Categorical color constancy for simulated surfaces

Maria Olkkonen   Department of Psychology, Justus Liebig University Giessen, Germany, & Department of Psychology, University of Pennsylvania, USA

Thorsten Hansen   Department of Psychology, Justus Liebig University Giessen, Germany

Karl R. Gegenfurtner   Department of Psychology, Justus Liebig University Giessen, Germany

Color constancy is the ability to perceive constant surface colors under varying lighting conditions. Color constancy has traditionally been investigated with asymmetric matching, where stimuli are matched over two different contexts, or with achromatic settings, where a stimulus is made to appear gray. These methods deliver accurate information on the transformations of single points of color space under illuminant changes, but can be cumbersome and unintuitive for observers. Color naming is a fast and intuitive alternative to matching, allowing data collection from a large portion of color space. We asked observers to name the colors of 469 Munsell surfaces with known reflectance spectra simulated under five different illuminants. Observers were generally as consistent in naming the colors of surfaces under different illuminants as they were naming the colors of the same surfaces over time. The transformations in category boundaries caused by illuminant changes were generally small and could be explained well with simple linear models. Finally, an analysis of the pattern of naming consistency across color space revealed that largely the same hues were named consistently across illuminants and across observers even after correcting for category size effects. This indicates a possible relationship between perceptual color constancy and the ability to consistently communicate colors.

Keywords: color appearance/constancy, color vision, color categories

Introduction

Perceived colors of objects do not change much when seen under different illuminations, even though the light coming to the eye from a surface depends both on surface reflectance and on the illumination. This ability, called color constancy, is an important factor in the recognition of objects in the real world. Color constancy has traditionally been characterized with either asymmetric matching, where two stimuli embedded in different contexts are matched (e.g. Arend & Reeves, 1986; Bäuml, 1999; Brainard, Brunt, & Speigle, 1997; Brainard & Wandell, 1992), or with achromatic settings, where a stimulus is made to appear achromatic under different illuminants (e.g. Bäuml, 1994; Brainard, 1998; Helson & Michels, 1948; Werner & Walraven, 1982). Asymmetric matches may be collected at any point in color space, but the stimulus chromaticities used have generally not spanned the whole range of colors, and the task of matching stimuli over different illuminants might not be intuitive for observers (cf. Brainard et al., 1997). Achromatic adjustments, on the other hand, are usually easy for observers, but data are collected only for one point in color space. Measuring only a few points in color space would be generalizable to the appearance of all colors only

in the case if the illuminant-induced color appearance changes were uniform over the whole color space. Speigle and Brainard (1999) compared data from achromatic settings and asymmetric matching and concluded that asymmetric matches could be predicted from achromatic settings, at least with identical viewing conditions. However, comparisons such as these are rare, and it remains unclear whether the transformations are indeed uniform when the stimuli span a larger portion of color space.

As stated above, people rarely make mistakes about surface colors in real life. However, color constancy in the laboratory is often less than perfect, except for conditions where the whole experimental room was lit by one illuminant (Hansen, Walter, & Gegenfurtner, 2007; Murray, Daugirdiene, Vaitkevicius, Kulikowski, & Stanikunas, 2006; Rinner & Gegenfutner, 2000). Generally, color constancy in the laboratory improves when the number of relevant cues to the illumination is increased (Jin & Shevell, 1996; Kraft & Brainard, 1999; Yang & Maloney, 2001). One reason for this seeming discrepancy between real life and laboratory conditions might be that studies on color constancy did not often differentiate between hue/saturation matches and surface color matches, even though this would seem like an important distinction to make when studying color constancy for naturalistic scenes. Consider a

scene where one region is lit by direct sunlight, and another region is in shadow (cf. Figure 1 in Zaidi & Bostic, 2008). We are usually able to tell which objects have the same surface color in the two regions, although we are also aware of the differences in the absolute chromaticities of the objects in the two regions that we judge as having the same underlying surface color. Indeed, subjects sometimes show less color constancy when they are asked to match hue and saturation of a stimulus than when they are asked to match surface reflectance (Arend & Reeves, 1986; Bäuml, 2001). An alternative to matching methods that overcomes this confusion between surface and hue matches are forced-choice paradigms, where subjects are asked to identify surfaces rather than to match them across two contexts (Bramwell & Hurlbert, 1996; Khang & Zaidi, 2002; Zaidi & Bostic, 2008). The advantage of these methods is that they measure subjects' ability to recognize surfaces as being the same when the illuminant is varied without making assumptions about subjective appearance.

Color naming is similar to typical forced-choice paradigms in the sense that it requires observers to choose one color name from a set of color names that best describes a given stimulus, without asking subjects to match appearance across contexts. In addition, it seems like an ecologically valid task for measuring real-world color constancy. Consider a green apple either under bluish skylight or under yellow-orange sunlight: the apple would most probably be called green under both illuminants, even though the light signal at the retina changes drastically due to the change in the illuminant spectrum. Low-level processes, such as chromatic adaptation at the photoreceptors, account for some of this constancy, but color categorization has also been hypothesized to play a role in stabilizing color appearance across viewing conditions (Jameson & Hurvich, 1989; Smithson, 2005).

