



ELSEVIER

Contents lists available at SciVerse ScienceDirect

Cognition

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

Color categories and color appearance

Michael A. Webster^{a,*}, Paul Kay^{b,c}^a Department of Psychology, University of Nevada, Reno, NV 89557, United States^b Department of Linguistics, University of California, Berkeley, CA 94720, United States^c International Computer Science Institute, Berkeley, CA 94704, United States

ARTICLE INFO

Article history:

Received 28 August 2010

Revised 28 October 2011

Accepted 14 November 2011

Available online xxx

Keywords:

Color

Language

Categorical perception

Perceptual grouping

ABSTRACT

We examined categorical effects in color appearance in two tasks, which in part differed in the extent to which color naming was explicitly required for the response. In one, we measured the effects of color differences on perceptual grouping for hues that spanned the blue–green boundary, to test whether chromatic differences across the boundary were perceptually exaggerated. This task did not require overt judgments of the perceived colors, and the tendency to group showed only a weak and inconsistent categorical bias. In a second case, we analyzed results from two prior studies of hue scaling of chromatic stimuli (De Valois, De Valois, Switkes, & Mahon, 1997; Malkoc, Kay, & Webster, 2005), to test whether color appearance changed more rapidly around the blue–green boundary. In this task observers directly judge the perceived color of the stimuli and these judgments tended to show much stronger categorical effects. The differences between these tasks could arise either because different signals mediate color grouping and color appearance, or because linguistic categories might differentially intrude on the response to color and/or on the perception of color. Our results suggest that the interaction between language and color processing may be highly dependent on the specific task and cognitive demands and strategies of the observer, and also highlight pronounced individual differences in the tendency to exhibit categorical responses.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The physical spectra on which our color vision depends vary continuously, yet we categorize colors in terms of a small set of discrete verbal labels. A long-standing question is whether the categories defined by language can influence how color is processed or perceived (Hardin & Maffi, 1997; MacLaury, Paramei, & Dedrick, 2007). Cross-cultural studies of color naming have demonstrated strong similarities in how speakers of different languages partition color space (Kay & Regier, 2003; Lindsey & Brown, 2006; Regier, Kay, & Cook, 2005), though differences in color boundaries across languages have also been emphasized (Davidoff, Davies, & Roberson, 1999; Roberson, Davidoff, Davies, & Shapiro, 2005). Comparisons of color naming across languages have

thus been used to argue for both independence and dependence of language and color processing (Kay & Regier, 2006).

Another common approach to examining the relationship between language and color perception has been to test for evidence of categorical perception (CP), the tendency to discretely represent stimuli so that they appear more similar or are less discriminable when they belong to the same vs. different categories (Harnad, 1987). Examples of CP have been reported for many stimulus dimensions, from the perception of phonemes (Liberman, Harris, Hoffman, & Griffith, 1957) to face perception (Beale & Keil, 1995; Etcoff & Magee, 1992), and the concept has been invoked repeatedly in the context of color vision (Bornstein, 1987). Recently a number of studies have reported CP effects in the reaction times for discriminating color differences. Individuals are faster at discriminating a target from distractors if the target and distractor colors fall in

* Corresponding author.

E-mail address: mwebster@unr.edu (M.A. Webster).

different linguistic color categories (e.g. blue and green) than if they fall within the same category (e.g. two shades of blue) (Drivonikou et al., 2007; Franklin et al., 2008; Gilbert, Regier, Kay, & Ivry, 2006; Roberson, Pak, & Hanley, 2008; Winawer et al., 2007). These differences are generally stronger for stimuli presented in the right visual field (thus biasing encoding toward the left hemisphere), are reduced by verbal interference, and are specific to the color categories of the individual's language (e.g. differing for English, Russian, or Korean). They can also become manifest after observers are trained to categorize colors according to new arbitrary divisions (Zhou et al., 2010). Analogous categorical effects are also observed in the recognition and discrimination of colors held in memory (Boynton, Fargo, Olson, & Smallman, 1989; Brown & Lenneberg, 1954; Ozgen & Davies, 2002; Pilling, Wiggett, Ozgen, & Davies, 2003; Roberson & Davidoff, 2000). Such effects suggest that language can shape the processing of color in tasks that are not – at least overtly – linguistic, and in this sense represent an example of linguistic relativity in color processing (Regier & Kay, 2009).

Previous studies have also examined CP and the perceptual representation of color by testing color discrimination under conditions in which response time and memory requirements are reduced. One approach has involved examining discrimination of chromatic stimuli that are near the detection threshold. For example, one can measure the saturation differences required to discriminate a stimulus from white, and then ask which of these just detectable colors can be discriminated from each other based on hue. Mullen and Kulikowski (1990) performed an experiment of this kind and found strong evidence for CP. Specifically, for colors at the threshold of detection they found three fixed boundaries in the spectrum, determining four spectral regions, such that only colors from distinct regions were discriminable from each other. However, Krauskopf, Williams, Mandler, and Brown (1986) instead probed discrimination at threshold with isoluminant pulses that varied along different directions in a color space defined by the spectral sensitivities of early color-opponent mechanisms. They found that two stimuli could be distinguished as soon as they were detectable whether they fell along the cardinal axes of the mechanisms or along intermediate orthogonal axes. The absence of CP in this case thus implied that at threshold different color directions might be encoded roughly uniformly rather than discretely through a small number of mechanisms. A further approach involving color discrimination has tested the smallest differences that can be distinguished between suprathreshold stimuli (for example the hue difference between two saturated blues). A recent example is a study by Roberson, Hanley, and Pak (2009), who sampled discriminations for a range of stimuli spanning blue and green. The thresholds remained approximately constant across the blue–green range with no evidence for a decrease at the blue–green boundary, and thus no evidence for a categorical effect. However, CP was observed in a second task in which the discriminations were made more difficult by spatially separating the two colors.

Interpreting the results of these studies is complex because an interaction between language and color could potentially arise at many different levels and thus could

depend critically on the nature of the task (Kay & Kempton, 1984; Pilling et al., 2003; Roberson & Davidoff, 2000). Most studies of categorical effects of color have used measures of *performance* – the speed or accuracy of the responses. These have the advantage of being objective but the disadvantage of introducing processes that may not be tied in a predictable way to color appearance. For example, measures of threshold discriminations are well suited for testing a strong instantiation of CP because small color differences are less likely to be categorized differently by language (Roberson et al., 2009). If one found these threshold discriminations distorted by linguistic boundaries, one could conclude that language was truly affecting perception. However, the converse does not hold: failure to find distortions at threshold does not indicate that perception is not affected. The conceptual problem here is that the task is designed to test for an influence of language under conditions chosen to minimize that influence: restriction to threshold judgments. This assumes or at least tests for the assumption that language has a very general and pervasive effect. However, it leaves untested the alternative that language when it is invoked could directly affect the perceptual representation of color. This representation is not reducible to discrimination thresholds, because the visual mechanisms functioning at threshold may not characterize perceptual processing at suprathreshold levels. Thresholds will be limited by noise and thus by the mechanism(s) with the best signal to noise ratio. There are few grounds for thinking that these same mechanisms are involved at suprathreshold levels or that their responses can be extrapolated in a predictable way. As an example, a common problem in color research is that detection of nominally isoluminant patterns could be mediated by luminance artifacts in the stimulus (Lennie, Pokorny, & Smith, 1993). However, the luminance-sensitive mechanisms underlying this detection may play little role in the suprathreshold appearance of these patterns. Moreover, measures of spectral sensitivity generally fail to predict measures of color appearance, such as the stimulus that appears achromatic or is seen as a pure or “unique” hue (Brainard et al., 2000; Webster, Miyahara, Malkoc, & Raker, 2000; Werner & Scheffrin, 1993).

Similarly, many of the suprathreshold studies that have reported categorical effects on color processing have measured either the speed of the response or the ability to retain colors in memory. This raises the possibility that language is not influencing the actual perception of color, but only the ability to store or react to it. A classic example of an interaction at the response stage is the Stroop effect, in which it takes longer to name the color a word is written in if it denotes a different color (MacLeod, 1991). The automatic tendency to respond to the printed word is thought to compete with the response to the display color. The linguistic effects on color search could reflect a form of the Stroop effect: if the stimuli are both visually and verbally encoded, then the verbal code might compete with the visual response when the two colors fall in the same category (verbally “same” though visually “different”) while facilitating the response to cross-category color pairs (verbally and visually “different”). (However, unlike the conventional Stroop effect with printed color words this

would require that the color alone would automatically trigger verbal coding.) This account has been noted as a basis for the influence of language on the speed and accuracy of perceptual discriminations for color (e.g. Gilbert, Regier, Kay, & Ivry, 2008; Pilling et al., 2003; Roberson & Davidoff, 2000; Roberson & Hanley, 2010), and remains a possible explanation of most of the reported effects, including the right visual field (left hemisphere) lateralization of linguistic category effects. By this account, then, the influence of verbal categories on the reaction times for color discriminations would be post-perceptual, thus leaving open the question of how these categories might affect the actual appearance of color.

In this study, our aim was to test for categorical effects on suprathreshold color appearance, by using tasks that are ostensibly more directly tied to the phenomenology of color. Previous studies have typically explored CP and color appearance by measuring similarity ratings between different color samples (e.g. picking the most different sample from a triad of colors), again with varying results (Davies & Corbett, 1997; Kay & Kempton, 1984; Lindsey & Brown, 2009; Pilling & Davies, 2004; Pilling et al., 2003; Roberson et al., 2005). These ratings have the potential to incorporate verbal categories into the observer's strategies and comparisons, for example if the observer weights these categories in judging the similarity (Kay & Kempton, 1984). Thus as in all tasks, judgments of color appearance could also potentially be susceptible to different forms of linguistic effects. In the present work, we assessed CP and color appearance with two measures of color similarity that are likely to differ in the extent to which verbal color categories might be invoked. For both we focused on the blue–green region of color space, which has shown categorical effects in previous studies of reaction times to color (Gilbert et al., 2006; Roberson & Hanley, 2010; Winawer et al., 2007). The first task involved measuring the effects of color on perceptual grouping. Grouping has the advantage that it taps processes that occur rapidly and with minimal attention (Bravo & Nakayama, 1992; Moore & Egeth, 1997; Treisman, 1982), and does not require overt reference to the nominal color. Color is among the strongest attributes affecting image segmentation and perceptual organization (Wolfe & Horowitz, 2004). Indeed, grouping by color is the principle exploited by classic screening tests for color blindness such as the pseudoisochromatic plates. In these plates the individual dots are perceptually “fused” to form numbers or figures that segregate from the background distractors based on their common color. More generally the power of color to link or segregate different regions of the visual field may in fact represent one of the primary functions of color vision (Mollon, 1989). We examined the tendency for colors to group as a function both of differences in stimulus chromaticity and of verbal labels, specifically to test whether chromaticities that shared a common label might be more likely to be grouped together.

