataset. By matching the datasets we eliminated the main effect of dataset ($F(1,86) = 1.07$, $p = 0.31$, $\eta_p^2 = 0.01$) and the 3-way interaction between Color Space, Coordinate System and Dataset ($F < 1$). This suggests that by equating the number of colors and largely matching their appearance we were able to substantially reduce the difference in model fits between the two datasets. Thus, the overall difference between fits for the NCL-126 and BCP-32 datasets were likely due to the difference in number of colors and/or sampling of colors within these datasets, not due to participants being from different populations (UK vs. US). However, the interaction between Color Space and Dataset still holds ($F(1,86) = 15.92$, $p < 0.001$, $\eta_p^2 = 0.17$): the extent to which the LabC model fit the data better than the CCLS model was greater for the BCP-32 dataset than the NCL-32 dataset. The main effects of Color Space, Coordinate System, and Degree reported above also still hold ($F(1,86) = 60.61$, 42.29, 78.37, $p < 0.001$, 0.001, 0.001, $\eta_p^2 = 0.41$, 0.33, 0.48, respectively), but the interaction between Color Space and Factor Degree was eliminated ($F(1,86) = 3.12$, $p = 0.082$, $\eta_p^2 = 0.04$).

To summarize these analyses, models that best fit the color preference data were constructed from a higher-level color space (CIELAB) using a cylindrical coordinate system (Lch), with 1st and 2nd degree factors. The most effective model for both the

[10] The a^* and b^* coordinates for the BCP-32 colors are scaled differently from those in Fig. 1 of Palmer and Schloss (2010) because we used a white point of 116 cd/m^2 and they used the background luminance (19.26 cd/m^2).

NCL-32 and BCP-32 datasets in the cross-validation test was the *2nd-harmonic Lch model* (labeled "LabC Cyl2" for brevity in Figs. 3 and 4, and Table 2). We note that this best-fitting model significantly improves on the performance of previous models such as Ling and Hurlbert's (2007) extended cone-contrast model (labeled "CCLS Cart" above) and that it differs from that model both in its underlying color space (physiological cone-contrast vs. standardized CIELAB) and in its use of cylindrical vs. Cartesian coordinates.[11] In the following section we further explore the 2nd-harmonic Lch model to understand which aspects of the color preference data it fits and where it falls short, and how the parameters of the model can be used to parsimoniously describe patterns in average and individual participants' color preferences.

3.5. Illustrating model predictions

The color preference data we used for evaluating the models described above are illustrated in Fig. 5A (from Schloss et al., 2015) and Fig. 5B (from Ling & Hurlbert, 2007). Fig. 5B plots only the subset of the NCL-126 data that contained all 10 hues within each value/chroma level. We chose this subset because it provided a way to demonstrate hue preference functions at each saturation and lightness level without having missing hues within the function. Both sets of color preferences show a typical profile with a peak at blue, a trough around yellow-green, with an especially pronounced dislike for dark yellow. In the NCL dataset, it is also apparent that the darker and less saturated blues are less preferred than the light saturated blues, consistent with the overall reduction in preference as lightness and saturation are decreased. This trend is not apparent in the BCP dataset, but it contains far fewer blues overall and no colors at relatively low lightness and saturation in comparison with the NCL dataset. Note that in the BCP dataset, the colors in one "cut" are not all at the same value and chroma level, so there is not a direct correspondence with the curves in the NCL dataset, which have constant value and chroma within each curve.

Fig. 5C and 5D show the predicted color preferences using the LabC Cyl2 representation for BCP-32 and NCL-126, respectively. To obtain the predictions plotted here, we first fit the model to each participant's color preference data, calculated the individual predictions from the model weights, and then calculated the mean of those predictions. Although only a subset of the NCL-126 data are shown (those that contained colors at each hue within the value/chroma level), the model fits were conducted on the full datasets for 126 colors. Both sets of predictions depicted in Fig. 5 are strongly correlated with the color preference data on which they were based ($r = 0.87$, $p < 0.001$ for both datasets). The models capture a broad peak around blues and cyans, a trough around yellows, and an overall preference for saturated colors.

