Vision Research xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Vision Research

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

Variations in normal color vision. VII. Relationships between color naming and hue scaling

Kara J. Emery [a], Vicki J. Volbrecht [b], David H. Peterzell [c], Michael A. Webster [a,d,*]

[a] Graduate Program in Integrative Neuroscience, University of Nevada, Reno, Reno, NV 89557, United States
[b] Department of Psychology, Colorado State University, Fort Collins, CO 80523, United States
[c] College of Psychology, John F. Kennedy University, Pleasant Hill, CA 94624, United States
[d] Department of Psychology, University of Nevada, Reno, Reno, NV 89557, United States

ARTICLE INFO

Article history:
Received 1 August 2016
Received in revised form 2 December 2016
Accepted 6 December 2016
Available online xxxx

Keywords:
Color vision
Color naming
Color categories
Color appearance
Hue scaling
Individual differences

ABSTRACT

A longstanding and unresolved question is how observers construct a discrete set of color categories to partition and label the continuous variations in light spectra, and how these categories might reflect the neural representation of color. We explored the properties of color naming and its relationship to color appearance by analyzing individual differences in color-naming and hue-scaling patterns, using factor analysis of individual differences to identify separate and shared processes underlying hue naming (labeling) and hue scaling (color appearance). Observers labeled the hues of 36 stimuli spanning different angles in cone-opponent space, using a set of eight terms corresponding to primary (red, green, blue, yellow) or binary (orange, purple, blue-green, yellow-green) hues. The boundaries defining different terms varied mostly independently, reflecting the influence of at least seven to eight factors. This finding is inconsistent with conventional color-opponent models in which all colors derive from the relative responses of underlying red-green and blue-yellow dimensions. Instead, color categories may reflect qualitatively distinct attributes that are free to vary with the specific spectral stimuli they label. Inter-observer differences in color naming were large and systematic, and we examined whether these differences were associated with differences in color appearance by comparing the hue naming to color percepts assessed by hue scaling measured in the same observers (from Emery et al., 2017). Variability in both tasks again depended on multiple (7 or 8) factors, with some Varimax-rotated factors specific to hue naming or hue scaling, but others common to corresponding stimuli for both judgments. The latter suggests that at least some of the differences in how individuals name or categorize color are related to differences in how the stimuli are perceived.

1. Introduction

The light spectrum varies continuously, yet we typically describe color with a small number of discrete verbal categories. The nature and basis of color categories has been extensively investigated. Anthropological studies have explored how the patterns of color naming vary across different cultures or linguistic groups (Kay, Berlin, Maffi, Merrifield, & Cook, 2009; MacLaury, 1997). There are strong correspondences across languages in how basic color terms are assigned to different regions of color space (Berlin & Kay, 1969; Kay & Regier, 2003; Lindsey & Brown, 2006; Regier, Kay, & Cook, 2005). These "universal" tendencies point to a common basis shaping color categories, perhaps anchored by the basic physiology of color processing in the visual system (Boynton & Olson, 1990; Kay & McDaniel, 1978) or by consistent properties of the visual environment (Jameson & D'Andreade, 1997; Regier, Kay, & Khetarpal, 2007; Yendrikhovskij, 2001). By this account, categories might be strongly constrained by the perceptual organization of color imposed by visual processing or the distribution of color signals in the environment. For example, a stimulus labeled as pure yellow might correspond to the undiluted response of a blue-yellow opponent process inherent to color coding (Hering, 1964), or to a special property of the light environment such as the locus of natural daylight (Mollon, 2006). Consistent with this, categorical responses to color can emerge in infants even before they learn a language (Yang, Kanazawa, Yamaguchi, & Kuriki, 2016). However, languages also differ widely in the number of basic color terms used to partition colors (Berlin & Kay, 1969), and can vary in the foci or best examples for different color terms both across (Pilling & Davies, 2004; Roberson, Davidoff, Davies, &

* Corresponding author at: Department of Psychology/296, University of Nevada, Reno, Reno, NV 89557, United States.
E-mail address: mwebster@unr.edu (M.A. Webster).

http://dx.doi.org/10.1016/j.visres.2016.12.007
0042-6989/© 2016 Elsevier Ltd. All rights reserved.

Shapiro, 2005; Roberson, Davies, & Davidoff, 2000) and within categories (Webster & Kay, 2007; Webster et al., 2002). These differences have emphasized the "cultural relativism" of color naming. This relativism could arise because verbal categories are presumably also shaped by interaction and communication among observers (Jameson & Komarova, 2009; Lindsey, Brown, Brainard, & Apicella, 2015; Steels & Belpaeme, 2005), so that categories are influenced by both perception and language [e.g. (Cibelli, Xu, Austerweil, Griffiths, & Regier, 2016)], as well as a variety of other factors or decision rules at the various levels of representing and categorizing color (Cropper, Kvansakul, & Little, 2013; Parraga & Akbarinia, 2016). Consequently this may weaken the potential links between color perception and color naming. For example, two observers might describe the same stimulus in the same way because of how they have learned to label the stimulus, even if they "see" them differently. Conversely, it is not evident whether people who describe colors differently do so because their percepts differ. In this study, we focused on how color categories are related to color percepts, by examining individual differences in color naming and color appearance.

A variety of methods have been used to measure color appearance with nonverbal responses. These techniques include: 1) hue scaling [e.g. (Gordon, Abramov, & Chan, 1994)], 2) hue similarity ratings [e.g. (Shepard & Cooper, 1992)], 3) hue cancellation [e.g. (Hurvich & Jameson, 1957)], and 4) unique hue measurements (which involves a subset of full cancellation functions) [e.g. (Larimer, 1974; Larimer, Krantz, & Cicerone, 1975)]. In this study, we focused specifically on hue scaling, in which observers decompose stimuli by rating the proportion of red, green, blue or yellow perceived in a stimulus. These hue judgments are for most observers intuitive and highly reliable (Gordon et al., 1994) and directly tap the phenomenal percepts of color, thereby permitting experimenters to study how color appearance varies across different stimulus conditions or observers. The hue-scaling responses are also assumed to reflect the underlying responses or spectral sensitivities of the red-green and blue-yellow opponent processes which remain a cornerstone of theoretical accounts of color perception (Abramov & Gordon, 1994; Werner & Wooten, 1979).