A possible shortcoming of color naming is its poor resolution. Humans are able to discriminate thousands of colors (Linhares, Pinto, & Nascimento, 2008; Nickerson & Newhall, 1943; Pointer & Attridge, 1998), but color names are limited to a few discrete categories, with the number varying somewhat between cultures (Berlin & Kay, 1969; Kay & Regier, 2003). However, Hansen et al. (2007) showed in a recent study that if the whole color space is covered by test stimuli, the categorical nature of the responses does not hinder very accurate estimates of the achromatic point in color space under each test illuminant. Indeed, the fitted achromatic points were very well constrained by the naming responses since each point—with a certain grain—of color space was taken into account.

Color naming has been previously used for investigating adaptation effects on color appearance (Hansen et al., 2007; Jacobs & Gaylord, 1967; Smithson & Zaidi, 2004; Speigle & Brainard, 1996; Troost & de Weert, 1991; Uchikawa, Emori, Toyooka, & Yokoi, 2002; Uchikawa, Uchikawa, & Boynton, 1989b; Uchikawa, Yokoi, & Yamauchi, 2004), effects of narrow achromatic backgrounds on color appearance (Uchikawa, Uchikawa, & Boynton, 1989a), as well as

changes in cone weights caused by incremental and decremental colored backgrounds (Chichilnisky & Wandell, 1999). Jacobs and Gaylord (1967) measured adaptation to spectral narrow-band lights and found the color naming method to be as accurate as and more intuitive than asymmetric matching for measuring adaptation effects. Troost and de Weert (1991) as well as Uchikawa et al. (2004) measured color constancy with both asymmetric matching and color naming and found higher color constancy performance with the color naming task. However, neither of these studies equated displays across the two tasks, which might explain the disagreement with the results from Speigle and Brainard (1996), who found comparable constancy for matching and naming. Hansen et al. (2007) investigated the effect of spatial and temporal context on color constancy with color naming. Hansen and colleagues found almost complete color constancy under full-field illumination, and gradually less constancy when the information to the illuminant was decreased.

None of the above studies employed either real or simulated surfaces with known reflectance spectra. Speigle and Brainard (1996) used real surfaces but they extended the stimulus gamut by combining a projected image with the surfaces, thus not analyzing the data in terms of surface reflectance. Troost and de Weert (1991) simulated their stimuli under various illuminants, but did not use surface reflectance functions and illuminant spectra for the simulation, but rather a type of von Kries transform. It is thus not possible to say from these studies which *surfaces* were classified in the same category over observers, illuminants or repetitions. Here we extend the color naming paradigm by using Munsell chips with known reflectance spectra simulated under different illuminations as approximations of natural surfaces. We analyze the data in a way that allows us to estimate i) whether all regions of color space remain equally perceptually stable under illuminant changes, ii) how this depends on the amount of contextual information, and iii) whether there is a relationship between naming consistency across illuminants (color constancy) and naming consistency across observers. We found that color constancy was high for some hues but not for others, and that naming consistency tended to be higher for stimuli remote from the category boundary. Also, naming consistency across illuminants was similar to naming consistency across observers in magnitude and in its pattern across stimulus hue, saturation, and lightness, indicating a possible relationship between color constancy and communication about color.

Methods

Observers

Three naive observers and one of the authors (MO) participated in the experiment. All had normal color

vision as tested with the Ishihara color plates and normal or corrected to normal visual acuity.

Apparatus

The experiment took place in a viewing chamber (2.2 m high × 2.4 m wide × 1.3 m deep). The stimuli were displayed on a Sony Multiscan GDM-F520 monitor with a spatial resolution of 1280 × 1024 pixels and a refresh rate of 100 Hz. The monitor was driven by an NVIDIA graphics card and had a color resolution of 8 bits per channel. The monitor was placed in a black tunnel behind the chamber and was viewed through a 10 × 8 degree aperture in the back wall. The whole back wall subtended 45 degrees horizontally and 64 degrees vertically. The chamber was illuminated with two sets of three fluorescent lamps (red, green and blue) placed behind a diffusing sheet on both sides of the chamber. The output of each of the monitor primaries across the whole voltage range was measured with a UDT Instruments model 370 optometer with a model 265 photometric filter. Spectra of the monitor primaries were measured with a Photo Research PR-650 spectroradiometer. The output of the lamps at different voltages and the chromaticities of the primaries were measured with the PR-650 spectroradiometer. The lamps and monitor phosphors were corrected for nonlinearities in the input/output relationship with look-up tables, and a transformation matrix was calculated to convert between the lamp and monitor primaries. The calibration of the set-up is described in detail in Rinner and Gegenfutner (2000).