Fig. 4E and 4F show the differences between the color preference data and the model predictions for each participant, averaged over participants. The prediction errors show that the best model still does not capture the extreme dip in preference for dark relatively saturated yellow, the small peak of preference at light, saturated yellow-oranges and the reduced preference for less saturated, darker blues. The deviations are largely due to the interactions between hue, chroma and lightness that cannot be captured by our additive model. For example, in the BCP-32 dataset (Fig. 5A), the model captures the similar hue preference profile for the saturated, light, and muted colors, and the overall elevated preference for the saturated colors (Fig. 5C). However, as shown in the residuals (Fig. 5E), the model does not capture deviations from this profile for the dark colors—it fails to predict how much people disliked dark yellow and like dark red and dark green (Palmer & Schloss, 2010). For the NCL dataset (Fig. 5B), the model again captures the general hue preference profile but does not account for deviations at different value/chroma levels (Fig. 5D). As shown in the residuals (Fig. 5F), it under-predicts preference for lighter higher-chroma cool hues and over-predicts preference for darker, lower-chroma cool hues. It also does not predict how much people dislike the darkish yellow (V/C of 5/6) that is similar to the BCP Dark Yellow. Future research will be necessary to capture these complexities in the color preference data.

A feature of the Cyl2 representation is that the preference profiles may be interpreted in terms of dominant hue angles and dominant hue axes, as described in Section 2.2 (Figs. 1 and 2). Fig. 6 shows the dominant hue angles (1st harmonic) and dominant axes (2nd harmonic) for each participant in both datasets.

For the 1st harmonic, each radial segment in Fig. 6 has an angle corresponding to the dominant hue for that individual and a radius corresponding to its weighting (relative importance of the dominant hue in describing the color preferences of that individual). The opposite of the dominant hue angle segment (not depicted) represents the least liked hue. For both datasets, the 1st harmonic weights confirm the weighting toward the blue side of the blue-yellow axis seen previously, meaning that people generally prefer bluish hues most and yellowish hues least.

For the 2nd harmonic, segments correspond to two opposing hues with equal preference strengths (each radial segment extends an equal length in opposite directions). The angle of the segment corresponds to the dominant opposing hue pair and the radius corresponds to the strength in preference for the opposing hues (i.e., amplitude of the harmonic). For both datasets, the 2nd harmonic reveals strongest weighting along an orangish-red to bluish-green axis, indicating that people tend to like both orangish-reds and bluish-greens. This pattern also implies that the troughs of the 2nd harmonics are around a greenish-yellow to bluish-purple axis. That means that people tend to dislike both greenish-yellows and bluish-purples. Had there been no tendency to like (or dislike) opposing hues, weights would be near zero on the 2nd harmonic, resulting in very small line segments in the 2nd harmonic plots in Fig. 6, which is not the case.[12]

The combination of a dominant blue 1st harmonic with a less pronounced reddish-orange to greenish-blue 2nd harmonic produces the typical hue preference function that has long been documented in the literature (Guilford & Smith, 1959): a broad peak in preference around blues and greens, moderate preference for reds, and a trough in preference around yellows, as shown in Fig. 2.

4. General discussion

Color space based models are powerful tools for describing and predicting color preferences because they provide a concise way of characterizing complex patterns of data over large sets of colors. For example, the 2nd-harmonic Lch (LabC Cyl2) model describes preferences for 126 colors in the NCL-126 data using six parameters and an additive constant. Once the parameters have been estimated for a particular individual or group, it is possible to predict preference for any color that can be specified within the space. The anticipated accuracy of the prediction can be estimated from the model fits using cross-validation.

We set out to determine which color space models most effectively describe and predict color preferences and found that the

[11] CCLS Cart, which is Hurlbert and Ling's (2007) extended cone-contrast model, is the same as the cylindrical model for stimuli that vary only in hue because the cone-contrast components are the sine and cosine functions in a circle of unit radius.

[12] Note that there is no a priori relationship enforced by the model between the directions of the dominant 1st and 2nd harmonic. The dominant peaks of the 1st and 2nd harmonic curves emerge independently from the data for each individual.