Several studies have explored the links between color naming and color perception. For example, measurements of unique hues (the stimuli that appear pure red, green, blue, or yellow) have been found to be closely similar to the focal colors or exemplars of a color category (Miyahara, 2003; Wooten & Miller, 1997). Furthermore, researchers have explored (with varying results) whether focal colors might be perceptually more salient (Kuehni, Shamey, Mathews, & Keene, 2010; Witzel & Franklin, 2014). A recent actively explored question is whether hues that share a common linguistic color category are perceptually more similar than equivalent stimulus differences that span two categories (Bornstein, 1987). These categorical effects have been examined in a number of tasks and across languages that differ in the number of basic color terms (so that larger differences are predicted in a language that distinguishes two hues than in languages that do not). Studies of categorical perception for color have revealed a number of cases where categories can influence both perception and performance (e.g. in the judged similarity or the time required to discriminate a pair of colors) (Franklin et al., 2008; Gilbert, Regier, Kay, & Ivry, 2006; Mullen & Kulikowski, 1990; Winawer et al., 2007). However, these categorical effects tend to be subtle and labile (Brown, Lindsey, & Guckes, 2011; Witzel & Gegenfurtner, 2011, 2013), and may depend strongly on whether the task used to measure them is limited by the properties of perceptual encoding versus the decision stages of the response or language processing (Kay & Kempton, 1984; Pilling, Wiggett, Ozgen, & Davies, 2003; Roberson & Davidoff, 2000; Roberson, Pak, & Hanley, 2008; Webster & Kay, 2012). Thus it remains unknown to what extent differences in naming patterns are tied to differences in perception. Moreover, in most cases, the aim of this work has been to examine how the category influences the percept, or potential top-down biases on color processing. Here we instead examined how appearance might shape the stimulus choices for the categories.

We explored the relationship between color appearance and color naming by examining how they vary across observers. Individuals with normal color vision differ markedly in how they label colors or rate their appearance (Bimler, Kirkland, & Pichler, 2004; Kuehni, 2004; Malkoc, Kay, & Webster, 2005; Webster, Miyahara, Malkoc, & Raker, 2000). Moreover, these differences are dramatic within a language, and in fact can be far greater than the differences across linguistic groups (Lindsey & Brown, 2009; Webster & Kay, 2007). Here we exploited these individual differences to ask how naming patterns vary and whether they covary with perceptual judgments, using hue scaling as an index of the observer's percepts. In the accompanying paper, we focus on the individual differences in hue scaling and what they reveal about the perceptual representation of color (Emery, Volbrecht, Peterzell, & Webster, 2017). In this paper, we instead consider what the inter-observer differences in hue-scaling functions imply about how individuals verbally partition color space, and what the patterns of these partitions themselves suggest about the structure of color appearance. As with the accompanying paper (Emery et al., 2017), we used correlational and factor-analytic methods to elucidate these systematic patterns of individual differences.

2. Methods

Observers participated in three experimental sessions that included color naming and hue scaling. The general methods are detailed in the accompanying paper (Emery et al., 2017), and summarized here.

2.1. Participants

The same observers completed the hue-naming and hue-scaling tasks. They included 26 adult university students, all with normal color vision and normal or corrected-to-normal acuity. An additional observer gave highly variable settings and was excluded from the analysis. Participation was with informed consent and followed protocols approved by the University of Nevada Institutional Review Board and conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

2.2. Apparatus and stimuli

Stimuli were displayed on a calibrated SONY 500PS CRT monitor controlled by a Cambridge Research System ViSaGe Stimulus Generator. The stimulus was a 2-deg uniform square displayed in the center of the screen, and shown on an achromatic 11.3 by 8.5-deg background with the same photometric luminance ($20\,cd/m^2$) and the chromaticity of Illuminant C. The observer viewed the display binocularly in an otherwise dark room from a distance of 200 cm, and responded with a handheld keypad.

The stimuli consisted of 36 chromaticities spaced in 10-deg steps along a circle of fixed contrast in cone-opponent space (Derrington, Krauskopf, & Lennie, 1984; MacLeod & Boynton, 1979), scaled to roughly equate contrasts along the chromatic cardinal axes of the space (LvsM, corresponding to angles of 0–180 deg, and SvsLM, corresponding to angles of 90–270 deg) (Fig. 1a). The scaled LvsM and SvsLM contrasts were related to the l_{mb} and s_{mb} coordinates in the MacLeod-Boynton chromaticity diagram by:

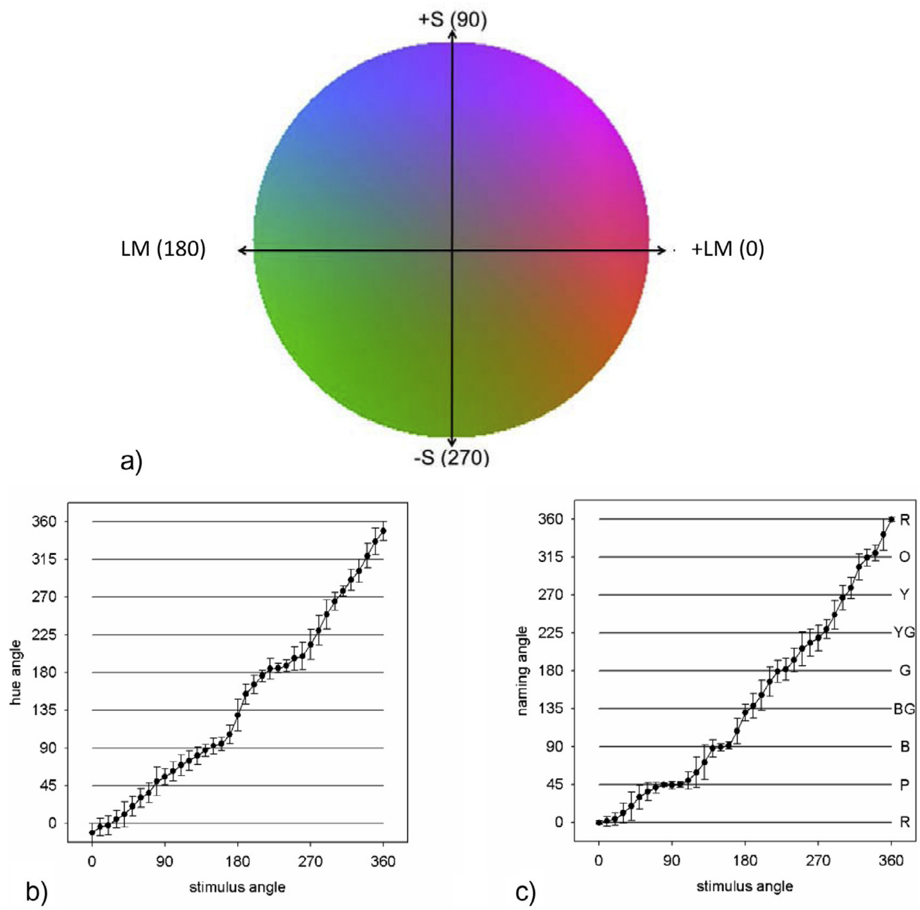

Fig. 1. Hue-scaling and color-naming functions. (a) The stimulus space with angles defined by directions within the LM and S cone-opponent plane. (b) Average hue angles (±1 sd) for 26 observers, expressing the relative blue-yellow (90–270°) to red-green (0–180°) proportions from scaling, as a function of the stimulus angle. (c) Average naming angles (±1 sd), showing the terms used to describe each stimulus angle, with the terms defined as 45° increments in the red-green vs. blue-yellow plane.

$$\mathrm{LvsM} = (l_{mb} - 0.6568) * 2754$$

$$\mathrm{SvsLM} = (s_{mb} - 0.01825) * 4099,$$

with the achromatic point at the chromaticity of illuminant C.