In the second task we re-analyzed hue scaling data that was collected as part of two previous studies (De Valois, De Valois, Switkes, & Mahon, 1997; Malkoc, Kay, & Webster, 2005). In hue scaling, observers rate the proportion of primary hues (red, green, blue, or yellow) in the stimulus (Boynton, Schafer, & Neun, 1964). This task has the advan-

tage that the stimuli are judged in isolation. Consequently, it directly measures color appearance rather than color differences between stimuli (though it requires comparing the contributions of different primaries within the stimulus). Moreover, the task differs from grouping in that observers must explicitly judge the appearance in terms of color labels. We examined how the proportions assigned to these categories varied with chromaticity to test whether the ratings tend to be more similar for chromatic differences that fall within than between linguistic color categories. To anticipate, the results for these two tasks turned out to be substantially different.

2. Methods

2.1. Apparatus

Stimuli were displayed on a SONY 500 PS color monitor controlled by a Cambridge Research Systems VSG 2/5 card, which allows colors to be specified with high resolution. The monitor was calibrated with a PhotoResearch PR650 spectroradiometer to allow accurate reproduction of colors, based on their colorimetric values. Observers viewed the display binocularly in a dark room from a distance of 114 cm, and used a handheld keypad to indicate their responses.

2.2. Stimuli

The grouping stimuli consisted of five circles arranged to form the four corners and center of an imaginary square (Fig. 1). Each circle subtended 1° while the sides of the square subtended 3° . The circles were shown against a gray background (with the chromaticity of Illuminant C) with the same luminance (25 cd/m^2), and were delimited from the background by narrow black borders. Luminances were defined photometrically and thus were not equated for individual observers. Control experiments in which actual luminance mismatches of 5% were introduced into the stimuli in a way that should favor a competing perceptual organization showed that this had no measureable effect on the settings, suggesting that potential luminance artifacts were unlikely to influence the results.

Chromaticities of the circles were defined by their angle in the CIELAB uniform color space, so that color differences across the range of colors tested were roughly equal. All stimuli had a constant difference from the background of 30 delta E, and sampled angles in CIELAB ranging from 115° (yellow–green) to 245° (bluish) in steps of 5° . Stimuli were always shown centered in the display and were directly fixated or viewed at $+8^\circ$ or -8° in the periphery by showing a black fixation cross on the screen.

2.3. Procedure

We measured three different tasks with the stimuli.

2.3.1. Perceptual grouping

In this task the circles at opposite diagonal corners of the square had the same color, while the two diagonals differed in color by a fixed angle of 30° (corresponding to a

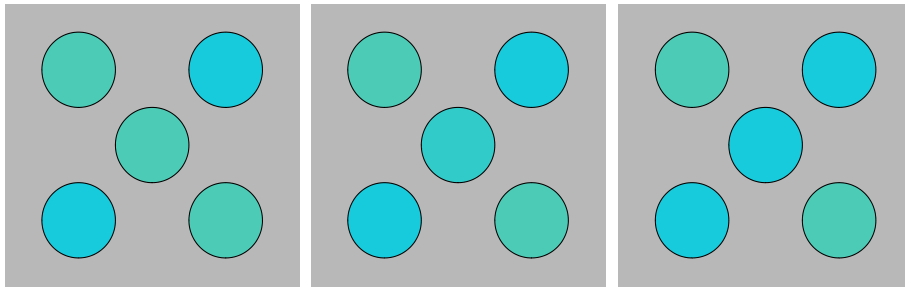


Fig. 1. An illustration of the stimuli. Corner colors along each diagonal have the same hue and differ from the opposite diagonal by a fixed hue angle. When the center color appears more similar to the colors along one diagonal (e.g. greenish on the left or bluish on the right) then the dots perceptually group along that axis. The center dot was adjusted to an intermediate color (e.g. center) until either grouping appeared equally salient. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

moderate fixed color difference (ΔE) of 15.5. This difference is comparable to the color differences that have been found previously to result in categorical effects in measures of reaction times (Gilbert et al., 2006). The center color was initially selected at random from a value between the two corner angles. Observers made a two-alternative forced choice response to indicate whether the perceived orientation was clockwise or counterclockwise. A staircase varied the center color angle to estimate the angle at which both orientations appeared equally likely, with the point of subjective equality estimated from the mean of the final 10 of 13 reversals in the staircase.

2.3.2. Reaction times

In a second task the corner colors were shown as before but the center color was chosen to be the same as one of the two corner pairs. Observers were instructed to respond as quickly as possible to again indicate the perceived orientation of the grouping. Each corner pair was shown 20 times with the reaction times based on the mean of all correct responses. Accuracy effects were not analyzed as incorrect responses were infrequent.

2.3.3. Color boundaries

The purpose of the final task was to estimate the observer's blue–green category boundary. The observer was shown the center color in isolation, and the color angle of this circle was varied with the same staircase as in the grouping task, but this time the observer was instructed simply to indicate the color category of the stimulus (i.e., by responding “blue” or “green”). Similar measurements were also made for the yellow–green and blue–purple boundaries in order to delineate the angular range of the blue and green categories.

The grouping and reaction time experiments were run in separate sessions with different subsets of observers, with color boundaries and stimulus condition measured in interleaved trials within the sessions for each. Separate sessions were also used to test hemispheric effects for the grouping. For all cases observers were adapted throughout to the gray background. Test stimuli were displayed for an interval of 0.5 s with the next stimulus shown 3 s after each response. Observers were tested over a range of different color angles with the order of angles and eccentricities (left or right

visual field) counterbalanced across trials. Trials measuring grouping and the color naming boundaries were intermixed within each session. Plotted results show the mean of four repeated settings for each stimulus condition for the grouping or boundary estimates or up to 20 settings for the reaction times.

2.4. Participants

Observers included author MW (who is left-handed and denoted in the figures as S1) and 12 right-handed students at the University of Nevada, Reno, with different individuals tested in different subsets of experiments. One student was aware of the aims of the project while the remaining participants were naïve. This student was tested only in the conditions of Figs. 5 and 6, and is labeled as S2. Note that S2 in Fig. 3 and Table 1 is instead a different, naïve observer. All had normal or corrected-to-normal acuity and normal color vision as assessed by the Cambridge Colour Test, and were native speakers of English. Participation was with informed consent and all procedures followed protocols approved by the university's Institutional Review Board.

3. Results

3.1. Perceptual grouping

3.1.1. Predicted categorical effects

Fig. 2 shows an example of how a CP effect might be manifest in the grouping task. Each pair of vertically separated circles plots the hue angles of a given pair of corner colors, which differ by a fixed angle of 30° within the color space (with the angle of the “greener” color below and the “bluer” color above). The curves plot the angle of subjective equality for the two possible groupings of the center color with either corner color, as a function of the mean angle of the two corner color pairs. In the absence of CP the chosen angle should, in the simplest case, follow the mean and thus should plot as a single straight line whether the corner colors fall within the same or different verbal categories. Conversely, if CP is complete, then observers should always set the center color at the category boundary when the two corner colors span the boundary, resulting in a line

Table 1

(a) Test for slope differences between color pairs that were across the blue–green boundary or fell within either the blue or green categories. (b) Average difference between settings for adjacent stimulus angles for color pairs that were across or fell within the blue or green categories.

(a) Observer	Slope (SE) across blue or green	Slope (SE) within blue/green	<i>t</i> (df)	<i>p</i>
S1	.75 (.035)	.934 (.013)	5.04 (80)	.001
S2	.765 (.065)	.909 (.015)	2.14 (80)	.05
S3	1.13 (.054)	1.168 (.015)	.68 (68)	NS
S4	1.07 (.053)	1.10 (.012)	.61 (71)	NS
S5	1.14 (.075)	1.15 (.019)	.12 (71)	NS
S6	1.03 (.048)	1.13 (.020)	1.93 (52)	NS
S7	1.23 (.047)	1.15 (.016)	−1.43 (67)	NS
S8	.907 (.037)	1.03 (.009)	3.23 (71)	.01
(b) Observer	Differences (SE) across blue/green	Differences (SE) within blue/green	<i>t</i> (df)	<i>p</i>
S1	3.84 (0.49)	5.15 (0.29)	2.33 (16)	.05
S2	3.91 (0.42)	5.09 (0.70)	1.00 (16)	NS
S3	5.56 (0.58)	6.18 (0.78)	0.52 (13)	NS
S4	5.92 (1.21)	5.16 (0.53)	−0.67 (14)	NS
S5	4.79 (0.55)	6.10 (0.74)	1.11 (14)	NS
S6	5.19 (0.62)	5.23 (1.10)	0.026 (9)	NS
S7	6.18 (0.36)	5.30(0.41)	−1.36 (13)	NS
S8	4.15 (0.40)	5.61 (0.34)	2.55 (14)	.05
All observers	5.48 (0.19)	4.94 (0.84)	1.59 (7)	NS

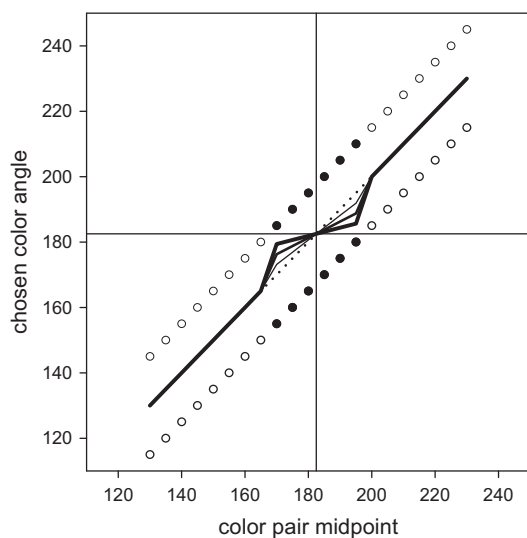


Fig. 2. Hue angles for grouping predicted by different degrees of categorical bias. For this example the blue–green boundary is at a hue angle of 182°. Filled dots show the hue angle of corner color pairs that span the boundary, while unfilled dots are for pairs that both fall within the blue or green categories. In the absence of bias the center hue angle at which either orientation of grouping is equally likely should remain midway between the two corner hues (dotted line). For stimuli that span the blue–green boundary, a categorical bias will increase the similarity of hues that fall within the same category and thus bias the settings toward the boundary, reducing the slope of the function. Predicted settings are shown for a bias of 0.25 (thin line), 0.5 (medium) or 0.75 (thick).

of zero slope for cross-category conditions – that is, for the six cross-category pairs of corner colors, denoted by filled circles on Fig. 2. Under complete CP observers should have no basis for discriminating between the corner colors when both instead fall within the same category, and thus the settings should vary randomly within the color category. Obviously, we normally can distinguish different

shades of blue or green so full CP is untenable. However, it remains possible that there is a bias toward CP, so that color pairs that fall within the same category appear more similar. This would act to bias the settings toward the category boundary for cross-category but not within-category corner pairs. For example, this could occur if the response is a weighted average of two decisions, an analog response sensitive to the chromatic differences in the stimuli and a binary response that classifies the stimuli as “same” or “different.” The different curves in the figure show the predicted responses for different weightings of these two decisions, and predict increasingly shallower slopes for cross-category pairs as the category difference is given more importance. In the analysis we fit curves of this form to the measured responses to estimate the relative weighting and thus the potential magnitude of a bias toward CP.