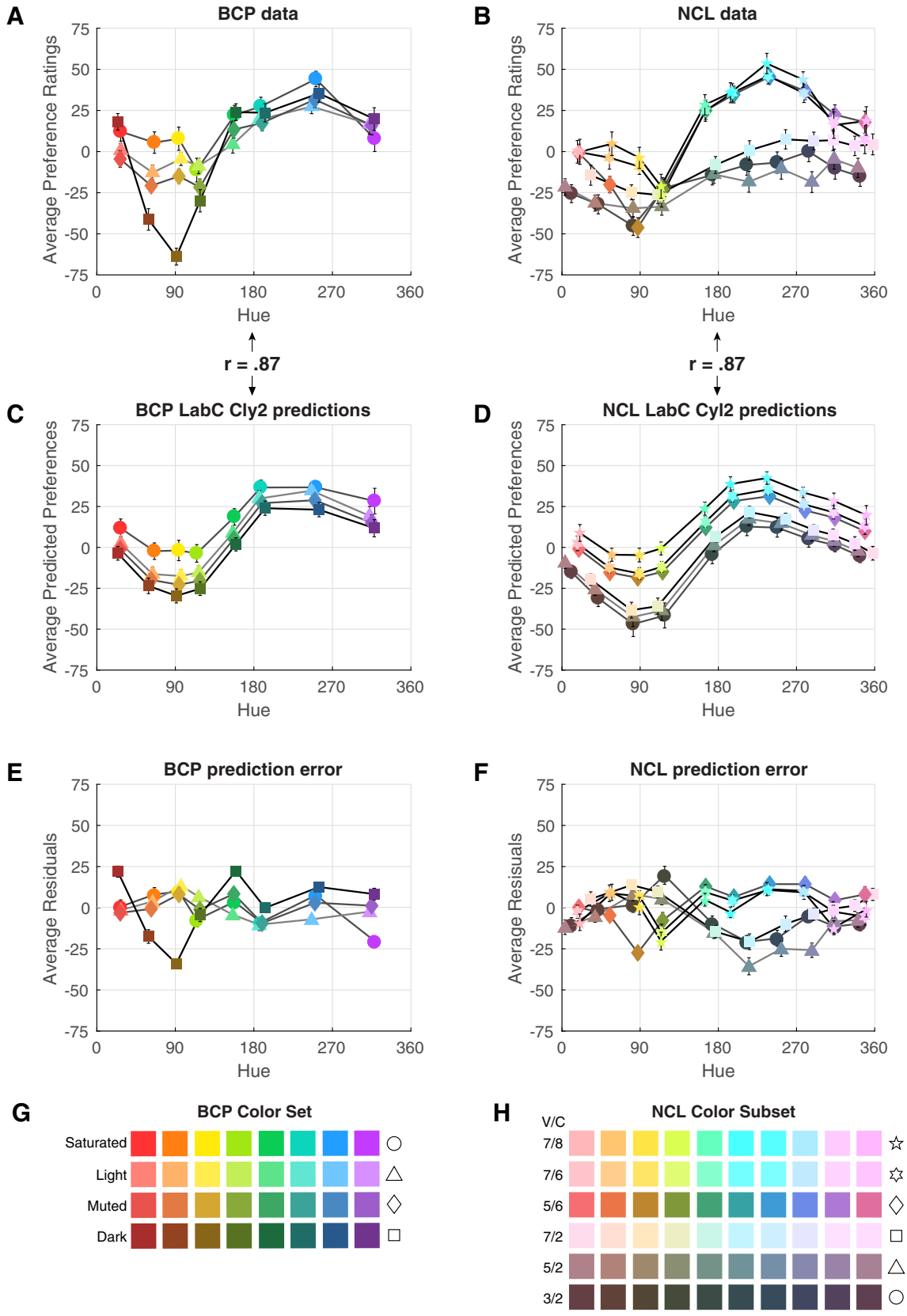

Fig. 5. Average preference ratings as a function of hue angle in CIELAB space (i.e., *h* in CIE Lch) (x-axis) for (A) the BCP-32 colors from Schloss et al. (2015), 48 observers, and (B) a 60-color subset of the NCL-126 colors from Ling and Hurlbert (2007), 40 observers, (scaled to range from −100 to +100). Model fits show predicted mean preferences for the BCP color set (C) and the NCL color set (D) for the same participants from the 2nd-harmonic Lch model (i.e., LabC Cyl2). Differences between the predicted and actual preferences for each color (actual – predicted), averaged over all individuals, are shown for the BCP (E) and NCL (F) color sets. Different lines connect hues at the same saturation/lightness level in each set (roughly for the BCP-32 set, and exactly for the NCL set), illustrated by rows in (G) and (H) (not colorimetrically accurate). The correlation (*r*) between the upper and lower graphs is the Pearson correlation coefficient between the mean preference data and the mean predicted preferences for the same participants, over all the colors shown.