2.3. Procedure

Testing was conducted across three sessions lasting less than one hour each. The first session included color screening with the Cambridge Colour Test, measurements of relative sensitivity to the LM and S chromatic directions, and two hue-naming tasks. The first naming task involved open-ended labeling of the stimuli in which observers were free to use any name, while the second restricted the available names. Only the procedure for the latter hue-naming task is described here, since the unrestricted names were not amenable to the present analysis (which required representing the names on a numerical scale). In the restricted-naming task, observers were shown the 36 hues in random order and were asked to label the stimuli using a set of eight color terms corresponding to the unique hues (red, green, blue, or yellow) or binary hues (purple, blue-green, yellow-green, or orange). The stimulus was shown as a 500-ms pulse interleaved with 1 s of the gray background. The eight color terms were displayed continuously at the bottom of the screen, and the observers responded by using a handheld keypad to scroll through and highlight the selected term. After responses had been collected for each of 36 stimuli, the stimuli were again presented in random order. This cycle continued until four responses were obtained for each stimulus. Except where noted, results reported are based on the observer's average responses to each stimulus.

In the last two sessions, observers scaled the hue for each of the 36 stimuli which they named. The hue-scaling task required each observer to judge the relative percentage of red, green, blue or yellow present in each stimulus using a handheld keypad to adjust percentages at the bottom of the screen which corresponded to the four primaries. The percentages were required to sum to 100%. The percentages were then converted into a hue angle within a perceptual red-green (0–180°) versus blue-yellow (90–270°) opponent-space:

$$\text{hue angle} = \tan^{-1}[(\text{blue-yellow})/(\text{red-green})].$$

For example, if an observer responded that a stimulus appeared 50% red and 50% yellow, the hue angle would be 315°. Four responses were again obtained for each stimulus, and the results reported are based on the observers' average responses. For a detailed description of the hue-scaling method, see Emery et al. (2017).

3. Results

3.1. Naming vs scaling: Average responses

Fig. 1 summarizes both the hue-naming and hue-scaling results. The top panel illustrates the LM vs S stimulus space, in which observers again judged the hue of 36 stimuli corresponding to

Please cite this article in press as: Emery, K. J., et al. Variations in normal color vision. VII. Relationships between color naming and hue scaling. *Vision Research* (2017), http://dx.doi.org/10.1016/j.visres.2016.12.007

different directions in the space. We refer to these variables as the *stimulus angles* of the set. The scaling and labeling of the stimulus angles in most cases varied monotonically with the stimulus angle. To map the relationships between the stimuli and rated percepts, in the case of scaling, the responses were converted into hue angles within a perceptually-based color space defined by a red-green (0–180°) and blue-yellow (90–270°) axis. The responses are thus represented by the *hue angle* within this space. Fig. 1b shows the hue angles averaged across the 26 observers (with the individual settings shown in Fig 1b in the accompanying paper). The standard deviations correspond to the differences between observers.

In order to compare the results from hue scaling and hue naming, the restricted color terms were represented using the same convention. In this case the eight terms were denoted values in steps of 45 deg from 0 deg (red) to 315 deg (orange). Consistent with the perceptual space, this places the labels for the unique hues along the axes of the space, and for the binary hues at angles midway between the adjacent unique hues [e.g. so that purple (45 deg) is between red (0 deg) and blue (90 deg)]. The settings for each observer were calculated as the mean of the four responses. (For example, an observer who called a stimulus red twice and purple twice had a response of 22.5 deg.) We refer to the resulting responses as the *naming angles*.

The average naming angles for all observers from the color-naming task are shown in the final panel of the figure (Fig. 1c), with the error bars again showing the standard deviation in the names across observers. Note that the curve for the naming should follow more of a staircase than the scaling since the responses represent eight discrete steps. This tendency is clear in the figure but blurred by both the differences between observers and inconsistent responses within observers, e.g. near the category boundaries.

3.2. Hue naming: Relationships between color categories assessed by correlational and factor analyses

Before examining how hue scaling and hue naming are related, we first consider some of the properties and implications of the naming responses themselves. From the naming functions for each observer we estimated the regions of the cone-opponent plane that were labeled by a given hue term (red, purple, blue, green, yellow, orange) or hue combination (blue-green, green-yellow). This was done by smoothing the function by fitting it with an 8th-order polynomial, and then estimating the stimulus angle corresponding to the category boundary between adjacent colors (e.g. the stimulus angle at which the response was 22.5° for the red-purple boundary). The polynomial was chosen to preserve the overall shape of the individual's hue-scaling function while dampening local variability which would otherwise dominate boundary estimates based on a more local interpolation such as a spline fit. These boundaries are shown in Fig. 2 for the 26 observers, and illustrate the substantial differences in how these observers categorized the stimuli using the eight supplied names. Notably, these differences were largest for the yellow-green/yellow and red-purple boundaries, while smaller for red-orange or blue/blue-green boundaries, differences that might partly reflect the relative proximity of adjacent focal colors within the cone-opponent plane.

If the labeling directly reflected variations in a small number of dimensions (e.g. red-green or blue-yellow) then the inter-observer differences in naming might be expected to show covariations across the different categories. For example, an observer whose unique red was rotated clockwise (toward purple) might have category boundaries for red-purple and red-orange also rotated, so that the two boundaries chosen for the red term are positively correlated. Moreover, if this rotation is a change in a linked red-green opponent axis, then these boundaries should also covary with their boundaries for green. Alternatively, in an observer for whom the red-green axis is relatively more salient compared to blue-yellow, the category boundaries for red (and green) should both be expanded and thus negatively correlated.