3.1.2. Grouping of centrally fixated colors

Fig. 3 shows the results of the grouping task. Each panel plots the settings for one of the eight observers tested. Again S1 was an author while the remaining seven observers were naïve. In each case the mean hue angle of the stimuli varied from 130° to 230°, thus spanning a wide range from yellow–green to blue. The crossed horizontal and vertical straight lines indicate the observer’s blue/green boundary. Consistent with previous reports of individual differences in color naming (Kuehni, 2004; Malkoc et al., 2005; Webster et al., 2000), there were substantial differences in the angles chosen for the blue–green boundary, with values varying from 172° to 205°. These differences were markedly larger than the within-observer variation from the repeated settings across the four sessions (average $sd = 5.6^\circ$). Observers also reliably differed in the boundaries chosen for blue–purple (261–289°) and yellow–green (108–147°). Notably, boundaries for the color grouping also showed consistent differences between observers, with overall slopes ranging from shallower (0.91) to steeper (1.17) than the nominal midpoint in the

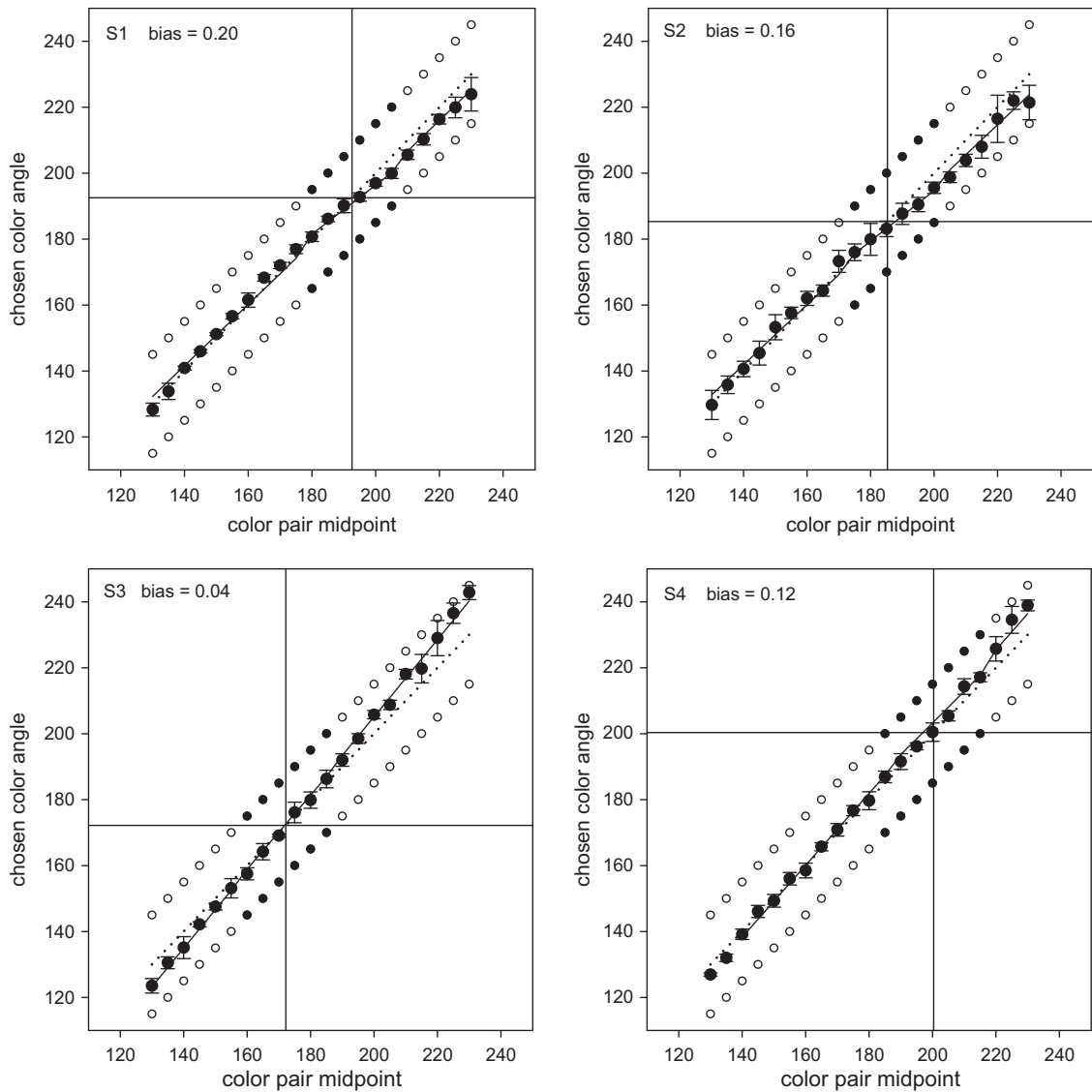


Fig. 3. Midpoint settings for grouping as a function of the mean hue angle of the corner colors (filled symbols ± 1 sd) for eight observers. Horizontal and vertical lines indicate the observer's blue–green boundary. Solid line shows fits to the midpoints by varying the categorical bias, with the estimated bias indicated in each panel.

CIELAB space. Because of these inter-observer differences we analyzed the results for each individual rather than pooling across observers, as in most previous studies.

We examined whether the grouping changed near the blue–green boundary in three ways. In the first case we fit a simple model of the form illustrated in Fig. 2 to estimate the relative influence of category differences in the stimuli. Because the settings for most observers systematically deviated from the midpoint between the corner angles in the CIELAB space, we fit a single line to the settings for corner pairs that fell within the green or blue categories (excluding angles that encroached on an individual observer's yellow–green boundary), and then fit the same line plus a bias toward the blue–green boundary for pairs that spanned the boundary. The fits were thus estimated from

$$\Theta_{\text{pred}} = m\Theta_c + b$$

(for within category pairs of corner colors)

$$\Theta_{\text{pred}} = (1 - \alpha)(m\Theta_c + b) + \alpha c$$

(for cross-category color pairs)

where Θ_c and Θ_{pred} are the mean of the two corner angles and the predicted midpoint chosen by the observer, respectively; m and b are the slope and intercept of the best-fit line to the settings for within category stimulus pairs; c is the observer's chosen blue–green boundary; and α is the relative weighting of the categorical response. The fitting thus varied the slope and intercept of the line and the magnitude of the CP bias.

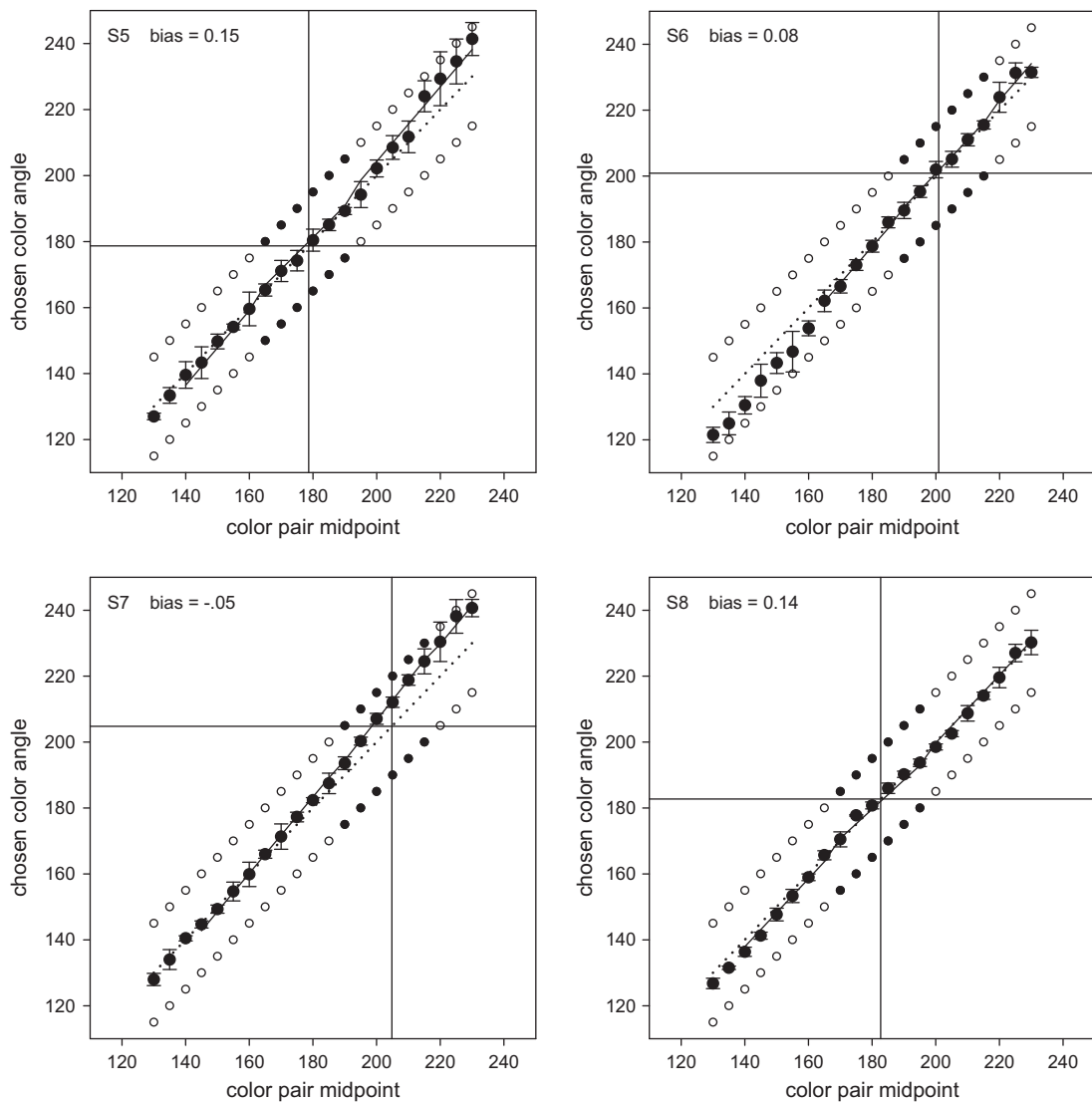


Fig. 3 (continued)

Stimuli straddling the boundary were nominally taken as pairs where both corner angles were within 30° of the individual observer's blue–green border. Variability in the perceived or chosen boundary would tend to blur the transition from within- to cross-category pairs, and thus tend to smooth the transition between the two fitted line segments. We therefore also repeated the fits for slightly narrower or broader ranges for defining the cross-category color pairs. However the estimated fits remained similar. Estimates of the bias ranged from $-.05$ to 0.20 for the eight observers and are indicated along with the fitted model within the panels showing each observer's settings. Overall the participants' settings thus trended toward a (very weak) CP effect, with an average bias of 0.10 (which was nevertheless significantly different from zero; $t(7) = 3.73$, $p < .01$). However, including this bias significantly improved the model fit for only 1 of the 8 observers (author S1, $F(22,22) = 0.44$, $p < .05$).