The experiments were written in Matlab (The Mathworks, Inc.) with the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997).

Stimuli

Stimuli were uniformly colored two degree discs presented in the center of the monitor screen. Stimulus colors were chosen from the matte Munsell collection. The Munsell color space is a color order system developed by the artist Albert Munsell (Munsell, 1912). Different colors are ordered with approximately equal perceptual distances along a hue, chroma (saturation) and value (lightness) dimension. Value varies between 0 (black) and 10 (white), chroma between 1 and some number depending on the hue and value of the particular sample. Hue varies in 100 steps (40 in the printed samples) around the hue circle. Altogether 1269 Munsell samples are reproduced in the matte collection of the Munsell book of color. The reflectances of the samples for the monitor simulations were acquired from the Database of the University of Joensuu Color Group (http://spectral.joensuu.fi/). All those Munsell chips with value between 4 and 7 were chosen that fitted the monitor gamut under all experimental illuminants. This resulted in 469 samples from the possible 1269 chips. The Judd-Vos corrected CIE chromaticities of the chips simulated under a neutral illuminant are shown in Figure 1A. The average luminance of the whole stimulus collection when simulated under a
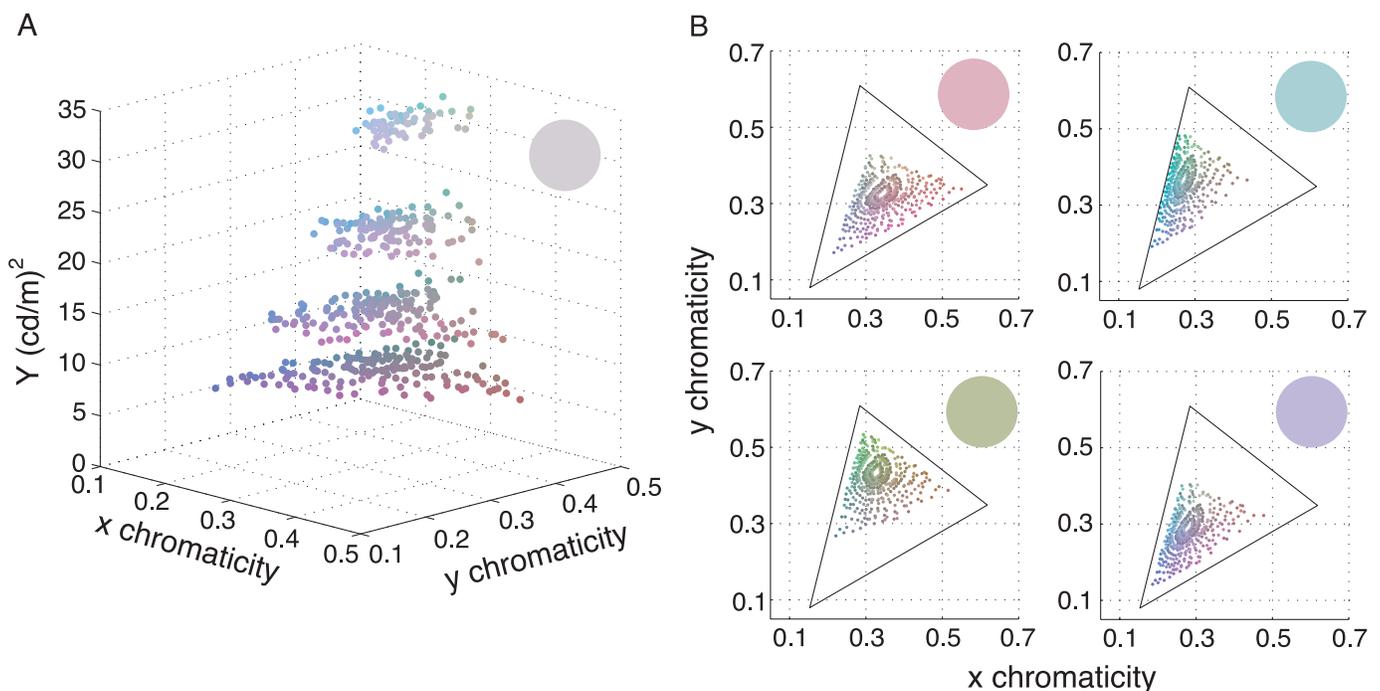

Figure 1. The Munsell chip collection used in this study. A: Munsell chips under a neutral illuminant are plotted in the CIE xyY space. Symbol colors indicate the color of the reflected light from each chip. B shows the projection of the chip chromaticities under four chromatic illuminants to the CIE x,y plane (illuminants from top left clockwise: reddish, bluish-green, greenish-yellow, violet). Insets show the color of a neutral chip under each illuminant. Black triangles indicate the monitor gamut.

neutral illuminant was 9.3, 14.8, 22.5 and 32.5 cd/m^2 for Munsell values 4, 5, 6 and 7, respectively.