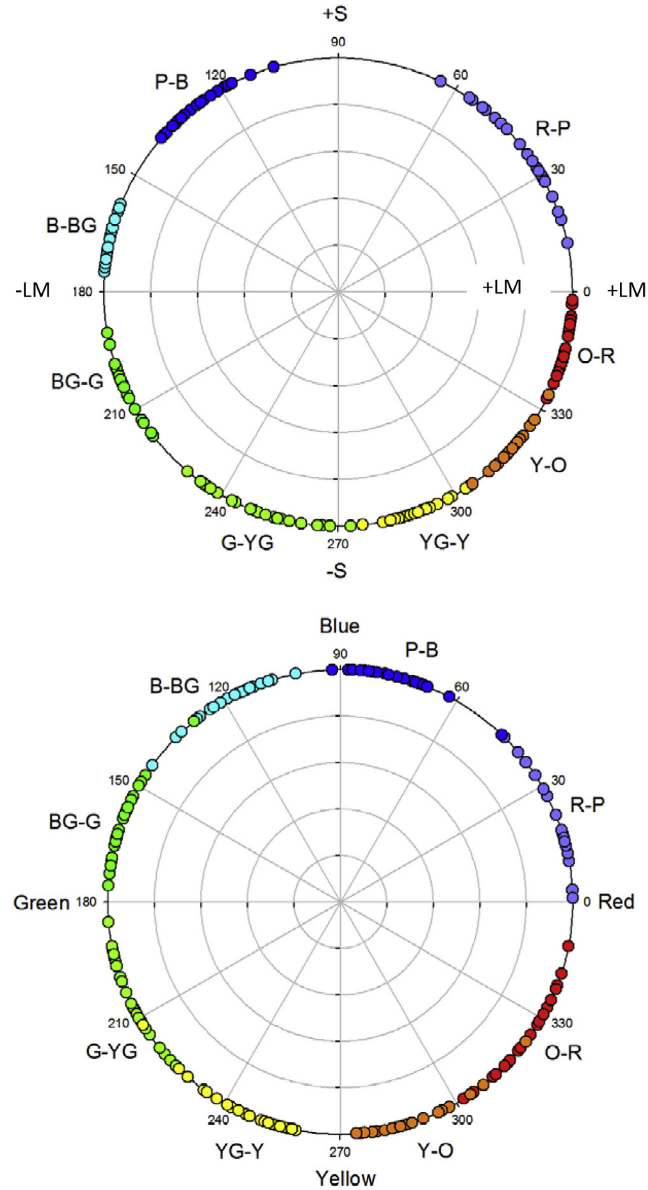

Fig. 2. Category boundaries for the color terms for individual observers, (a) shown by their angle within the LM vs S cone-opponent space, or (b) within the red-green vs. blue-yellow perceptual-opponent space. Points plot the estimated transition (50% probability of naming) between successive adjacent categories (e.g. O-R orange-red boundary).

Table 1 shows that the actual correlations of naming boundaries are very different from the above predictions. After a Bonferroni correction for multiple comparisons (experimentwise error $p < 0.05$), only one of the 28 pairs of boundaries were significantly correlated. This corresponded to the two boundaries demarcating blue-green from blue or green, and was positive, suggesting that observers differed by a rotation in the stimulus angles labeled as blue-green. There was also a suggestive positive correlation between the blue-green/green boundary and yellow-green/yellow, and negative between the orange-red and yellow-green/yellow boundaries. However, a more striking feature is that variations in the remaining categories were largely independent, and there is little sign of a relationship between the categories of the opposite opponent pairs. Moreover, apart from blue-green, the

Table 1
Correlations between the category boundaries for different color terms, based on the boundary angles within the LM vs. S space.

	O-R	R-P	P-B	B-BG	BG-G	YG-G	YG-Y	Y-O
O-R	1	0.005	0.061	−0.174	−0.286	−0.396	−0.515	0.433
R-P		1	0.088	0.414	0.267	−0.067	−0.137	−0.323
P-B			1	0.233	−0.238	−0.127	−0.182	−0.379
B-BG				1	0.649*	−0.064	0.045	−0.295
BG-G					1	0.527	0.221	−0.237
G-YG						1	0.451	−.131
YG-Y							1	−0.073
Y-O								1

* Significant after Bonferroni correction for multiple comparisons, $p < 0.002$.

two bounding sides of a single category were largely unrelated. Thus whether an individual is more likely to call a stimulus red than purple did not predict whether they would label a stimulus as red or orange.

To further assess these variations, we applied a factor analysis to the full color-naming functions for the observers, based on the naming angles for each of the 36 stimulus angles. Factors were extracted using principal component analysis (because the matrix was not positive definite), and rotated with the Varimax criterion to approximate simple structure. (An oblique rotation using the Direct Oblimin criterion yielded a similar pattern with only weak correlations between factors, suggesting the actual factors are largely orthogonal as assumed by the Varimax rotation.) As described in Emery et al. (2017), factors were included for rotation based on systematically-tuned loadings across adjacent variables. This resulted in 5–6 factors (60% of the variance), with loadings shown in Fig. 3. These factors tend to be monopolar (e.g. with significant loadings sharing the same sign), a pattern which is again more indicative of potential rotations in the locations of the color categories rather than changes in their angular subtense. While variable, some of the factors clearly exhibited only a single and relatively circumscribed peak. This is consistent with the weak correlations between different categories as illustrated in Table 1, and again suggests that the partitions that observers selected for labeling the colors were controlled by multiple processes, some of which are narrowly tuned for specific boundaries.

In the accompanying paper (Emery et al., 2017), we detailed how these multiple and narrowly-tuned factors are inconsistent with conventional color-opponent models based on two underlying red-green and blue-yellow dimensions, and instead point to higher-dimensional representations of color where circumscribed regions of color space are perceived in ways that may be coded independently. The narrow factors and weak intercorrelations observed here point to similar conclusions, and suggest that even in color naming, there is little evidence that the category boundaries are mediated in a deterministic or pervasive way by an underlying red-green and blue-yellow representation of color.

3.3. Relationships between color naming and hue scaling: Correlational and factor analyses of combined hue naming and scaling data

In the next set of analyses, we examined how hue naming and hue scaling were related. Again, our aim was to ask whether differences in how the colors were labeled could be tied to differences in how the colors were perceived. For example, differences in category boundaries could occur because individuals differed in their rules for assigning color names. Thus one observer might require more yellow added to red before they will assign the term "orange." Such differences should in theory show less relation to hue scaling because the variations in strategy are unrelated to the underlying percepts. Alternatively, the categories could differ

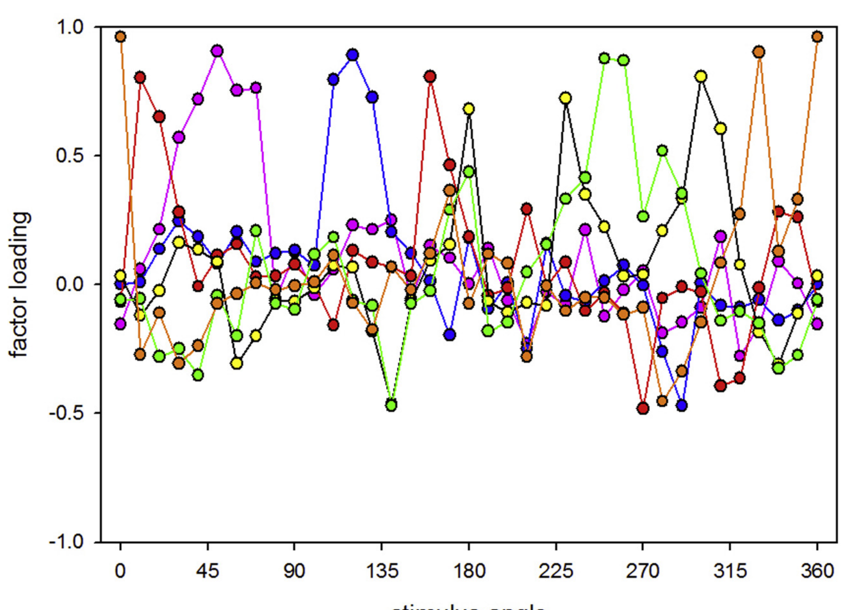

Fig. 3. Varimax-rotated factor loadings for the first 6 factors extracted from PCA of the color-naming choices for the 26 observers. This factor analysis was computed on the responses from color naming alone.