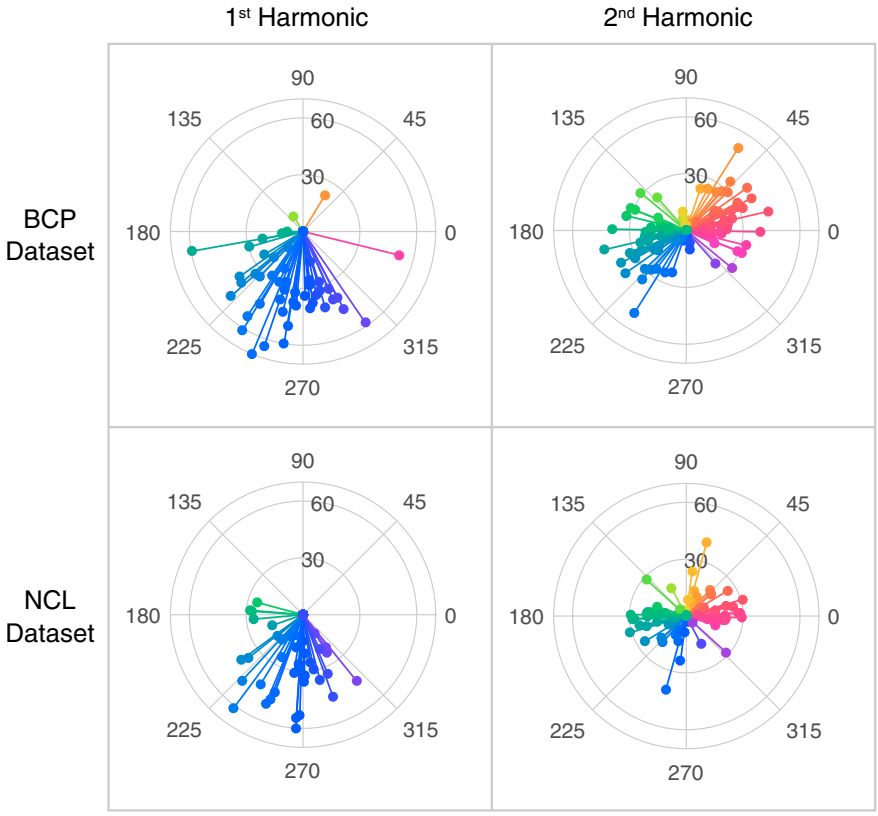

Fig. 6. Weights on the 1st harmonic, which indicate the dominant hue angle and strength of its contribution for each participant in the BCP and NCL data sets, and weights on the 2nd harmonic, which indicate the dominant axis for each participant.

2nd-harmonic Lch model was most effective. The model was constructed from CIELAB space in cylindrical coordinates space including first and second harmonics of hue. It accounted for 76% of the variance average preferences in the BCP-32 dataset and 74% of the variance in average preferences in the NCL-126 dataset. When used to fit individual subjects' color preferences and averaging over those individual fits, the model accounted for 66% of the variance for the BCP-32 dataset and 56% for the NCL-126 dataset. These fits are substantially greater than the fits from models that were previously reported in the literature (see Fig. 3 and Table 2).

Some of the improvement in the fit over previous models comes from the use of cylindrical coordinates, as verified by the main effect of coordinate system in the comparison of model fits for the cross-validated data. This coordinate system provides a way to encode hue, chroma, and lightness as separate, independent factors, eliminating the conceptual redundancy in previous Cartesian models (Hurlbert & Ling, 2007; Ling & Hurlbert, 2007; Palmer & Schloss, 2010), which coupled saturation with hue. The improvement with cylindrical coordinates is visible for both color spaces, although generally weaker for the cone-contrast models. The addition of second harmonic components improves the fit, still further, by enabling double peaks in preference to emerge as a function of hue. This effect is stronger for the LabC models than for the CCLS models, evidenced by the interaction between color space and factor degree in the model fits for the BCP-32 and NCL-126 datasets.