The experiment was run under two viewing conditions. In the full-cue viewing condition, the monitor background had the same chromaticity as the surrounding wall illuminated by the lamps. The luminance of the monitor background and of the surrounding wall was 18 cd/m^2, which corresponded to the mean luminance of the whole stimulus collection under neutral illumination. Thus, half of the stimuli (values 4 and 5) were decrements relative to the background, and half (values 6 and 7) increments. In this viewing condition, the immediate background of the stimulus had the chromaticity of the illuminant. In the reduced-cue viewing condition, the monitor background was black, and only the wall was illuminated with a mean luminance of 18 cd/m^2. In this condition, all stimuli were increments relative to the local surround. The only information about the illuminant came from the peripheral border between the black monitor background and the illuminated wall, and from the mean chromaticity of the simulated stimulus collection.

Illuminant simulation

Five illuminants were generated by combining the output of three fluorescent lamps in different proportions. Four of the illuminants were chosen from the cardinal axes of DKL color space, and can be described as reddish, bluish-green, violet and greenish-yellow. The fifth illuminant was metameric to the standard daylight illuminant D65 (see Table 1 for Judd-Vos corrected CIE chromaticities of all illuminants). The DKL color space is a linear transformation of the cone excitation space based on the Smith and Pokorny cone fundamentals (Derrington, Krauskopf, & Lennie, 1984; Krauskopf, Williams, & Heeley, 1982; MacLeod & Boynton, 1979; Smith & Pokorny, 1975). On one axis, the L and M cone excitations vary in opposition so as to keep their sum constant (perceptually reddish-greenish), and on the second axis the S cone excitations vary in opposition to the sum of the L and M cone excitations (perceptually yellowish-violet). The sum of the L and M cone excitations (luminance) varies

Illuminant	x	Y	Y (cd/m^2)
Neutral	.298	.341	87.9
Red	.342	.323	95.4
Bluish-green	.277	.364	118
Greenish-yellow	.325	.432	96.2
Violet	.277	.288	94.3

Table 1. Judd-Vos corrected CIE chromaticities of the different illuminants measured off a white reference surface. The mean luminance of the whole stimulus collection under the experimental illuminants was 18 cd/m^2 on average.

on the third axis. The DKL axes were scaled according to the maximum contrast in the cones produced by the monitor primaries and normalized to the range [−1 1].

For simulating the Munsell chips, the spectra of each illuminant were measured with a PR-650 spectroradiometer off a white reference surface (Photo Research SR-2) that was placed on the back wall of the viewing chamber. The chip collection was simulated under each illuminant by taking a wavelength by wavelength product between the reflectance spectra of the chips and each illuminant spectrum and using this value to derive Judd-Vos corrected XYZ values and device-dependent RGB values. The stimulus collection simulated under D65 is shown in Figure 1A, and under the four chromatic illuminants in Figure 1B.

Procedure

Observers viewed the display in the front end of the chamber from a distance of 187 cm with their heads stabilized with a chin rest. After observers had adapted to the illumination for two minutes, the first stimulus was displayed on the screen for 500 ms. Observers' task was to assign the color of the stimulus to one of nine color categories (green, turquoise, blue, purple, red, orange, yellow, brown, gray). The color names were given in German (grün, türkis, blau, lila, rot, orange, gelb, braun, grau). We chose these categories (8 basic color terms + turquoise) because they resulted in the clearest division of the DKL color space in a previous study (Hansen et al., 2007). Observers were instructed to classify any achromatic chip in the gray category independent of lightness. Observers responded at their own pace by pressing one of nine keys, after which the next stimulus appeared. Each of the 469 chips was categorized once under each illuminant. The chips were presented in a randomized order. Different illuminants were run in separate sessions, whose order was counterbalanced across observers. The experiment was replicated after six months to verify test–retest reliability within observers.

In addition to color naming, observers chose the prototypes, i.e. the best examples, for each category out of the whole collection of real Munsell chips under daylight illumination.

Data analysis
Fitting category boundaries

In order to divide the naming space of each observer into categories, we fitted category boundaries to the color naming data for each subject and each illumination condition. For the fitting, the stimuli were represented according to their surface reflectance so that the only difference between conditions was the pattern of color names given to the stimuli. The boundaries were modeled as straight lines constrained to converge on one point in color space. The fitting was done for each Munsell value

separately. The chromaticities of the stimuli under each illumination were transformed from XYZ values to the DKL color space to facilitate the comparison to the study of Hansen et al. (2007). The categories for orange and brown were pooled for the boundary fitting, because their centroids, i.e. average chromaticities, were generally the same but depended on stimulus luminance.