Please cite this article in press as: Emery, K. J., et al. Variations in normal color vision. VII. Relationships between color naming and hue scaling. *Vision Research* (2017), http://dx.doi.org/10.1016/j.visres.2016.12.007

because two observers both apply the same rule (e.g. the same proportion of yellow must be perceived), but differ in the stimulus angles required to elicit these percepts. In this case the functions for labeling should be tied to the scaling.

We again assessed this in various ways. If perceptual differences did indeed contribute to the variations in color categories, then these variations should be reduced when the categories are expressed using the observers' underlying hue-scaling functions. That is, we should see less spread compared to the category boundaries shown in the stimulus-based, cone-opponent space of Fig. 2a, when the boundaries are instead plotted in terms of the red-green and blue-yellow dimensions of the perceptually-based color space. To examine this, the hue angles corresponding to each individual's category boundaries were estimated (by again fitting an 8th-order polynomial to their hue-scaling function, and then evaluating the function at the stimulus angles corresponding to their boundaries). These are shown in the second panel of Fig. 2. However, this transformation did not significantly reduce the variability in the settings (and instead slightly increased the average standard deviation of the boundaries, from 9.1 deg for the stimulus angles to 10.7 deg for the hue angles). In other words, the differences in labeling are not a simple and direct consequence of the differences in hue scaling.

The representation of naming angles in terms of the hue angles is also of interest with regard to how different categories are defined. For example, how much yellow or blue must be introduced to a red before it starts to be called orange or purple? If the primary colors (red, green, blue and yellow) required fairly pure stimulation to warrant a label, then these labels should occupy smaller arcs of the perceptual color space than the binary hues. In contrast, if the names are equally deployed, then the span of each name should cover 45 deg. The average arcs ranged from 38 deg for red to 57 deg for purple, with otherwise little to suggest that binary hues filled more or less of the space than the primary hues. Thus the different naming categories appear to be used with similar salience.

As a second test of the relationship between appearance and naming, we compared the correlations between the hue-scaling and color-naming functions. This was done by correlating the deviations of each observer from the mean at each stimulus angle and for each task, with the correlations assessed between pairs of observers rather than pairs of stimuli. Suppose that naming for all of the stimuli was determined by a consistent rule based on the perceived proportions of red, green, blue or yellow. In this case we should expect that correlations between the tasks should be higher when compared for the same observer than two different observers. The cross-task correlation for the same observer was weak, averaging $r = 0.276$, and varied widely across observers (ranging from $r = -0.09$ to 0.63). This average was nevertheless substantially greater than the average correlation between the observers ($r = -0.014$) a difference that was highly significant ($t(349) = 6.15$, $p = 1E-9$). Thus by this criterion there were suggestive but potentially observer-dependent relationships between the scaling and naming.

The preceding analyses argue against a strong global link between individual differences in scaling and naming, in which the two are related by a common and consistent rule applied across the space, at least when pooled across all observers. However, as we showed in the previous section, hue naming as well as hue scaling do not themselves appear to vary due to global or broadly tuned factors. Thus there still might be correspondences between the two tasks but in ways that might vary idiosyncratically for different stimulus regions. To test this, we performed a further factor analysis, this time including both data sets as the observed variables (so that there were now a total of 72 items and 26 observers). Factors were extracted through principal component analysis and rotated with the Varimax criterion. (As before similar results with largely orthogonal factors were also obtained by an oblique rotation.) The analysis produced 7–8 systematically-tuned factors which collectively accounted for 66% of the variance across the two tasks. The loadings for these factors are plotted in Fig. 4, with each panel showing a single factor. The two curves within each plot show how the single factor loaded on either hue angles (solid curve) or naming angles (dashed) again as a function of the stimulus angle.

The factor loadings are in some cases varied and complex, but there is a trend for the factors to load only on a small subset of the variables and to roughly tile the space of different stimulus angles. Importantly, some of the factors tend to load on similar stimulus angles for both of the tasks. For example, this is readily apparent for the first factor (Factor 1 in Fig. 4), which accounted for the most variance (11.8%) in the pooled dataset. This factor is characterized by a very similar pattern for both the hue-scaling and hue-naming data, loading primarily on stimulus angles in the yellow-green region of the space. The polarity of the loadings is arbitrary, and is such that an observer with a higher score on this factor tends to scale the yellow-green stimuli as more yellowish and also is more likely to label them as yellow. Suggestive correspondences also occur for some of the other factors. For example, Factors 5 and 7 each exhibit a narrow peak roughly around the two boundaries for blue, again with similar peaks for the hue-naming and hue-scaling responses. To assess these correspondences, we calculated the correlations between the factor loadings for the two tasks. The resulting matrix is given in Table 2. For Factors 1 and 5, there is a strong relation between the two tasks ($r = 0.76$ and 0.70 respectively, both significant after a Bonferroni correction for multiple comparisons with experimentwise error set at $p < 0.05$), and 4 of the 5 highest positive correlations are between the naming and scaling loadings within the same factors. Thus, these results suggest that at least some aspects of how the stimuli were categorized are significantly related and potentially attributable to individual differences in how they were scaled. On the other hand, in several other cases (e.g. factors 3, 4, or 6) there is little relationship between the pattern for scaling and naming, including factors which tend to load only on the variables for one of the tasks; and as noted, where correspondences do occur they tend to affect only localized stimulus angles. This indicates that the relationships between naming and appearance are complex, and potentially governed by different rules for different chromatic regions, a pattern which is again consistent with the large number of factors required to account for either the scaling or category boundaries taken on their own.