(The average bias across the seven naïve observers remained significant when S1 was excluded; $t(6) = 3.21$, $p < .01$.)

As a second test, we fit regression lines to different segments of the function to ask whether the slope differed for color pairs that spanned the blue–green boundary or were confined to either category (Table 1). These slopes did not significantly differ between the within-blue and within-green segments for any observers, and were significantly shallower for the cross-category pairs for only 3 of the 8 observers. We also compared the differences between the settings for adjacent stimulus angles for pairs that crossed the blue–green boundary or were within either category (Table 1). Again these differences should be smaller for the cross-category pairs if there is a categorical bias. However, the difference was significant for only 2 of the 8 observers and did not reach significance when pooled across observers. The results are thus consistent with at

most only a very weak categorical effect of the color differences spanning the blue–green border.

3.1.3. Choice of color space

Measures of categorical effects require testing stimulus differences that are equated by some criterion and then asking whether they remain equivalent on a given task when the differences fall within or between categories. As noted, we defined the stimuli by their angles in the CIELAB uniform chromaticity diagram in order to roughly equate the magnitude of color differences between the two corner angles across the range of angles tested. Perceptual metrics are by their very design chosen to equate perceptual differences in the stimulus. Relying on these metrics to test for CP – a change in perceptual difference – thus presents a potential circularity (and one which has been present in many previous studies). If the space has correctly scaled perception for the task at hand, then performance on the task will already be uniform and will mask any nonuniformities in the underlying response to color (or alternatively, could also give rise to artifactual categorical effects if the space is not scaled appropriately for the individual). However, spaces like CIELAB were derived from threshold color discriminations and apply global transformations to the stimulus chromaticities. As a result, they do not in general embody the local distortions that should result from CP around different local category boundaries.

To test the influence of the choice of color space around the blue–green boundary, we re-analyzed the stimuli within a cone-opponent space that varies linearly with cone excitation. The space was a variant of the MacLeod and Boynton (1979) and Derrington, Krauskopf, and Lennie (1984) diagrams in which chromaticities were defined by variations in S cones and relative excitation of the L and M cones at constant luminance. In our version the origin corresponded to the chromaticity of Illuminant C and contrasts along the two axes were scaled to roughly equate sensitivity to the S and LvsM axes as detailed in Webster et al. (2000). Fig. 4a shows the nonlinear relationship between the stimuli in the CIELAB and the cone-opponent space. To ask whether this could have distorted the color differences around the blue–green border, we calculated the stimulus angles that observers should have chosen if in our grouping task they were setting the central disk to be equidistant from the two corner colors in terms of the linear cone-opponent signals. The predicted settings are shown in Fig. 4b for the scaling assumed in Webster et al. (2000), while the dashed lines show the predictions if the S signals are either increased or decreased by a factor of 2. We included these scale differences because there is no established metric for equating signals along the S and LvsM axes and because this scaling can vary across stimuli and observers. It is also important to consider because the blue–green boundary is near the $-L + M$ axis (180°) and thus might be more susceptible to changes with scaling. Finally, we applied the same fitting procedure to the model data, assuming a blue–green border at 185° in CIELAB, roughly the mean value for the observers. This produced an estimated bias of only 0.03 for the original scaling

and biases ranging from 0.01 to 0.07 for the twofold increase or decrease in S contrasts, respectively.

This analysis shows that the CIELAB space introduces very little bias in the predicted grouping settings across the blue–green boundary relative to a linear cone space. In turn, this means that the space did not undo and thus “hide” a large effect that would have been observed by representing color differences in a linear cone-excitation space. Alternatively, the very weak biases introduced by using the CIELAB metric suggests that the small biases we found in the observer’s settings may in part include an artifact of the stimulus space, weakening further the evidence for a clear CP effect in the grouping task. Finally, we note again that observers showed consistently different results in the grouping. Thus the nominally uniform space was not uniform for individual observers.

3.1.4. Hemispheric effects

Several studies of search times for color differences have found that categorical effects are stronger in one hemisphere (Drivonikou et al., 2007; Franklin, Drivonikou, Bevis et al., 2008; Gilbert et al., 2006, 2008; Roberson et al., 2008; Siok et al., 2009). In right-handed English-, Mandarin- and Korean-speaking adults the cross-category advantage occurs for colors presented in the right visual field, thus biasing processing toward the left hemisphere – again consistent with an influence of language. Interestingly, in infants this bias is reversed, in that cross-category effects are instead lateralized to the right hemisphere (Franklin, Drivonikou, Bevis et al., 2008; Franklin et al., 2008). However, other studies have found comparable effects for stimuli presented to both hemispheres (Brown, Lindsey, & Guckes, 2011; Witzel & Gegenfurtner, 2011).

We examined whether there might be an effect of visual field on color grouping. To test this we repeated the same task but with a fixation point added to the display so that the circles were now centered at 8° in the left or right visual field. In order to avoid producing potential differences in the chromatic stimuli across conditions (e.g. because of monitor inhomogeneities), the grouping stimuli remained centered on the screen while fixation crosses were added to either side. The presenting field was counterbalanced across staircases and the corner color angles were restricted to span a range closer to the blue–green boundary. Fig. 5 shows individual settings for the six observers tested. As before there are consistent inter-observer differences in the slopes relating the chosen color boundaries to the mean corner color angle, ranging from ~ 0.6 to 1.0. As noted above, these slopes should be shallower for corner colors spanning the blue–green boundary, and in this case we asked whether they are shallower within an observer depending on which visual field the stimuli were displayed in. To test this we again compared only the settings for corner colors that spanned the color boundary. Notably the color boundaries were very similar but not always identical at the two eccentricities; when they differed the cross-category color pairs were defined separately for each hemisphere. However, for none of the six observers was there a significant difference between the slopes for their settings at the two field locations (Table 2). Moreover, settings at the two sides also did not differ for any observer

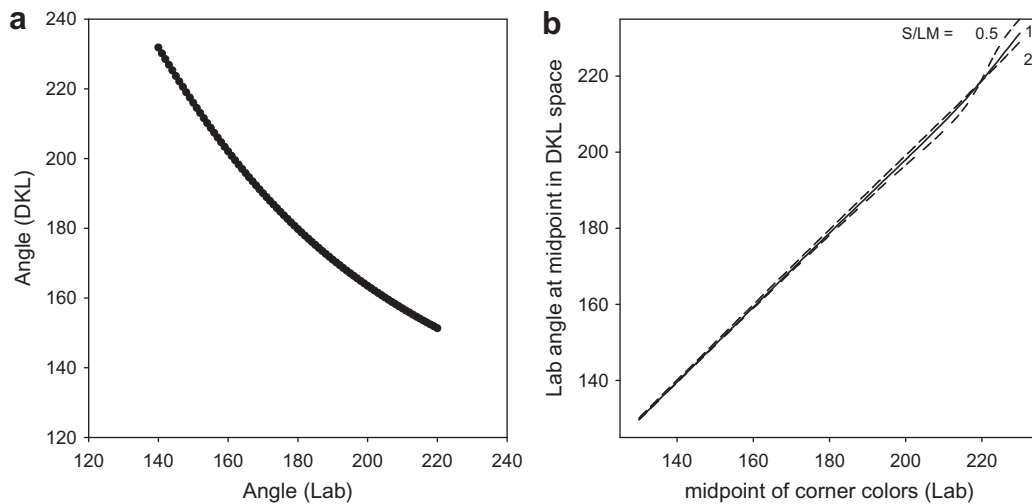


Fig. 4. (a) Relationship between hue angles in the CIELAB space and a scaled version of the LvsM and S cone-opponent space. (b) Midpoint settings predicted for the stimuli defined by their CIELAB coordinates if the midpoint depends on equal distances in the cone-opponent responses. Settings are shown for the nominal relative scaling of LvsM and S contrasts or for a twofold increase or decrease in the assumed S contrast.

based on an analysis of the differences in the hue angles chosen for adjacent stimulus pairs. Thus the data provide little evidence for a hemispheric difference in the perceptual grouping of color.

3.1.5. Reaction times

In the final measurements we used the same stimuli to compare reaction times for colors that spanned the blue–green boundary, to test whether for our conditions a clearer categorical effect might occur when the form of the response was changed. As noted, in this case the corner colors were again separated by a fixed angle of 30° while the center color was set equal to either pair. Observers responded as quickly as possible to indicate the orientation of the matched colors.

Fig. 6 shows the results for five observers as well as the mean results across observers, for patterns presented in the left or right visual field. The settings averaged across observers exhibit a weak minimum at angles near the blue–green border, though again this border varied over a substantial range (22°) between observers. For each individual we compared the mean reaction times for color pairs that spanned their blue–green boundary vs. pairs that both fell within the blue or green regions. Response times were shorter for 3 of the 5 observers in the right visual field (though one of these observers (S1) is left-handed) (Table 3). In contrast, differences did not reach significance for any of the observers for the left visual field. Differences also did not reach significance for either hemisphere when pooled across observers. The pooled results thus did not clearly reveal a categorical effect in the right visual field (left hemisphere) that has been found in some previous reports. Moreover, like the midpoint settings, hemispheric differences were not robust across individual observers, and in this case were found for only one of the three naïve observers (though this null finding relative to previous studies could also be because we tested only a small sample of observers).

3.2. Hue scaling

Again, an advantage of the grouping task is that it allowed us to assess the relative similarity of moderate, suprathreshold color differences in a judgment that did not require an overt coding of color category and where performance on the task was not limited by speed or accuracy. To explore an alternative suprathreshold appearance task, we turned to an analysis of measurements from previous studies of hue scaling, which instead requires observers to explicitly label the colors. In this color naming task, observers judge the relative proportion of red, green, blue, or yellow primaries in different chromatic stimuli. For example, an orange hue typically appears as a mixture of red and yellow, but the amount of red or yellow varies systematically as the chromaticity changes between pure red and yellow. In this sense, the blue–green boundary as defined above measures one point on the hue scaling function, where the proportion of blue and green are scaled equally. Ratings across the spectrum have been used to define the response characteristics of the phenomenal red–green and blue–yellow opponent processes (Abramov, Gordon, & Chan, 1991; Boynton et al., 1964; De Valois et al., 1997; Webster, Malkoc, Bilson, & Webster, 2002), and thus provide a measure of how color appearance varies with the direction in color space. We asked whether these ratings show evidence for categorical effects at angles that fall within or between different verbal color categories.