The category boundaries and achromatic points in each illumination condition were determined with a grid search procedure (Hansen et al., 2007). Best fitting boundaries were defined as boundaries that led to the most chips being classified in the correct category and least chips outside of the correct category. Preliminary category boundaries were defined as the average of the centroids of two adjacent categories. Boundary angles were varied in 1 degree steps around the preliminary boundary, and the amount of correct and false classifications was calculated for each boundary angle. The convergence point, to which all boundaries were constrained to converge, was varied on a discrete grid with a spacing of 0.01 DKL units. This procedure would result in one vector for each convergence point with the classification errors for each possible angle for each category. The classification errors were pooled across categories for each convergence point, and the convergence point along with the boundaries with the fewest overall classification errors were chosen.

Transformations in color categories

We modeled effects of illumination changes on color naming by testing two different types of transformations in color categories. We used the category boundaries from the neutral condition as a baseline and sought an optimal fit between the baseline boundaries and naming data under each chromatic illuminant by 1) shifting the naming data two-dimensionally in the isoluminant plane of the DKL color space until the boundaries were optimally aligned with the data, or by 2) shifting and scaling the naming data in the isoluminant plane until an optimal alignment was found. Shifting meant adding or subtracting two independently varied constants to/from the x and y coordinates of the data. Scaling meant multiplying the x and y coordinates of the data by two independently varied coefficients. Numerical search in Matlab was used to minimize fitting error. For this analysis, the stimuli were represented according to their chromaticities under each illuminant (i.e., the product of the illumination and surface reflectances).

Color constancy

Color constancy was characterized in several ways. Firstly, the frequency with which each Munsell chip was classified in the same category across the neutral and each chromatic illuminant was calculated for each observer. This index for naming consistency could take on values between 0 (chip classified differently under all illuminations) and 1 (chip classified similarly under all five illuminations),

and is similar to the index defined by Troost and de Weert (1991). We also calculated naming consistency across observers for the neutral illuminant to compare color constancy with inter-observer agreement.

The fact that some color categories are larger than others might influence naming consistency to some extent: in a wide category, some stimuli might not shift to a different category even under a large illuminant change. In that case, even 0% color constancy would not lead to 0% naming consistency. We corrected for this effect in the naming data by calculating a lower bound for naming consistency across illuminants as follows. First, baseline category boundaries were fitted to the color naming data under the neutral illuminant for each subject and each Munsell value separately. Next, the light signals reflecting off each stimulus under each chromatic illuminant were calculated and categorized based on the baseline category boundaries. Finally, a consistency index was calculated with these simulated data as described above. The naming consistency indices derived from the raw naming data were then corrected by first subtracting the simulated consistency indices and then rescaling so that the maximum consistency value remained the same. This correction procedure is essentially the same as ignoring the chip-illuminant combinations for which consistent naming would be expected even in the complete absence of constancy.

The observer consistency index was corrected in a similar manner. As there are no stimulus chromaticity changes across observers, we modeled the effect of category size by rotating the category boundaries in the neutral condition by a random amount uniformly chosen between -60 and $+60$ degrees separately for each observer, after which the observer consistency index was calculated based on the rotated boundaries. The rationale was that the largest categories would be most immune to the rotation and thus would give us an estimate of the effect of category size on the naming consistency across observers.

In addition to naming consistency, color constancy was characterized as the stability of the achromatic points under illuminant changes. We defined achromatic points both as the centroids of the gray category, and as the convergence points of the boundary fitting procedure, since these do not always coincide (Ekroll, Faul, Niederée, & Richter, 2002). A standard color constancy measure that quantifies the change in the achromatic point relative to the magnitude of the illuminant change was calculated for each illuminant change from neutral for both types of achromatic point (Equation 1). For this measure, the chromaticities of the chips under the different illuminants rather than the surface reflectances were used.

$$\mathrm{CI}_{\mathrm{achrom}} = \frac{\mathbf{S_c} \cdot \mathbf{S_p}}{\|\mathbf{S_p}\|}. \tag{1}$$

In Equation 1, vector $\mathbf{S_c}$ is the observed shift in the achromatic point from D65 to a given test illuminant and

vector $\mathbf{S_p}$ the predicted shift of the achromatic point given perfect constancy. Projecting $\mathbf{S_c}$ to $\mathbf{S_p}$ gives the common component of the observed shift in the direction of the predicted shift, and the constancy index is derived by dividing the projection by the magnitude of the predicted shift. Numbers close to 1 indicate high constancy, and numbers close to 0 low constancy.

Calculating color constancy from centroids of other categories than gray is problematic due to unpredictable changes in the form of the stimulus gamut under illumination changes (Speigle & Brainard, [1996]). As a third type of analysis, we calculated color constancy for the chromatic categories with a procedure that does not depend on category centroids. We defined the width of each category by the angle it subtended in DKL color space, and quantified color constancy as the change in angle from the baseline condition for each test condition (Ling, Allen-Clarke, Vurro, & Hurlbert, [2008]). Let θ_{BL} and θ_T be the sizes in degrees of a given category in the baseline and test

condition, respectively, and $\theta_{overlap}$ be the size in degrees of the overlapping portion of the category in the baseline and test condition. The color constancy index is then defined as

$$C_{chrom} = \frac{\theta_{overlap}}{(\theta_{BL} + \theta_T)/2}. \tag{2}$$

If a particular category occupies the same portion of color space under two illuminants, size of the overlapping portion (nominator) is the same as the mean size of the two categories (denominator), and constancy will be close to 1. In the absence of overlap, constancy will be close to 0.