4. Discussion

That individuals might describe the colors of the same stimulus differently was recently brought to the fore by the worldwide debate over the color of #thedress (Brainard & Hurlbert, 2015). But despite the surprised reactions to this image, such differences are not uncommon. Here we asked whether these differences in naming can be tied to differences in appearance, or whether they instead reflect differences in how observers label their percepts. In the accompanying paper, we analyzed individual differences in hue scaling, and showed that these differences depend on multiple factors that each influence scaling over a narrow range of chromatic directions (Emery et al., 2017). We further showed that this pattern is inconsistent with conventional accounts of color appearance, in which hue is represented by the relative responses in two underlying dimensions encoding red-green and blue-yellow. Instead, we argued for a representation in terms of multiple mechanisms, in which there is no need for a fixed relationship between

Please cite this article in press as: Emery, K. J., et al. Variations in normal color vision. VII. Relationships between color naming and hue scaling. *Vision Research* (2017), http://dx.doi.org/10.1016/j.visres.2016.12.007

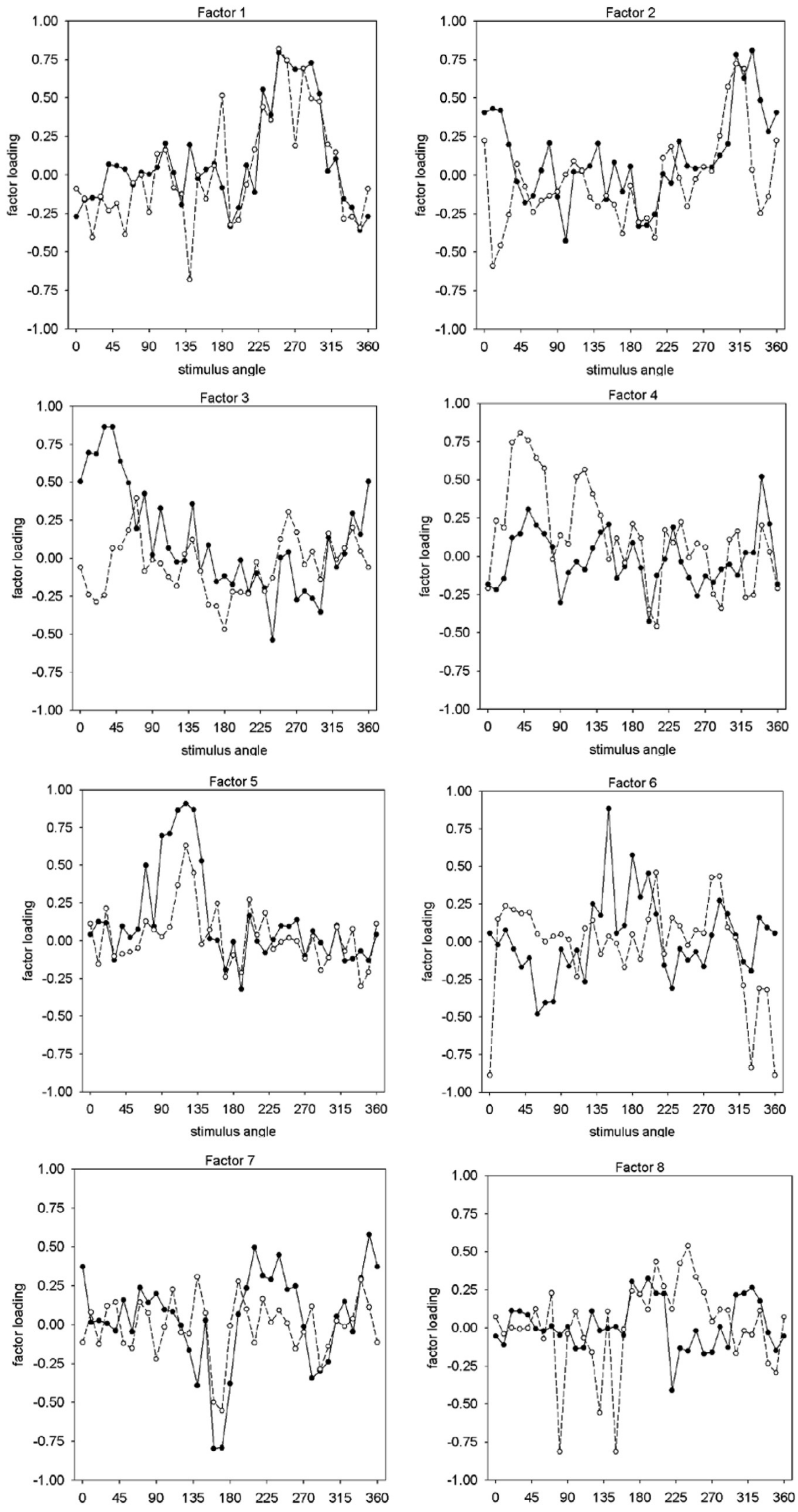

Fig. 4. Factor analysis including responses from both hue scaling and color naming. Panels plot the loadings for the first 8 components following Varimax rotation, with each individual factor shown in a separate panel. Solid lines show the loadings of the factor on the stimulus angles for the hue-scaling task, while dashed lines show the loadings of the factor on the stimulus angles for the hue-naming task.

Please cite this article in press as: Emery, K. J., et al. Variations in normal color vision. VII. Relationships between color naming and hue scaling. *Vision Research* (2017), http://dx.doi.org/10.1016/j.visres.2016.12.007

Table 2
Correlations between the factor loadings for scaling and naming, for the factors shown in Fig. 4. Highlighted cells show the correlations between the loadings of the same factor on the two different tasks.

	Scaling							
Naming	factor 1	factor 2	factor 3	factor 4	factor 5	factor 6	factor 7	factor 8
factor 1	0.760*	-0.036	-0.527	-0.256	-0.114	-0.003	-0.026	-0.187
factor 2	0.324	0.369	-0.300	-0.010	-0.039	-0.116	0.040	0.028
factor 3	0.245	0.180	0.107	0.258	0.148	-0.385	0.300	-0.340
factor 4	-0.178	-0.123	0.529	0.451	0.343	-0.344	-0.048	-0.142
factor 5	-0.139	-0.087	-0.032	-0.324	0.696*	-0.099	0.010	-0.085
factor 6	0.391	-0.457	-0.078	-0.138	0.106	0.092	-0.136	-0.020
factor 7	-0.209	0.114	0.153	0.339	0.091	0.013	0.468	-0.126
factor 8	0.309	-0.153	-0.230	-0.330	-0.148	-0.157	0.068	0.057

color assignments and the underlying neural activity. In the present work, we first asked the question of whether this pattern also persists in the color-naming task. The results suggest that the individual differences in color naming behave similarly, with the basic color categories varying predominantly in an independent manner and with the two boundaries defining a single category often uncorrelated. Thus there is again no evidence for an opponent-chromatic code or one where some color terms are more special while others more derivative. This is in spite of the fact that some of the color terms used, like blue-green, were derived from explicitly combining the basic color categories. The finding that both the unique and binary color categories tend to float freely relative to each other points again to a representation where different color categories behave more or less like qualitatively distinct objects, rather than quantitative or metrical variations emerging from an underlying scaffolding in terms of the conventional red-green and blue-yellow dimensions of color-opponent theory. This is not inconsistent with a basic perceptual organization of color in terms of red-green and blue-yellow, or with the relationships between colors implied by measurements such as similarity ratings (Indow, 1988; Shepard & Cooper, 1992). Clearly most observers would report that an orange hue is more similar to red than it is to green. However, while there were some correlations between color categories, the overall pattern we observed suggests that what constitutes "orange" cannot be reduced to a small number of rigidly prescribed rules dictated by red-green and blue-yellow responses.