Fig. 7 illustrates a highly simplified example of categorical effects on hue scaling. In this case the chromatic angle of the stimulus (abscissa) is defined as the direction in a scaled version of the MacLeod–Boynton cone-opponent space, equivalent to the stimulus angles shown previously in Fig. 4. Again this space represents the chromaticity in terms of two cardinal axes that correspond to differential excitation in the L and M cones at constant luminance [an axis of $0(+L)^\circ$ to $180(-L)^\circ$], or relative excitation in the S cones at constant luminance [an axis of $90(+S)^\circ$ to $270(-S)^\circ$]. The

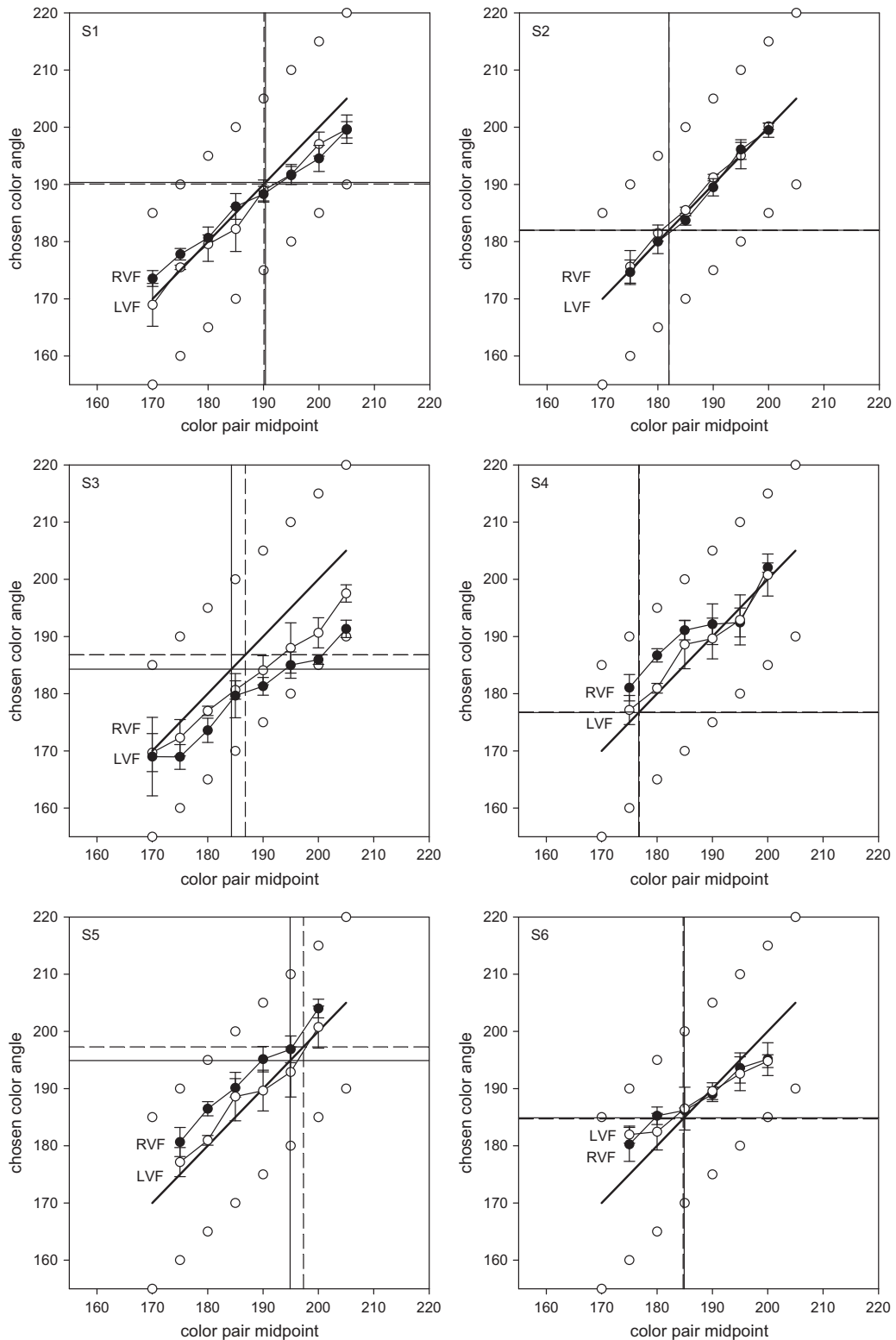


Fig. 5. Midpoint settings for color grouping for stimuli shown in left (unfilled symbols) or right (filled symbols) visual field for 6 observers. Horizontal and vertical lines show the selected blue–green boundary in the left (dashed) or right (solid) fields.

angle of the perceptual response (ordinate) is instead in terms of the phenomenal red vs. green (an axis of 0° [pure

red] to 180° [pure green]) and blue vs. yellow [an axis of 90° (blue) to 270° (yellow)] hue responses. Suppose that

Table 2

Slopes for colors spanning the blue–green boundary compared for stimuli presented in the left or right visual fields.

Observer	Left field		Right field		<i>t</i> (df)	<i>p</i>
	Mean	SE	Mean	SE		
S1	.85	.061	.70	.041	1.93 (44)	NS
S2	.97	.054	1.05	.055	–.986 (36)	NS
S3	.733	.069	.703	.085	.27 (44)	NS
S4	.86	.143	.75	.11	.56 (28)	NS
S5	.81	.070	.88	.071	–.67 (28)	NS
S6	.57	.083	.61	.065	–.37 (36)	NS

the red–green and blue–yellow responses were aligned with the cardinal axes and varied sinusoidally with the stimulus angle in the cone–opponent space (which would correspond to mechanisms that combine the cone signals linearly). These response functions are illustrated in Fig. 7a. In the hue scaling experiment observers are typically shown a set of colors one at a time, and for each report the relative proportions of each primary they perceive. In the absence of separate saturation judgments, the responses for each stimulus are thus normalized so that they sum to a constant value (1 in this example). These normalized functions are illustrated in Fig. 7b. Finally, the functions can be also represented by plotting the angle in color appearance space (where red vs. green = 0–180° axis and blue vs. yellow = 90–270° axis) as a function of the stimulus angle (i.e., appearance angle = $\tan^{-1}[(\text{blue–yellow response})/(\text{red–green response})]$ (Malkoc et al., 2005). For the linear response mechanisms this results in color appearance angles that change linearly with the stimulus angle, and the hue scaling function should thus follow the diagonal line (Fig. 7c). A categorical effect might again be manifest as a tendency to discretely represent stimuli that are dominated by a common primary as the same. For example, the dashed lines in Figs. 7a and 7b plot the responses if there is a 0.5 bias to report the hue as 100% of the dominant primary (i.e., the response is the average of the linear response and a discrete response to the dominant primary). The stepped functions in Fig. 7c show the predicted angles for different magnitudes of this bias. Like the grouping predictions, the slope changes at the category boundary, but in the hue scaling the predicted effect is for shallow slopes within the category (where for example the strength of “blue” or “green” appears to change more slowly with stimulus angle) while steeper at the boundary.

Actual hue scaling functions are more complicated for several reasons. First, it is well known that the perceptual axes are not aligned with the cardinal axes that characterize early post-receptor color coding (Krauskopf, Williams, & Heeley, 1982). For example, the blue–yellow axis is roughly midway between the LvsM and S axes. Second, the red vs. green or blue vs. yellow responses are not sinusoidal or symmetric, and thus inconsistent with a simple pair of linear opponent responses (De Valois et al., 1997). Finally, for the purpose of analyzing verbal categories an additional complication is that the hue circle includes two additional basic color terms (orange and purple), which observers did not use for the scaling. Nevertheless, it is possible to ask whether perceived hue as measured by this task

changes more rapidly at boundaries between the different color terms.

To analyze this, we used the hue scaling functions measured previously by Malkoc et al. (2005) for 59 observers. These sampled the LvsM and S plane at intervals of 15°. The colors were displayed in a 2° field on a white background, and shown individually for 1 s each. Fig. 8a and b shows the average measured scaling function and the corresponding inferred chromatic response functions. Note again that ordinate values of 0°, 90°, 180°, and 270° correspond to the stimuli that on average appeared pure red, blue, green, or yellow respectively, and vertical solid lines mark the corresponding LvsM and S stimulus angles. In the study of Malkoc et al. observers also named each hue using one of the six basic color terms. The purple and orange foci are taken from the modal values for the color naming for these hues as reported in their study. The dashed lines correspond to the color boundaries between adjacent terms. Between the scaling primaries (e.g. blue and green) these correspond to the stimulus angle where the two components appeared equal. For boundaries that included purple and orange, they were instead estimated from the color naming and the stimulus angles at which the two color terms were used with equal frequency, again as reported in Malkoc et al. (2005).

The average function shows evident scalloping. We estimated the local slope of the hue scaling from the derivative of piecewise polynomial fits to the curve. These slopes are shown for the average curve in Fig. 8c, and show a clear tendency for color appearance to vary most gradually with stimulus angle around pure red, blue, and green while showing a sharp transition at the blue–green border. On the other hand, appearance changed most rapidly around unique yellow. This might argue against a simple CP effect as a basis for the variations in color appearance, though an alternative is that yellow occupies only a narrow angular range of the cone–opponent space and was thus sampled too coarsely to define the hue changes within the yellow category. The results of Malkoc et al. (2005) do not allow us to distinguish between these alternatives for yellow. Moreover, the average function may tend to smooth the discontinuities in individual curves if the color boundaries vary across observers.