It is noteworthy that the coordinate system that best captured variations in color preference in the models that we tested also corresponds to the dimensions that characterize human perceptual experience (hue, saturation/chroma, and lightness). Although the evidence for explicit neural encoding of hue, saturation, and lightness, and their interactions, is still lacking, recent studies in non-human primates indicate that color encoding at higher stages of visual processing differs from early stages in no longer representing colors by linear responses tuned to cardinal axes. Instead, neural activity in higher areas shows narrow tuning to hue, and represents hue differences similarly to distances in the perceptually uniform space CIELUV (Bohon et al., 2016).

The question remains why the 2nd-harmonic Lch model fits the data overall better than the 2nd-harmonic cone-contrast model, given that both have enhancements provided by decoupling hue and saturation, and by including higher-order functions of hue. The answer must lie in the structure of the underlying color spaces (i.e. cone-contrast and CIELAB space). Fig. 7 illustrates differences between the two spaces. Fig. 7A shows the coordinates of a uniform grid of colors in the chromaticity plane of $L^* = 80$ and Fig. 7B shows those same coordinates transformed into cone-contrast space. Note that it is not possible to display colors at all chromaticity values in Fig. 7A at $L^* = 80$; this is a schematic diagram to illustrate the mathematical relation between CIELAB and cone-contrast space. Fig. 7C and 7D show what a circle in CIELAB space maps to in cone-contrast space. It is apparent that the circle is compressed for negative values of $S - LM$ and expanded for positive values of $S - LM$. Differences between colors are also more compressed on the negative side of LM compared to the positive side. Fig. 7E and 7F illustrate differences between the two spaces in encoding chromaticity at different lightness levels. As the lightness level is reduced in cone-contrast space, the chromaticity plane shrinks, but there is no corresponding shrinking in CIELAB space.

The nonlinear nature of the transformation between the spaces results in different model predictions. For example, consider two pairs of colors: two greens far from the origin and two blues closer to the origin, with the same CIELAB chroma difference between the colors in each pair. Because the compression of chroma increases with greater distance from the origin, the difference in CIELAB chroma between the two greens would transform to a smaller relative difference in cone-contrast space, in comparison with the