Note also that present results were restricted to nominally equiluminant hues of fixed saturation. It remains to be examined whether other attributes of color vary in similar ways. For example, conventional color models treat lightness and hue as largely separable dimensions, but it is well known that yellow exists only as a lighter color while decrements instead appear brown, and recent results suggest that the nulls for unique yellow and brown are different (Buck, 2015; Vincent, Kale, & Buck, 2016). However, it is not known whether individual differences in the hues of increments and decrements vary independently, nor how color naming or scaling might differ between observers as a function of saturation. We are currently exploring these issues.

The second question we addressed is how the variations in color naming and color appearance are related, and specifically whether individual differences in naming could be tied to differences in percepts. As we noted in the Introduction, there is a very large literature on color naming and its potential links to perception, with evidence for both strong universal tendencies [e.g. (Lindsey & Brown, 2006; Regier et al., 2005)] as well as large individual [e.g. (Kuehni, 2004; Lindsey & Brown, 2009; Webster et al., 2000)] and cultural differences [e.g. (Roberson et al., 2000)]. Our focus was on individuals with a common culture and language, and was restricted to comparing the individual differences in color naming and hue scaling. This has the advantage that we could assess normal variations across tasks that could be readily compared. However, it is first important to reconsider what each task is measuring. With regard to scaling, a conceptual issue is to what extent hue scaling itself is an actual measure of appearance, for it could be argued that the task is a variant on color naming where the observers are judging the salience of a small set of categories. Moreover, these judgments show some susceptibility to categorical biases with regard to the primaries used for scaling, and these biases could reflect an intrusion of linguistic coding (Webster & Kay, 2012). Nevertheless, responses from hue-scaling experiments reveal spectral sensitivities that closely resemble the functions measured by more nonverbal tasks such as hue cancellation [e.g. (Gordon et al., 1994; Werner & Wooten, 1979)], and these sensitivities do vary continuously with the chromatic angle of the stimulus. Thus it seems that the hue-scaling procedure provides a reasonable proxy for at least some aspects of the observers' color percepts. With regard to color naming, it is also important to ask whether the restricted-term task is a valid measure of naming, and not simply a variant on scaling, where observers are responding whether the stimuli appeared to be predominantly one of the basic color primaries or a mixture (Boynton, Schafer, & Neun, 1964). The constrained terms we used were important for allowing categories that were well understood by all observers and for allowing a common metric for the analysis (i.e., hue and naming angle). However, not all terms (e.g., orange and purple) made explicit references to the scaling primaries. Moreover, the terms appeared adequate to fully and intuitively represent the space, such that observers could readily label all of the stimuli with the eight terms provided. Thus, the hue-naming task used in our study captures important aspects of color naming as it is normally engaged. Finally, whether the two tasks measure different things is perhaps best answered by our finding that they did in part lead to different patterns of inter-observer variability.

Our results suggest that individual differences in hue naming cannot be simply reduced to differences in how the colors are perceived. Specifically, there was no indication of a strong global correspondence between the color naming and hue scaling, and the variations included factors that loaded primarily on only one of the tasks. This dissociation may perhaps seem obvious in that color naming must include shared knowledge and learning through language, while color appearance, except in the strongest Whorfian sense, is constrained more by the physiology or experience of the individual. Further, it is possible that the necessarily categorical nature of naming reflects representations that are fundamentally distinct from the form of the representations for a continuous perceptual dimension. More surprising is that, conversely, we also found evidence that color naming and hue scaling are susceptible to common influences, for a number of factors loaded on corresponding stimuli for both tasks. We emphasize again that these tasks were measured on different days and, as described above, required ostensibly different judgments about the stimuli. Like the variations in the individual tasks, the relationships between the two tasks tended to be confined to restricted chromatic angles, with the most prominent correspondences around yellowish-green and blue regions of the color space. This complexity may again reflect the observation that both the hue scaling and hue naming varied in different ways across different chromatic regions. This is in line with similarity ratings that have also revealed the influence of multiple dimensions which may differentially affect different subsets of hues (Chang & Carroll, 1980; Komarova & Jameson, 2013). We did not attempt to model the specific relationship between scaling and naming because it would thus require different rules for different regions of color space. However, at least for some of these chromatic regions, our results add to the debate over categorical influences on color perception, which has focused on whether perceptual differences occur within or between different color categories, by pointing to suggestive links between how stimuli are perceived and to which verbal categories they are assigned. Specifically, they raise the possibility that two observers who label stimuli differently, may sometimes do so because they see them differently.

Acknowledgments

Supported by EY-10834 (MW).