To test for categorical effects we again focused on the blue–green boundary in order to compare the hue scaling with the grouping task, and fit a model of the form shown in Fig. 7c to the stimulus angles ranging from the blue–purple to yellow–green border. The fits varied the slope and offset of the curve and the magnitude of the bias, and followed procedures similar to those used to fit the grouping settings. The prediction for the average curve is shown by the thick line in Fig. 8b and indicates a CP bias of 0.23 across the blue–green boundary. This is larger than the average CP effects indicated by the grouping, and in fact underestimates the degree of bias in individual observer’s settings, which spanned a very wide range from 0.02 to 0.92 with a mean of 0.35 (Fig. 8d). Again this underestimate occurs because the average curve tends to smooth between the different category boundaries of individuals. The distribution of categorical biases in the hue settings is substantially different from 0 ($t(58) = 15.79, p < 0.0001$). We also confirmed the

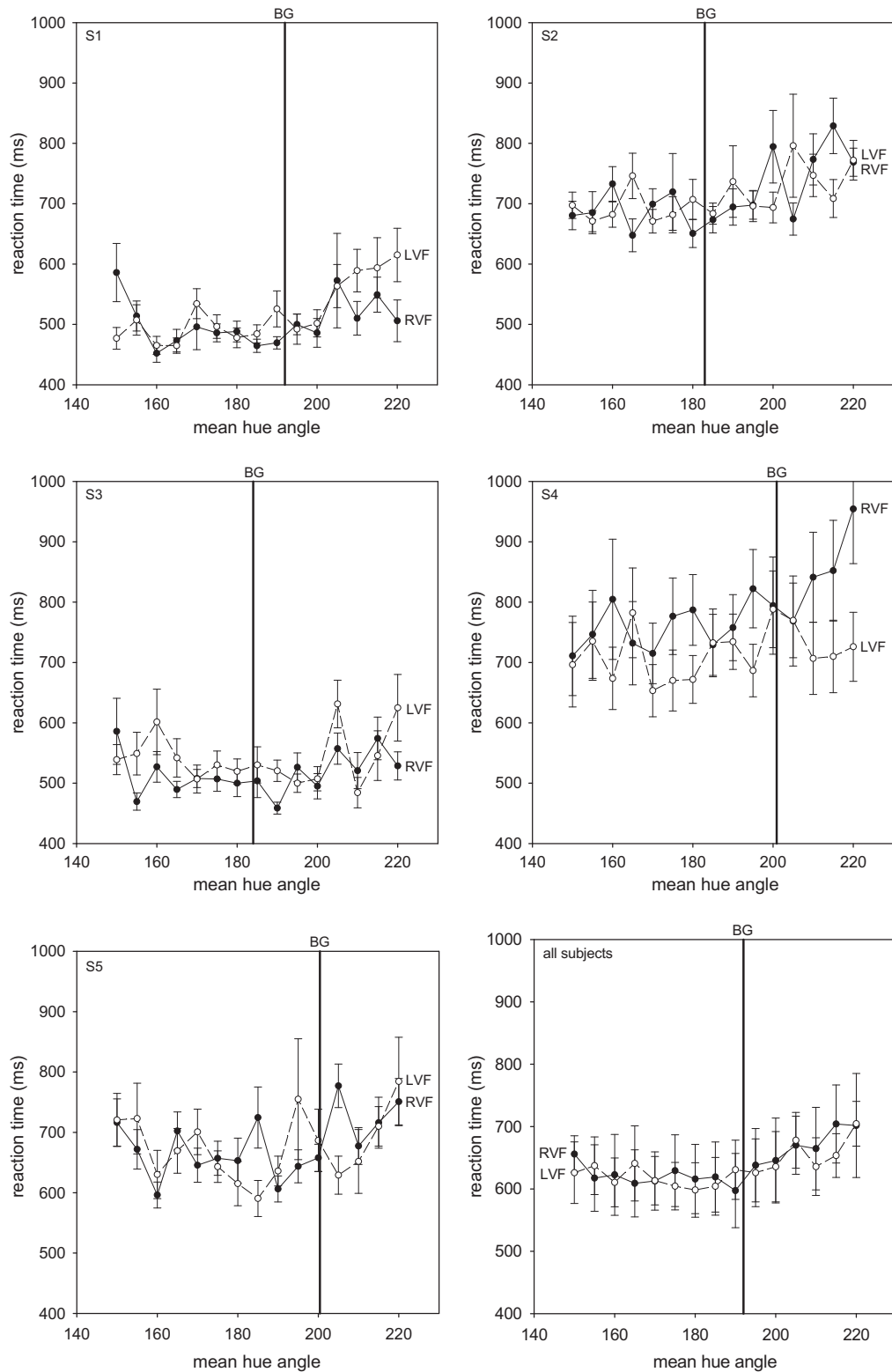


Fig. 6. Mean reaction times (± 1 se) for judging the grouping of the colors as a function of the mean hue of the corner colors for five observers and for the mean settings across observers (lower right). Stimuli were viewed at 8° eccentricity in the right (filled) or left (unfilled) visual field. Vertical lines denote each observer's selected blue-green boundary.

Table 3

Reaction times (ms) for grouping judgments for patterns presented in the right or left visual field.

Observer	Right field				<i>t</i> (218)	<i>p</i>
	Between		Within			
	Mean	SE	Mean	SE		
S1	480	8.1	508	11.4	1.98	<.05
S2	685	12.2	735	12.8	2.85	<.01
S3	475	12.6	511	12.4	2.08	<.05
S4	726	31.6	668	21.3	−1.52	NS
S5	687	15.4	663	12.0	−1.23	NS
	Left field					
S1	501	10.7	512	9.8	.74	NS
S2	704	18.1	725	15.8	.89	NS
S3	507	12.4	539	15.0	1.68	NS
S4	666	26.4	624	19.1	−1.30	NS
S5	696	31.0	671	15.0	−.74	NS

biases a second way by again comparing the change in the hue angle of the settings for the stimulus pairs that crossed each observer's blue–green boundary vs. the adjacent pairs on either side of the boundary. In this case the perceived hue change should be larger as the stimuli cross the boundary, and this difference was again found to be highly significant (mean and standard error of hue change across = 31.7 ± 1.97 , within = 18.16 ± 1.01 , $t(175) = 6.81$, $p < 0.001$). Fig. 9 shows the settings for the observers that showed the strongest or weakest biases and observers near the median. The results suggest that – unlike the effects we found for grouping – in the hue scaling task individuals vary widely in the tendency to categorize the stimuli, and moreover show that some individuals approach nearly complete categorical responses. (The individual plots in Fig. 9 suggest a potential relationship between the magnitude of an observer's categorical bias and the separation between their unique blue and green. However, separation and bias were not significantly correlated across the full set of observers, and for observers with strong categorical biases the location of pure blue and green are not well defined.)

Finally, to ensure that the categorical biases in hue scaling were not specific to the testing procedures used by Malkoc et al. (2005) (or to the hedonistic confines of their Reno test site), we analyzed comparable hue scaling data from a study by De Valois et al. (1997). These authors similarly measured hue scaling for stimuli as a function of their angle in the LM vs. S chromatic plane (at intervals of 22.5°). We again estimated the bias in the blue–green region for the 4 observers whose results were reported in their Fig. 3, after converting the independent percentages shown for each primary into angles in the red–green vs. blue–yellow space. These biases ranged from 0.16 to 0.31 with a mean of 0.25, and are shown for the four subjects as black bars in Fig. 8d. These are consistent with the moderate biases exhibited by many of the observers in the Malkoc et al. study, and thus suggest that the hue scaling procedure is generally susceptible to a categorical-like bias around the blue–green boundary. Interestingly, two of the participants in the De Valois et al. work included the authors who – as highly experienced color scientists – would have been especially well attuned to the task of

abstracting the component colors from the stimuli. It is therefore significant that even these observers tended to show a categorical effect (0.28 for author RD and 0.16 for KD). This argues against the possibility that naïve subjects were simply prone to categorically label the stimuli – even if these labels were clearly inconsistent with the observers' percepts.

4. Discussion

In this study we used two different measures of color appearance to test for categorical effects in color perception. Before considering the implications of the results, four points should be emphasized. First, our measures do not necessarily assess the effects of language on color. Indeed, many studies of CP have been interpreted as revealing the underlying discreteness of the neural representation and have not considered linguistic boundaries. For example, in classical opponent color theory these boundaries are determined by the neural responses (Hurvich & Jameson, 1957) (though neural responses that could give rise to these boundaries have yet to be clearly identified). The strongest behavioral evidence for linguistic effects has been the demonstration that categorical effects are reduced during concurrent verbal interference (Gilbert et al., 2006; Roberson & Davidoff, 2000; Winawer et al., 2007). We did not attempt these controls in the current study because in the empirical measurements we made any categorical effects in the grouping task were not significant for most observers. A second related point is that categorizing and verbally labeling the stimulus are not necessarily synonymous. For example, studies of category learning and categorical perception have pointed to dissociations between categorization and naming (Goldstone & Hendrickson, 2010; Hanley & Roberson, 2011). A third point is that any conclusions should be tempered by the fact that evidence for CP has been inconsistent even when the task is similar across studies. For example, studies of reaction times to colors have differed in whether they find CP and in whether or not it differs between the hemispheres (Brown et al., 2011; Franklin, Drivonikou, Bevis et al., 2008; Gilbert et al., 2006; Roberson et al., 2008; Winawer et al., 2007; Witzel & Gegenfurtner, 2011). Thus the conditions under which verbally mediated CP is manifest remain in question. A final point is that individual differences in both the hue scaling and grouping tasks – as well as in the location of observers' color boundaries – were substantial. These differences have been found in many previous studies of color naming (Webster & Kay, 2007), but are often hidden in analyses of group effects. The dramatic differences we observed among individuals in hue scaling are a case in point. These individual differences may themselves provide important clues about perceptual processes (Wilmer, 2008), and specifically about the interaction between language and color (Lindsey & Brown, 2009a, 2009b), and they are also important for evaluating the robustness and bases of categorical effects.

As noted, categorical effects on color processing have been tested in a variety of different paradigms. We focused on two paradigms to measure the phenomenal appearance of the stimuli, i.e., to try to measure CP in tasks and