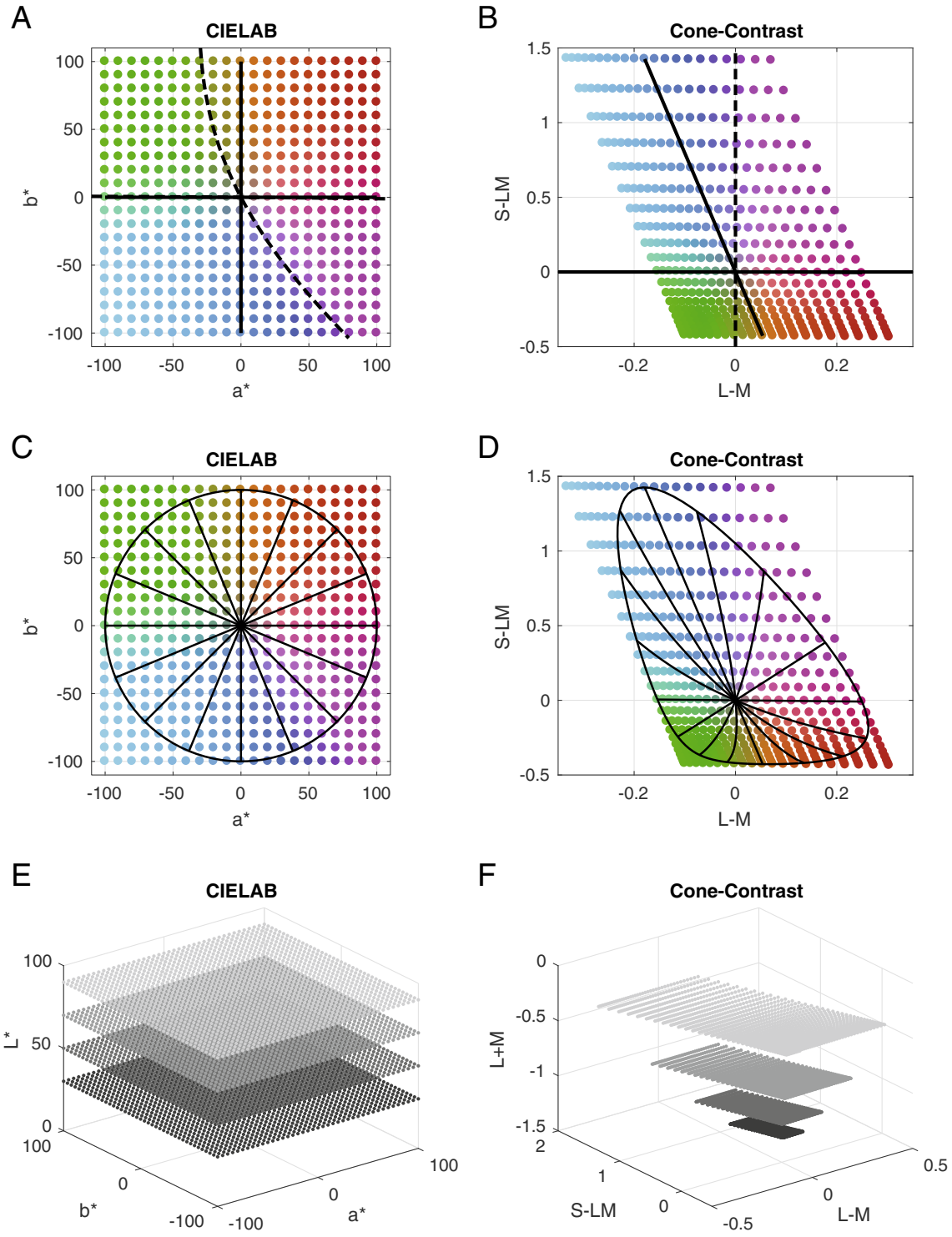

Fig. 7. An equally spaced grid of points in an iso-lightness plane in CIELAB (A) and a transformation of the points in A into cone-contrast coordinates (B). Thick black lines in the cone-contrast plane indicate the transformed a^* (horizontal) and b^* (oblique) axes. Dashed lines in the CIELAB iso-lightness plane indicate the transformed $S - LM$ axis (oblique) and LM (horizontal) axis. The same points are shown drawn in (C) overlaid with a constant-chroma hue circle in CIELAB space and a transformation of that circle into cone-contrast coordinates (D). Grids of points like those in (A) and (C) are shown at different CIE L^* levels in (E). The points from E are transformed into cone-contrast space in (F).

chroma difference between the blues. Therefore, a model based on cone-contrast space would predict a smaller effect of chroma for the former. It is noteworthy that the part of color space where the BCP-32 and NCL-32 datasets disagree most is in the saturated green-yellow-orange region. This also happens to be the area where the conversion from CIELAB to cone-contrast causes the most compression. This effect may further contribute to disagreements in fits between both models (Color Space x Dataset interaction described in Section 3.4).

The 2nd-harmonic Lch model will be useful in future work for characterizing and analyzing patterns of color preference data. For example, it may be used to track how color preferences change along particular dimensions over time, such as the seasonal variations in color preferences previously reported by (Schloss, Nelson,

Parker, Heck, & Palmer, in press). It may also be used to characterize compactly the differences in color preference between different populations, for example, between typically and atypically developing individuals (Cranwell, 2016; Hurlbert, Loveridge, Ling, Kourkoulou, & Leekam, 2011). A caveat for such applications is that the model descriptions are only as good as the model fits—a model that is a poor fit to the data will not be successful at describing the data. One important aspect of the 2nd-harmonic Lch model to note is that the dominant hue outputs are angular in nature. Therefore, to analyze the outputs (e.g., compare dominant hue angles across groups) standard statistical techniques, such as *t*-tests, would need to be replaced by their circular equivalents (e.g. the Watson-Williams test (Zar, 2009, Section 27.4)).

Although we have focused on modeling color preferences in the present study, the 2nd-harmonic Lch model can be used to describe and predict any kind of judgments about colors provided the criteria outlined in the introduction are met. For example, the model might be used to characterize patterns of color-emotion associations such as the amount of happiness, anger, and calmness associated with colors (e.g., Dael, Perseguers, Marchand, Antonietti, & Mohr, 2016; Palmer, Schloss, Xu, & Prado-León, 2013; Valdez & Mehrabian, 1994).