References

Abramov, I., & Gordon, J. (1994). Color appearance: On seeing red–or yellow, or green, or blue. *Annual Review of Psychology, 45*, 451–485.
Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.
Bimler, D., Kirkland, J., & Pichler, S. (2004). Escher in color space: Individual-differences multidimensional scaling of color dissimilarities collected with a gestalt formation task. *Behavior Research Methods, Instruments, & Computers, 36*(1), 69–76.
Bornstein, M. H. (1987). Perceptual categories in vision and audition. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 287–300). Cambridge: Cambridge University Press.
Boynton, R. M., & Olson, C. X. (1990). Salience of chromatic basic color terms confirmed by three measures. *Vision Research, 30*(9), 1311–1317.
Boynton, R. M., Schafer, W., & Neun, M. E. (1964). Hue-wavelength relation measured by color-naming method for three retinal locations. *Science, 146*, 666–668.
Brainard, D. H., & Hurlbert, A. C. (2015). Colour vision: Understanding #thedress. *Current Biology, 25*(13), R551–R554.
Brown, A. M., Lindsey, D. T., & Guckes, K. M. (2011). Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *Journal of Vision, 11*(12–12), 11–20.
Buck, S. L. (2015). Brown. *Current Biology, 25*(13), R536–R537.
Chang, J. J., & Carroll, J. D. (1980). Three are not enough: An INDSCAL analysis suggesting that color space has seven (£1) dimensions. *Color Research and Application, 5*(4), 193–206.
Cibelli, E., Xu, Y., Austerweil, J. L., Griffiths, T. L., & Regier, T. (2016). The Sapir-Whorf hypothesis and probabilistic inference: evidence from the domain of color. *PLoS One, 11*(7) e0158725.
Cropper, S. J., Kvansakul, J. G., & Little, D. R. (2013). The categorisation of non-categorical colours: A novel paradigm in colour perception. *PLoS One, 8*(3) e59945.
Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *Journal of Physiology, 357*, 241–265.
Emery, K., Volbrecht, V. J., Peterzell, D. H., & Webster, M. A. (2017). Variations in normal color vision. VI. Factors underlying individual differences in hue scaling and their implications for models of color appearance. *Vision Research*. http://dx.doi.org/10.1016/j.visres.2016.12.006.
Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R., Kay, P., & Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National academy of Sciences of the United States of America, 105*(9), 3221–3225.
Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National academy of Sciences of the United States of America, 103*(2), 489–494.
Gordon, J., Abramov, I., & Chan, H. (1994). Describing color appearance: Hue and saturation scaling. *Percept Psychophys, 56*(1), 27–41.
Hering, E. (1964). *Outlines of a theory of the light sense*. Cambridge: Harvard University Press.
Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review, 64*(Part 1 (6)), 384–404.
Indow, T. (1988). Multidimensional studies of Munsell color solid. *Psychological Review, 95*(4), 456–470.
Jameson, K. A., & D'Andreade, R. G. (1997). It's not really red, green, blue, yellow: An inquiry into perceptual color space. In C. L. Hardin & L. Maffi (Eds.), *Color categories in thought and language* (pp. 295–319). Cambridge: Cambridge University Press.
Jameson, K. A., & Komarova, N. L. (2009). Evolutionary models of color categorization. I. Population categorization systems based on normal and dichromat observers. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 26*(6), 1414–1423.
Kay, P., Berlin, B., Maffi, L., Merrifield, W. R., & Cook, R. (2009). *The world color survey*. Stanford: CSLI.
Kay, P., & Kempton, W. M. (1984). What is the Sapir-Whorf hypothesis? *American Anthropologist, 86*, 65–79.
Kay, P., & McDaniel, C. K. (1978). The linguistic significance of the meanings of basic color terms. *Language, 54*, 610–646.
Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National academy of Sciences of the United States of America, 100*(15), 9085–9089.
Komarova, N. L., & Jameson, K. A. (2013). A quantitative theory of human color choices. *PLoS One, 8*(2) e55986.
Kuehni, R. G. (2004). Variability in unique hue selection: A surprising phenomenon. *Color Research and Application, 29*, 158–162.
Kuehni, R. G., Shamey, R., Mathews, M., & Keene, B. (2010). Perceptual prominence of Hering's chromatic primaries. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 27*(2), 159–165.
Larimer, J. (1974). Opponent-process additivity–I: Red-green equilibria. *Vision Research, 14*(11), 1127–1140.
Larimer, J., Krantz, D. H., & Cicerone, C. M. (1975). Opponent process additivity. II. Yellow/blue equilibria and nonlinear models. *Vision Research, 15*(6), 723–731.
Lindsey, D. T., & Brown, A. M. (2006). Universality of color names. *Proceedings of the National academy of Sciences of the United States of America, 103*(44), 16608–16613.
Lindsey, D. T., & Brown, A. M. (2009). World Color Survey color naming reveals universal motifs and their within-language diversity. *Proceedings of the National academy of Sciences of the United States of America, 106*(47), 19785–19790.
Lindsey, D. T., Brown, A. M., Brainard, D. H., & Apicella, C. L. (2015). Hunter-gatherer color naming provides new insight into the evolution of color terms. *Current Biology, 25*(18), 2441–2446.
MacLaury, R. E. (1997). *Color and cognition in Mesoamerica: Constructing categories as vantages*. USA: University of Texas Press.
MacLeod, D. I., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America, 69*(8), 1183–1186.
Malkoc, G., Kay, P., & Webster, M. A. (2005). Variations in normal color vision. IV. Binary hues and hue scaling. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 22*(10), 2154–2168.
Miyahara, E. (2003). Focal colors and unique hues. *Perceptual and Motor Skills, 97*(3 Pt 2), 1038–1042.
Mollon, J. D. (2006). Monge (The Verriest Lecture). *Visual Neuroscience, 23*, 297–309.
Mullen, K. T., & Kulikowski, J. J. (1990). Wavelength discrimination at detection threshold. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 7*(4), 733–742.
Parraga, C. A., & Akbarinia, A. (2016). NICE: A computational solution to close the gap from colour perception to colour categorization. *PLoS One, 11*(3) e0149538.
Pilling, M., & Davies, I. R. (2004). Linguistic relativism and colour cognition. *British Journal of Psychology, 95*(Pt 4), 429–455.
Pilling, M., Wiggett, A., Ozgen, E., & Davies, I. R. (2003). Is color "categorical perception" really perceptual? *Memory & Cognition, 31*(4), 538–551.
Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National academy of Sciences of the United States of America, 102*(23), 8386–8391.
Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National academy of Sciences of the United States of America, 104*(4), 1436–1441.
Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition, 28*(6), 977–986.
Roberson, D., Davidoff, J., Davies, I. R., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology, 50*(4), 378–411.
Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology: General, 129*(3), 369–398.
Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition, 107*(2), 752–762.
Shepard, R. N., & Cooper, L. A. (1992). Representation of colors in the blind, color-blind, and normally sighted. *Psychological Science, 3*(2), 97–104.
Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: a case study for colour. *Behavioural Brain Science, 28*(4), 469–489 (Discussion 489–529).
Vincent, J., Kale, A. M., & Buck, S. L. (2016). Luminance-dependent long-term chromatic adaptation. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 33*(3), A164–A169.
Webster, M. A., & Kay, P. (2012). Color categories and color appearance. *Cognition, 122*(3), 375–392.

Webster, M. A., & Kay, P. (2007). Individual and population differences in focal colors. In R. E. MacLaury, G. V. Paramei, & D. Dedrick (Eds.), *Anthropology of color* (pp. 29–53). Amsterdam: John Benjamins.

Webster, M. A., Miyahara, E., Malkoc, G., & Raker, V. E. (2000). Variations in normal color vision. II. Unique hues. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 17*(9), 1545–1555.

Webster, M. A., Webster, S. M., Bharadwaj, S., Verma, R., Jaikumar, J., Madan, G., & Vaithilingham, E. (2002). Variations in normal color vision. III. Unique hues in Indian and United States observers. *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 19*(10), 1951–1962.

Werner, J. S., & Wooten, B. R. (1979). Opponent chromatic mechanisms: Relation to photopigments and hue naming. *Journal of the Optical Society of America, 69*(3), 422–434.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National academy of Sciences of the United States of America, 104*(19), 7780–7785.

Witzel, C., & Franklin, A. (2014). Do focal colors look particularly "colorful"? *Journal of the Optical Society of America A: Optics, Image Science, and Vision, 31*(4), A365–A374.

Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision, 11*(12–16), 11–24.

Witzel, C., & Gegenfurtner, K. R. (2013). Categorical sensitivity to color differences. *Journal of Vision, 13*(7), 1.

Wooten, B., & Miller, D. L. (1997). The psychophysics of color. In C. L. Hardin & L. Maffi (Eds.), *Color categories in thought and language* (pp. 59–88). Cambridge: Cambridge University Press.

Yang, J., Kanazawa, S., Yamaguchi, M. K., & Kuriki, I. (2016). Cortical response to categorical color perception in infants investigated by near-infrared spectroscopy. *Proceedings of the National academy of Sciences of the United States of America, 113*(9), 2370–2375.

Yendrikhovskij, S. N. (2001). Computing color categories from the statistics of natural images. *Journal of Imaging Science and Technology, 45*, 409–417.