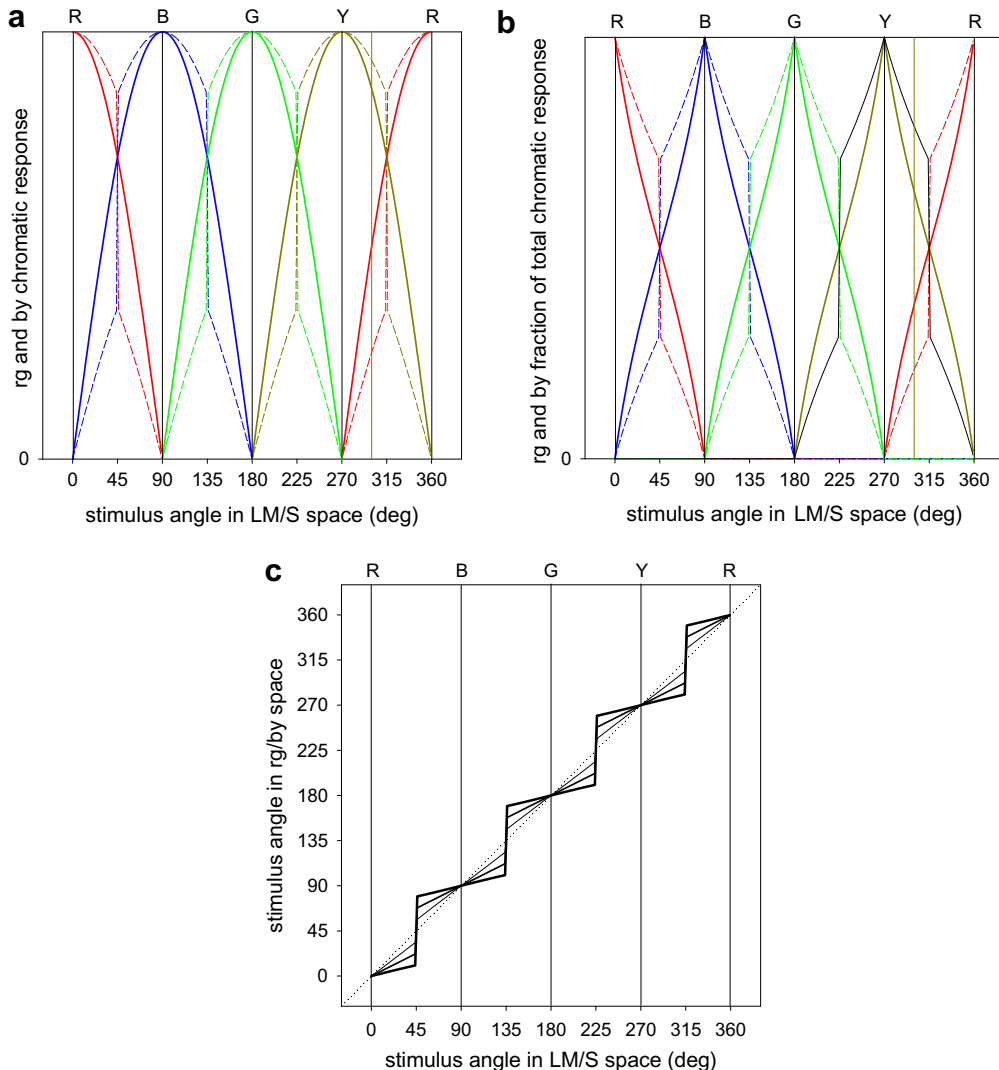


Fig. 7. Hypothetical hue scaling functions. (a) Chromatic response functions for red–green and blue–yellow. Solid lines show the responses predicted by linear mechanisms and vary as the cosine of the preferred angles in cone–opponent space. Dashed lines show the predicted responses if they include a categorical bias (=0.5) to respond only to the dominant hue. (b) The same response functions normalized to show the proportion of each primary hue at each wavelength, as typically reported in the hue scaling task. (c) Predicted scaling function for different magnitudes of bias. With no bias the function has a constant slope. A categorical bias will increase the similarity of hues within the same category (e.g. blue or green), reducing the slope of the function within the category and increasing the slope at the category boundary. Predictions are shown for a bias of 0.25 (thin solid line), 0.5 (medium) or 0.75 (thick). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

responses that might more directly reflect judgments of color appearance. An advantage of the grouping task we used is that the observer is not asked to overtly classify this appearance or color similarities but must somehow access this through whatever processes mediate perceptual organization. At least for the specific conditions and color boundary (blue–green) we tested, the results suggest that the color boundary exerts only a weak influence on the percept. If this is true of other color boundaries then the effect of color differences on perceptual organization – perhaps one of the most important functions of color vision – may be largely seamless and innocent of linguistic influence. In other words, in the case of grouping, color differences may show very little effect of language or

categorization and instead behave as an undifferentiated continuum. For the blue–green boundary this is all the more surprising because, as noted above, it falls close to the $-L$ pole of the $LvsM$ axis of cone–opponent space (Malkoc et al., 2005). Stimuli that lie on either side of the boundary correspond to increments or decrements in S cone signals, which are carried by distinct post–receptoral pathways (e.g. (Tailby, Solomon, & Lennie, 2008)). Thus, there are both neural and linguistic bases for a potential effect at the blue–green boundary, but neither appears substantial.

There was nevertheless a weak but significant categorical bias in the grouping results. Again, 7 of the 8 observers tested showed a bias in the direction predicted by CP, a

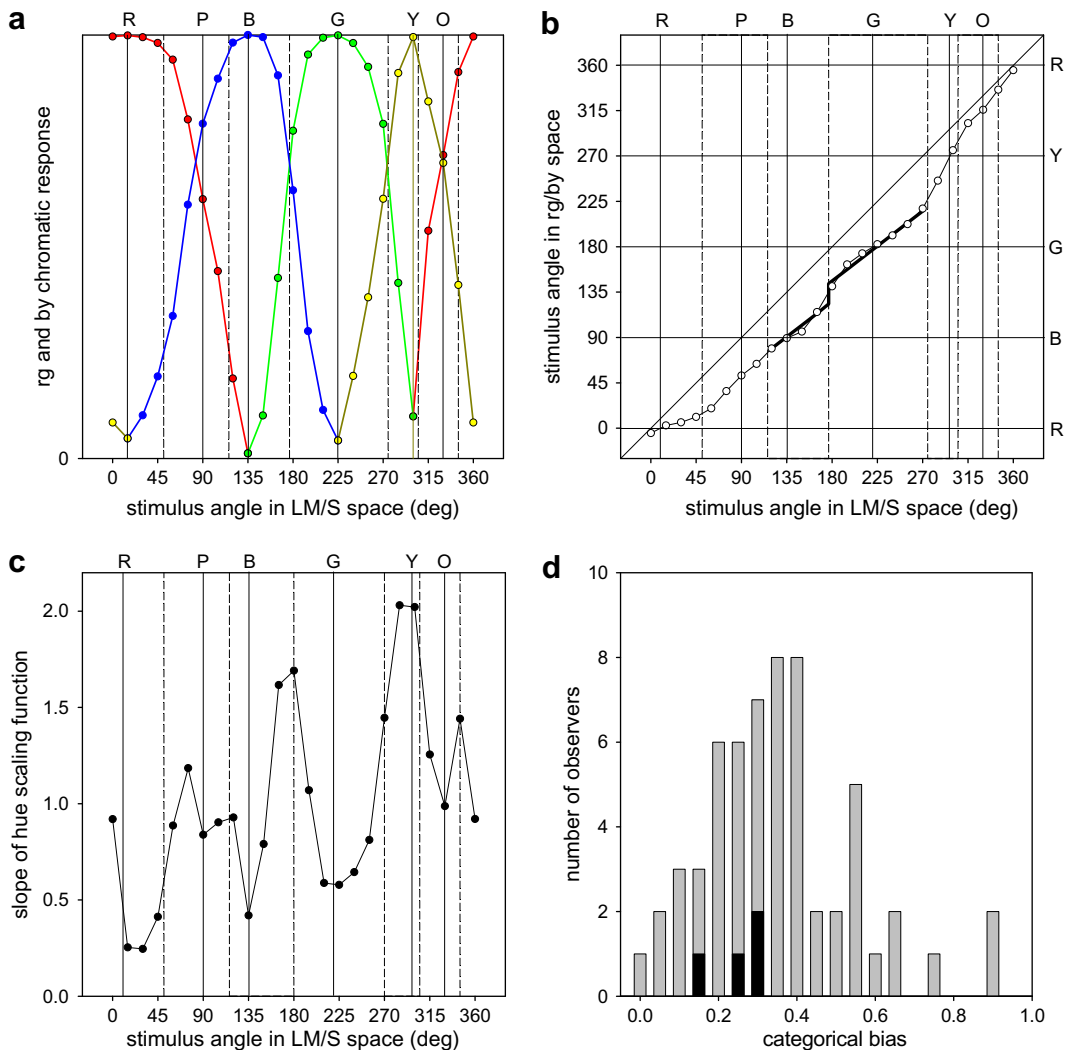


Fig. 8. Categorical biases in hue scaling. (a) Estimates of the average normalized red–green and blue–yellow chromatic response functions from the study of Malkoc et al. (2005). (b) The corresponding mean hue scaling function (unfilled symbols). Vertical and horizontal lines show the stimulus angles corresponding to red, purple, blue, green, yellow or orange foci (solid) or the boundaries between adjacent categories (dashed). Thick fitted line shows the blue and green settings predicted by the estimated bias of 0.23. (c) Local slope of the average hue scaling function. (d) Categorical biases estimated in the blue and green settings for each of the 59 observers tested. Black bars show biases estimated for four additional observers from the study of De Valois et al. (1997). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

proportion that is also significant by a simple sign test ($p = .035$). This raises the question of whether the CP effects for grouping are actually weaker than the CP effects that have been reported previously for reaction times. A direct comparison is difficult because it would require assessing the effective contrast of stimuli from how reaction times vary as a function of the magnitude of the stimulus differences, and typically only a small number of stimulus levels are tested. Our analyses do provide a measure of the actual magnitude of CP. If this magnitude is similar then our results suggest that it was too weak to significantly bias the settings for the majority of observers and that on average it was consistent with only a modest (~10%) contribution of the color category to the effective contrast of our stimuli. This is at odds with the strong

relativistic notion that color appearance is determined by linguistic categories, but does not preclude the potential for a weak linguistic influence on the perception or behavioral response to color.

A similar analysis of hue scaling suggested a very different magnitude of effects. In this case stronger categorical biases were evident in many observers' judgments, with an average bias of 0.35. How can this be reconciled with the much weaker effects we found for perceptual grouping of color? One possibility is that color appearance as measured by scaling draws on different signals from the mechanisms driving spatial grouping. The role of color can in fact vary widely in different visual judgments (Livingstone & Hubel, 1988), yet these differences usually involve the relative importance of luminance signals, which as noted