Results

Naming data for one subject is shown in Figure 2 for the four Munsell values separately. Stimulus chromaticities
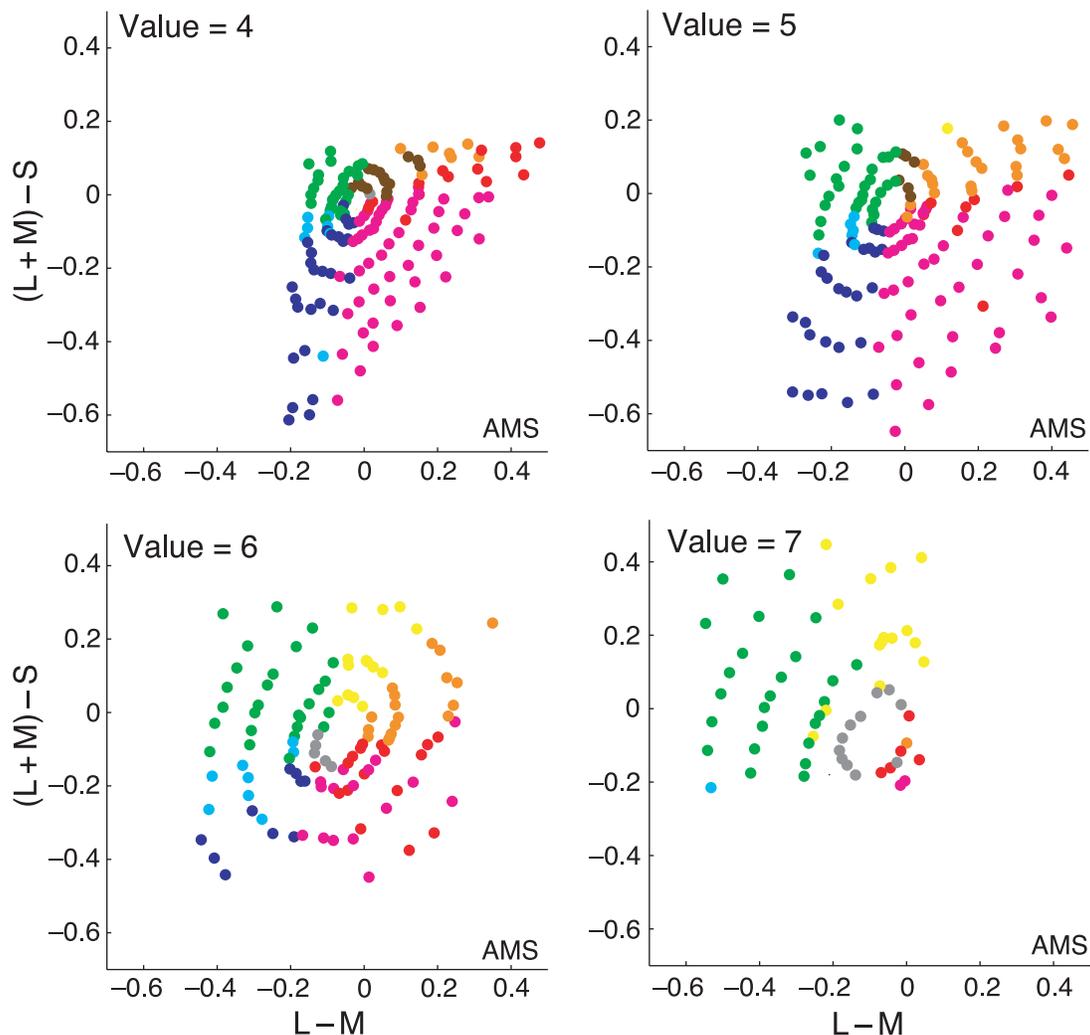

Figure 2. Naming data for subject AMS under the neutral full-field illuminant. Munsell chip chromaticities are plotted in the isoluminant plane of DKL color space. Chips with Munsell values 4 and 5 are plotted in the top row, and chips with values 6 and 7 in the bottom row. Symbol color denotes the color name given to a particular chip.

are plotted in the isoluminant plane of the DKL color space, and symbol colors denote the color names given to a particular stimulus. As expected, some categories were only present at certain luminance levels. For instance, there was no yellow category at the lowest luminance level (upper left panel). There was also no blue category on the highest luminance level (lower right panel), but as Figure 2 shows, there were practically no chips in the blue-purple part of color space that would fit the monitor gamut on the highest luminance level.