For future work that uses the 2nd-harmonic Lch model to describe and predict patterns of color preferences, it will not be necessary to use the k-fold cross-validation procedure implemented here. We used k-fold cross-validation as a means to fairly compare different candidate models and to select the best one. In particular, we used it *across colors* to ensure that the fits produced were not overfitting the data. Any of the models described in this paper can be applied directly (without using k-fold cross-validation) to new data, for example to predict a subject's preference for an unseen color given their preference profile for a known set of colors. However, it would be beneficial to use k-fold cross-validation in future studies aimed at model comparison, especially when those models have different numbers of parameters.

An open question that would warrant further model comparisons concerns the type of model that will best capture interactions in color preference data. As described in Section 3.5, there are well-documented interactions between dimensions of color in color preference judgments, particularly between hue and lightness. The present models additively combine weights along the different factors without accounting for such interactions, and the 2nd-harmonic Lch model might further be improved by building ways to model these more complex patterns.

Although color space models are extremely useful for describing and predicting color preferences, they do not provide explanatory accounts for color preferences. That is, they do not explain how color preferences are formed, why they differ between typical trichromatic individuals, and why they change over time. However, by providing a parsimonious way of describing color preferences they may elucidate key patterns in color preferences that effective theories of color preferences seek to explain.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.visres.2017.07.001.

References

Adams, R. J. (1987). An evaluation of color preference in early infancy. *Infant Behavior and Development, 10*(2), 143–150.

Bimler, D., Brunt, J., Lanning, L., & Bonnardel, V. (2014). Personality and gender-schemata contributions to colour preferences. In W. Anderson, C. P. Biggam, C. Hough, & C. Kay (Eds.), *Colour Studies: A Broad Spectrum* (pp. 240–257). Amsterdam: John Benjamins Publishing Company.

Bohon, K. S., Hermann, K. L., Hansen, T., & Conway, B. R. (2016). Representation of perceptual color space in macaque posterior inferior temporal cortex (the V4 complex). *Eneuro, 3*(4).

Choungourian, A. (1968). Color preferences and cultural variation. *Perceptual and Motor Skills, 26*(3c), 1203–1206.

CIE. (2004). Colorimetry, 3rd ed.

Cranwell, M. B. (2016). *Colour perception in autism spectrum condition and williams syndrome* (PhD). Newcastle upon Tyne, UK: Newcastle University.

Dael, N., Perseguers, M.-N., Marchand, C., Antonietti, J.-P., & Mohr, C. (2016). Put on that colour, it fits your emotion: Colour appropriateness as a function of expressed emotion. *The Quarterly Journal of Experimental Psychology, 69*(8), 1619–1630.

Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *Journal of Physiology, 357*, 241–265.

Dittmar, M. (2011). Changing colour preference with ageing: A comparative study on younger and older native Germans aged 19–90 years. *Gerontology, 47*, 219–226.

Eskew, R. T., McLellan, J. S., & Giulianini, F. (1999). Chromatic detection and discrimination. In K. R. Gegenfurtner & L. T. Sharpe (Eds.), *Color Vision: From Genes to Perception* (pp. 345–368). Cambridge, UK: University Press.

Eysenck, H. J. (1941). A critical and experimental study of color preference. *The American Journal of Psychology, 54*, 385–391.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The Elements of Statistical Learning*. Berlin: Springer. Vol. 1: Springer series in statistics.

Guilford, J. P., & Smith, P. C. (1959). A system of color-preferences. *The American Journal of Psychology, 72*(4), 487–502.

Helson, H., & Lansford, T. (1970). The role of spectral energy of source and background color in the pleasantness of object colors. *Applied Optics, 9*(7), 1513–1562.

Hurlbert, A. C., & Ling, Y. (2007). Biological components of sex differences in color preference. *Current Biology, 17*(16), 623–625.

Hurlbert, A. C., Loveridge, C., Ling, Y., Kourkoulou, A., & Leekam, S. (2011). Color discrimination and preference in autism spectrum disorder. *Journal of Vision, 11*, 429. http://dx.doi.org/10.1167/11.11.429.

Hurlbert, A. C., & Owen, K. A. (2015). Biological, cultural, and developmental influences on color preference. In A. J. Elliot, M. D. Fairchild, & A. Franklin (Eds.), *Handbook of Color Psychology* (pp. 454–480). Cambridge: Cambridge University Press.

Kuehni, R. G., & Schwarz, A. (2008). *Color Ordered: A Survey of Color Systems from Antiquity to the Present*. Oxford: Oxford University Press.