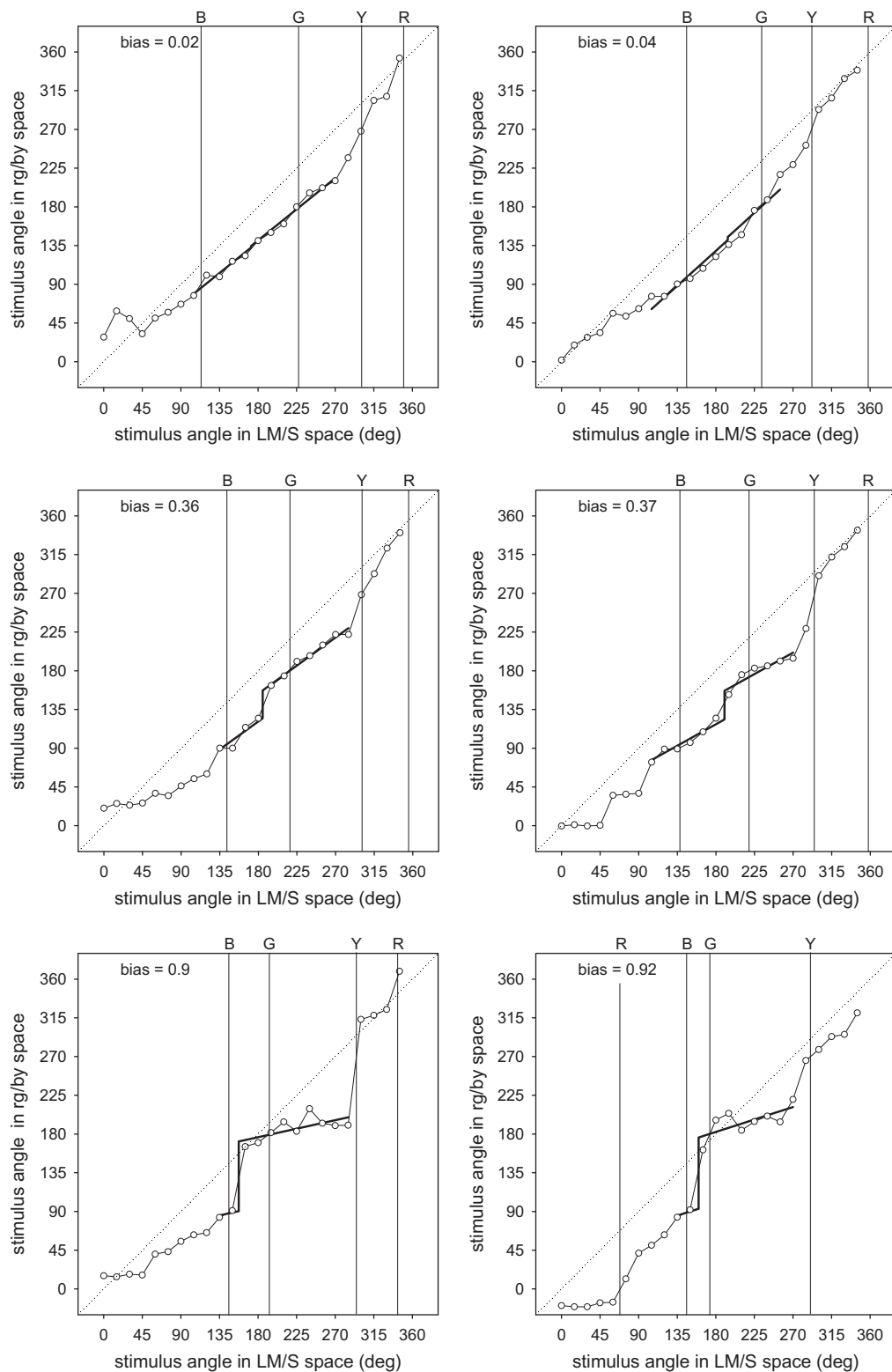


Fig. 9. Individual hue scaling functions and the estimated biases for observers who showed the least categorical bias in the blue and green ratings (top row), the most bias (bottom row), or biases close to the mean for all observers (middle row).

were unlikely to significantly contribute to our grouping results. An alternative possibility is that verbal categories

were influencing the scaling ratings rather than the percept. That is, in the act of deciding the relative amounts

of blue and green, say, in a stimulus, recourse to the labels “blue” and “green” may itself have led many observers to give more weight to the categorical component in their response. This possibility is reminiscent of the effects reported for color similarity judgments by Kay and Kempton (1984). They measured perceptual differences for color triads that spanned the blue–green boundary in speakers of English and Tarahumara (Uto-Aztecan family, Mexico), a language which lacks separate words for green and blue. English speakers reliably picked a categorical singleton as odd-man-out, in violation of jnd distances. Tarahumara speakers respected jnd distances in their triads judgments, but did show a small CP effect when jnd distances were exactly equated. However, the CP effect for English speakers was eliminated when only two colors of the triad were shown at a time and explicitly described as two shades of the same color (green or blue). These results suggested that verbal categories affected the cognitive strategy used to judge the stimuli rather than directly affecting the percept. By this account, then, language was influencing the scaling response and not the appearance, and categorical effects might necessarily be more strongly manifest when the stimuli are named (Pilling et al., 2003).

Still a further possibility is that verbally encoding the stimuli in hue scaling caused the perceptual representation of the stimulus to be recoded. As discussed above, verbal interference has been found to eliminate categorical color effects in studies where they have been observed. Hue scaling may instead represent a case where a verbally mediated mode of responding is required, since subjects must explicitly decide how much of each category (blue or green) is present in the stimulus (i.e., the task by its very nature requires labeling the colors and thus could not be done under verbal interference). Just as blocking language may reveal only a prelinguistic mode of processing, obligating it in the hue scaling task could potentially reshape the appearance of the stimuli. A related possibility is that categorical effects are more likely to arise when observers overtly compare different colors because these comparisons will elicit some degree of linguistic response (e.g. weighing the perceptual differences between color stimuli in the Kay and Kempton study, or the relative magnitudes of different color categories in hue scaling). In this case tasks like grouping may simply not elicit this response because the grouping task requires the subject to organize the perceptual field but does not require overt comparison of colors. Finally, a further possibility is that the actual act of scaling the hues is itself subject to nonlinearities that could distort the response to the colors even if the color percepts are undistorted (e.g. so that the observer over-reports the contribution of the dominant color even though they do not “over-perceive” it).

One implication of these varying accounts is that hue scaling functions, which have played a central role in characterizing color opponency (Abramov & Gordon, 1994), must be interpreted with caution because they may have the potential to misrepresent the “perceptual” component of the stimulus. More generally, whether and how linguistic biases intrude on the responses to visual stimuli may depend in critical yet subtle ways on both the specific task

and the instructions, which may in part be why a clear conclusion about the role of language on color perception has remained elusive. Regardless of its resolution, viewed from a cognitive perspective, it may turn out in the long run to be less important to know if linguistically driven category effects are strictly speaking perceptual than to know how automated they are, under what conditions they are suppressed, and what role they play in our everyday interactions with the environment and with each other.

Acknowledgments

Supported by NIH EY-10834. We are grateful to Angela Brown, Debi Roberson, and an anonymous reviewer for their very helpful comments.

References

- Abramov, I., & Gordon, J. (1994). Color appearance: On seeing red- or yellow, or green, or blue. *Annual Review of Psychology*, 45, 451–485.
- Abramov, I., Gordon, J., & Chan, H. (1991). Color appearance in the peripheral retina: Effects of stimulus size. *Journal of the Optical Society of America A. Optics and Image Science*, 8(2), 404–414.
- Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, 57(3), 217–239.
- 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. B., Fargo, L., Olson, C. X., & Smallman, H. S. (1989). Category effects in color memory. *Color Research and Application*, 14, 229–234.
- 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., Roorda, A., Yamauchi, Y., Calderone, J. B., Metha, A., Neitz, M., et al. (2000). Functional consequences of the relative numbers of L and M cones. *Journal of the Optical Society of America A. Optics and Image Science and Vision*, 17(3), 607–614.
- Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics*, 51(5), 465–472.
- Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal Psychology*, 49(3), 454–462.
- 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.
- Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, 398(6724), 203–204.
- Davies, I. R., & Corbett, G. G. (1997). A cross-cultural study of colour grouping: Evidence for weak linguistic relativity. *British Journal of Psychology*, 88(Pt 3), 493–517.
- De Valois, R. L., De Valois, K. K., Switkes, E., & Mahon, L. (1997). Hue scaling of isoluminant and cone-specific lights. *Vision Research*, 37(7), 885–897.
- Derrington, A. M., Krauskopf, J., & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *Journal of Physiology*, 357, 241–265.
- Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., et al. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences of the United States of America*, 104(3), 1097–1102.
- Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, 44(3), 227–240.
- 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.
- Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, 105(47), 18221–18225.
- 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.

- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language*, *105*(2), 91–98.
- Harnad, S. (Ed.). (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In *Categorical perception: The ground work of cognition* (pp. 1–25). Cambridge: Cambridge University Press.
- Hardin, C. L., & Maffi, L. (Eds.). (1997). *Color categories in thought and language*. Cambridge: Cambridge University Press.
- Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Interdisciplinary Reviews: Cognitive Science*, *1*, 65–78.
- Hanley, J. R., & Roberson, D. (2011). Categorical perception effects reflect differences in typicality on within-category trials. *Psychonomic Bulletin & Review*, *18*(2), 355–363.
- Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review*, *64*(Part 1(6)), 384–404.
- Kay, P., & Kempton, W. M. (1984). What is the Sapir–Whorf hypothesis? *American Anthropologist*, *86*, 65–79.
- 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.
- Kay, P., & Regier, T. (2006). Language, thought and color: recent developments. *Trends in Cognitive Sciences*, *10*(2), 51–54.
- Krauskopf, J., Williams, D. R., & Heeley, D. W. (1982). Cardinal directions of color space. *Vision Research*, *22*(9), 1123–1131.
- Krauskopf, J., Williams, D. R., Mandler, M. B., & Brown, A. M. (1986). Higher order color mechanisms. *Vision Research*, *26*(1), 23–32.
- Kuehni, R. G. (2004). Variability in unique hue selection: A surprising phenomenon. *Color Research and Application*, *29*, 158–162.
- Lennie, P., Pokorny, J., & Smith, V. C. (1993). Luminance. *Journal of the Optical Society of America A. Optics and Image Science*, *10*(6), 1283–1293.
- Lieberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368.
- 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. (2009a). Color difference scaling at the blue–green color category boundary as a test of the Sapir–Whorf hypothesis [Abstract]. *Journal of Vision*, *9*, 340A.
- Lindsey, D. T., & Brown, A. M. (2009b). 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.
- Livingstone, M., & Hubel, D. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, *240*(4853), 740–749.
- MacLaury, R. E., Paramei, G. V., & Dedrick, D. (Eds.). (2007). *Anthropology of colors*. Amsterdam: John Benjamins.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203.
- MacLeod, D. I., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America A. Optics and Image Science*, *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 and Image Science*, *22*(10), 2154–2168.
- Mollon, J. D. (1989). Tho' she kneel'd in that place where they grew. The uses and origins of primate colour vision. *Journal of Experimental Biology*, *146*, 21–38.
- Moore, C. M., & Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 339–352.
- Mullen, K. T., & Kulikowski, J. J. (1990). Wavelength discrimination at detection threshold. *Journal of the Optical Society of America A. Optics and Image Science*, *7*(4), 733–742.
- Ozgen, E., & Davies, I. R. (2002). Acquisition of categorical color perception: A perceptual learning approach to the linguistic relativity hypothesis. *Journal of Experimental Psychology: General*, *131*(4), 477–493.
- 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. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*(10), 439–446.
- 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.
- 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., & Hanley, J. R. (2010). Relatively speaking: An account of the relationship between language and thought in the color domain. In B. C. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience* (pp. 183–198). New York: Oxford University Press.
- Roberson, D., Hanley, J. R., & Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition*, *112*(3), 482–487.
- 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.
- Siok, W. T., Kay, P., Wang, W. S. Y., Chan, A. H. D., Chen, L., Luke, K. K., et al. (2009). Language regions of brain are operative in color perception. *Proceedings of the National Academy of Sciences of the United States of America*, *106*, 8140–8145.
- Tailby, C., Solomon, S. G., & Lennie, P. (2008). Functional asymmetries in visual pathways carrying S-cone signals in macaque. *Journal of Neuroscience*, *28*(15), 4078–4087.
- Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 194–214.
- 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., Malkoc, G., Bilson, A. C., & Webster, S. M. (2002). Color contrast and contextual influences on color appearance. *Journal of Vision*, *2*(6), 505–519.
- 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 and Image Science and Vision*, *17*(9), 1545–1555.
- Werner, J. S., & Schefrin, B. E. (1993). Loci of achromatic points throughout the life span. *Journal of the Optical Society of America A*, *10*(7), 1509–1516.
- Wilmer, J. B. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial Vision*, *21*(6), 561–579.
- 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., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision*, *11*(12–16), 11–24.
- Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501.
- Zhou, K., Mo, L., Kay, P., Kwok, V. P., Ip, T. N., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(22), 9974–9978.