Naming consistency

Across illuminants

Figure 3 illustrates the overall uncorrected naming consistency for observer AA and AG under the full-cue (A)

and reduced-cue (B) viewing conditions. Symbols are colored according to the mode color category assigned to that particular stimulus. Symbol size denotes the degree of consistency across illuminants.

Under full-field illumination, on average 50% of the chips had an uncorrected consistency index of 1, i.e. these chips were classified in the same category under all five illuminants. Consistency index averaged over stimuli and observers was 0.8. In the reduced-cue viewing condition, on average 27% of the chips had a consistency index of 1. Consistency index averaged over all chips and observers in this condition was .65.

The simulation for the lower bound of naming consistency across illuminants is shown in Figure 4 with the red dashed curve. Simulated naming consistency was clearly highest for greenish hues, which reflects the large size (over 90 degrees) of the green category. On average
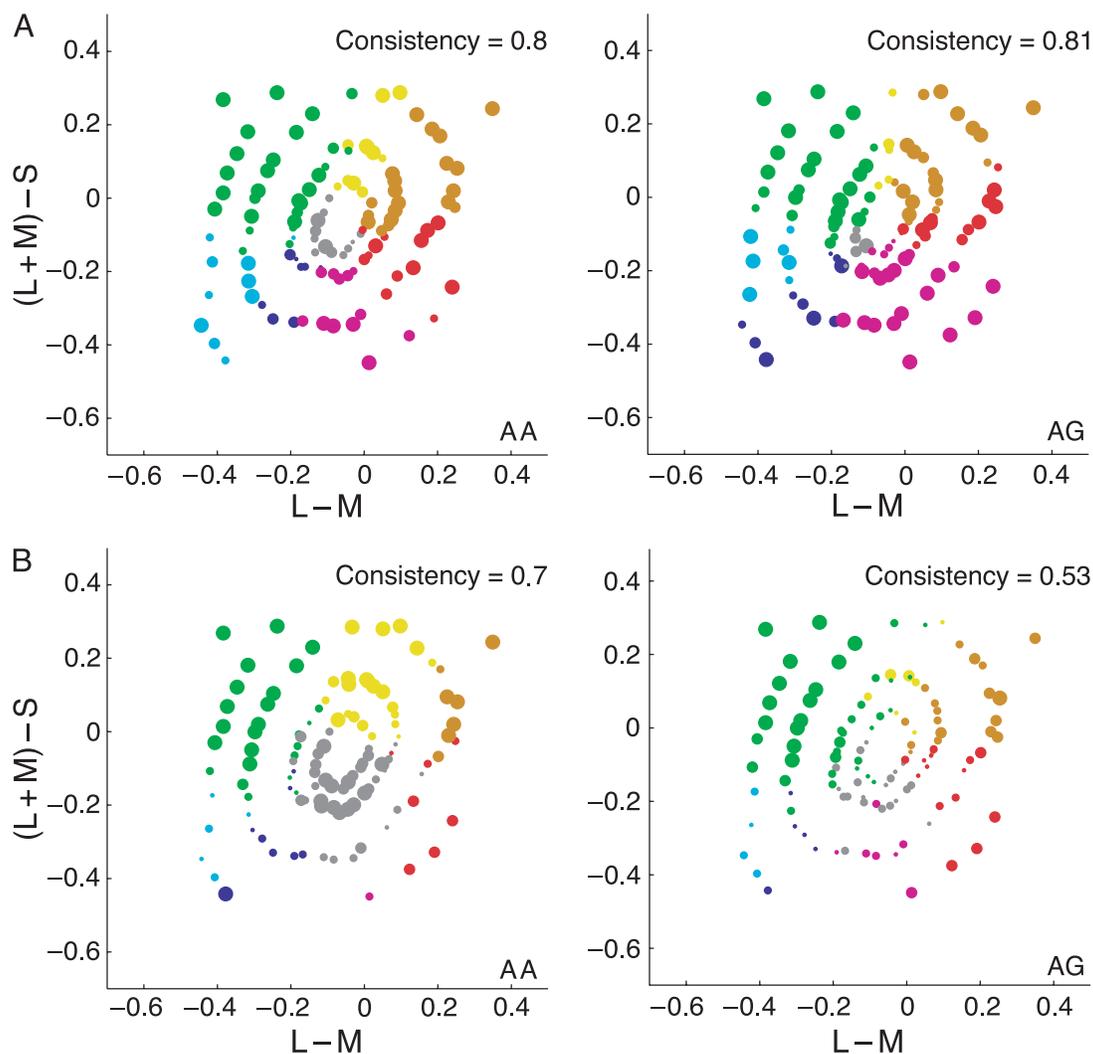