Lennie, P., & Movshon, J. A. (2005). Coding of color and form in the geniculostriate visual pathway. *Journal of the Optical Society of America, 22*, 2013–2033.

Ling, Y. L., & Hurlbert, A. C. (2007). A new model for color preference: universality and individuality. In: *Proceedings of the 15th color imaging conference*, Albuquerque, NM, pp. 8–10.

McManus, I. C., Jones, A. L., & Cottrell, J. (1981). The aesthetics of colour. *Perception, 10*, 651–666.

Munsell, A. H. (1921). *A grammar of color*. New York: Van Nostrand Reinhold.

Ou, L. C., Luo, M. R., Woodcock, A., & Wright, A. (2004). A study of colour emotion and colour preference. Part III: Colour preference modeling. *Color Research & Application, 29*(5), 381–389.

Palmer, S. E., & Schloss, K. B. (2010). An ecological valence theory of human color preference. *Proceedings of the National Academy of Sciences, 107*(19), 8877–8882.

Palmer, S. E., Schloss, K. B., Xu, Z. X., & Prado-León, L. R. (2013). Music-color associations are mediated by emotion. *Proceedings of the National Academy of Sciences, 110*(22), 8836–8841.

Pereverzeva, M., & Teller, D. Y. (2004). Infant color vision: Influence of surround chromaticity on spontaneous looking preferences. *Visual Neuroscience, 21*, 389–395.

Reddy, T. V., & Bennett, C. A. (1985). Cultural differences in color preferences. *Proceedings of the Human Factors Society, 29*(6), 590–593.

Saito, M. (1981). A cross-cultural survey on colour preference. *Bulletin of the Graduate Division of Literature of Waseda University, 27*, 211–216.

Saito, M. (1996). A comparative study of color preferences in Japan, China and Indonesia, with emphasis on the preference for white. *Perceptual and Motor Skills, 83*(1), 115–128.

Schiller, F., & Gegenfurtner, K. R. (2016). Perception of saturation in natural scenes. *Journal of the Optical Society of America a-Optics Image Science and Vision, 33*(3), A194–A206.

Schloss, K. B., & Palmer, S. E. (2017). An ecological framework for temporal and individual differences in color preferences. *Vision Research, 141*, 95–108.

Schloss, K. B., Nelson, R., Parker, L., Heck, I. A., & Palmer, S. E. (in press). Seasonal variations in color preference. *Cognitive Science*.

Schloss, K. B., Hawthorne-Madell, D., & Palmer, S. E. (2015). Ecological influences on individual differences in color preference. *Attention, Perception, & Psychophysics, 77*(8), 2803–2816.

Schloss, K. B., & Palmer, S. E. (2015). Ecological aspects of color preference. In A. J. Elliot, M. D. Fairchild, & A. Franklin (Eds.), *Handbook of Color Psychology* (pp. 435–453). Cambridge: Cambridge University Press.

Schloss, K. B., Strauss, E. D., & Palmer, S. E. (2013). Object color preferences. *Color Research & Application, 38*(6), 393–411.

Sorokowski, P., Sorokowska, A., & Witzel, C. (2014). Sex differences in color preferences transcend extreme differences in culture and ecology. *Psychonomic Bulletin & Review, 21*(5), 1195–1201.

Stamm, J. S. (1955). Fourier analyses for curves of affective value of color as functions of hue. *The American Journal of Psychology, 68*(1), 124–132.

Strauss, E. D., Schloss, K. B., & Palmer, S. E. (2013). Color preferences change after experience with liked/disliked colored objects. *Psychonomic Bulletin & Review, 20*(5), 935–943.

Taylor, C., & Franklin, A. (2012). The relationship between color–object associations and color preference: Further investigation of ecological valence theory. *Psychonomic Bulletin & Review, 19*(2), 190–197.

Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General, 123*(4), 394–409.

Wyszecki, G., & Stiles, W. S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae* (2nd ed.). New York: John Wiley & Sons.

Yokosawa, K., Schloss, K. B., Asano, M., & Palmer, S. E. (2016). Ecological effects in cross-cultural differences between US and Japanese color preferences. *Cognitive Science, 40*, 1590–1616.

Zar, J. H. (2009). *Biostatistical analyses* (5th ed.). London: Prentice Hall.