Figure 3. Naming consistency over illuminants for observer AA (left column) and AG (right column). The stimulus collection at value level 6 is plotted in the isoluminant plane of the DKL color space. Symbol size indicates the amount of consistency. A: Full-cue viewing condition. B: Reduced-cue viewing condition.
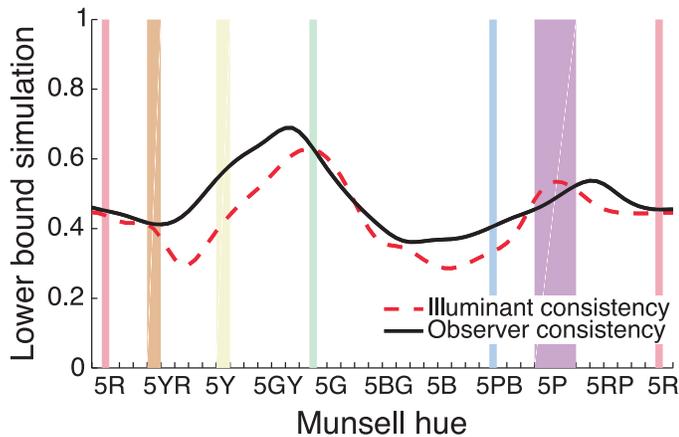
Figure 4. Lower bound simulation for consistency across illuminants (dashed red curve) and across observers (solid black curve) plotted for Munsell hue. The smooth curve was derived by averaging simulated values over two adjacent hues and interpolating between these averages. Vertical bars indicate the range of prototypical hues (from left to right red, orange, yellow, green, blue, purple, red).

4% of the chips remained in the same color category under all illuminants in the simulation, compared to the measured rate of 50%. The lower bound simulation for observer consistency is shown with the solid black curve. No chips remained in the same category across all observers, but on average, the simulated observer consistency was of similar magnitude as the simulated illuminant consistency.

Figure 5 summarizes uncorrected (A) and corrected (B) naming consistency for the full-cue (gray curves) and reduced-cue (black dashed curves) conditions as a function of stimulus hue, chroma, and value. The upper panel of Figure 5A shows the uncorrected illuminant consistency index for Munsell hue averaged over observers. Uncorrected naming consistency peaked near all prototypical hues except for the red one. Consistency increased as a function of stimulus saturation (lower left panel), and remained roughly similar for all lightness levels (lower right panel). Reducing cues to the illuminant (black dashed curves in Figure 5A) caused a decrease in overall naming consistency, but the pattern as a function of hue, chroma, and value remained similar.

The corrected naming consistency indices are shown in Figure 5B. All values above zero in this representation imply higher consistency than would be expected by the mere categorical stability of stimulus chromaticities under illuminant changes. Zero indicates the level of consistency in the absence of color constancy. The upper panel of Figure 5B shows the corrected naming consistency index as a function of Munsell hue averaged over observers. Even after the lower bound correction, naming consistency was not uniform across Munsell space but peaked around

blue, green and orange hues. Consistency was worst for bluish-green and red hues. Naming consistency dropped only slightly from the full-cue (gray line) to the reduced cue (black dashed line) condition; the largest drop was for bluish and yellowish hues. Overall, the pattern of the data in the full-cue and reduced-cue conditions was similar (r = .58, p < .0001).

Comparing 5A to 5B shows that the pattern of the data in 5A could be partly explained by the different category widths causing different amounts of baseline naming consistency. This is especially evident for the green category, for which the uncorrected index in the full-cue condition was close to 1, but the corrected index was around .4 to .6. Whereas the uncorrected index seemed to coincide with most prototypical hues, the corrected index coincided mostly only with the blue and orange prototypes.

The lower left panel of Figure 5B shows corrected naming consistency as a function of Munsell chroma. After the lower bound correction, naming consistency was especially high for medium levels of chroma for both viewing conditions. There was a difference between the viewing conditions only for chromas below 6. For higher chromas, consistency was as good in the reduced-cue condition as in the full-cue condition. The lower right panel of Figure 5B shows the effect of stimulus lightness on corrected naming consistency. In both viewing conditions, consistency was lowest for the highest value. There was also a large drop in consistency from the full-cue to the reduced cue condition for chips at Munsell value 6.

Across repetitions

The experiment was repeated six months after the first data collection to evaluate the consistency of categories over time. Figure 6A shows a similar plot to Figure 3 for subject AA where categories under the neutral full-field illuminant are compared for the two repetitions rather than for the five illuminants. For this observer, consistency over time in the baseline condition was .78, and varied from .58 to .80 (mean .75) between observers. Figure 6B shows the overall uncorrected naming consistency over illuminants and over time for the full-cue viewing condition. Naming consistency was as good or better across illuminants as it was across repetitions. Consistency was overall lower for the reduced-cue condition (6C and 6D) but also approximately of the same magnitude across repetitions and across illuminants.

Across observers

In addition to calculating naming consistency for each observer across illuminants and repetitions, we calculated a consistency index across observers for the neutral illuminant to see how categorical color constancy related to inter-observer agreement on color names. Figure 7A
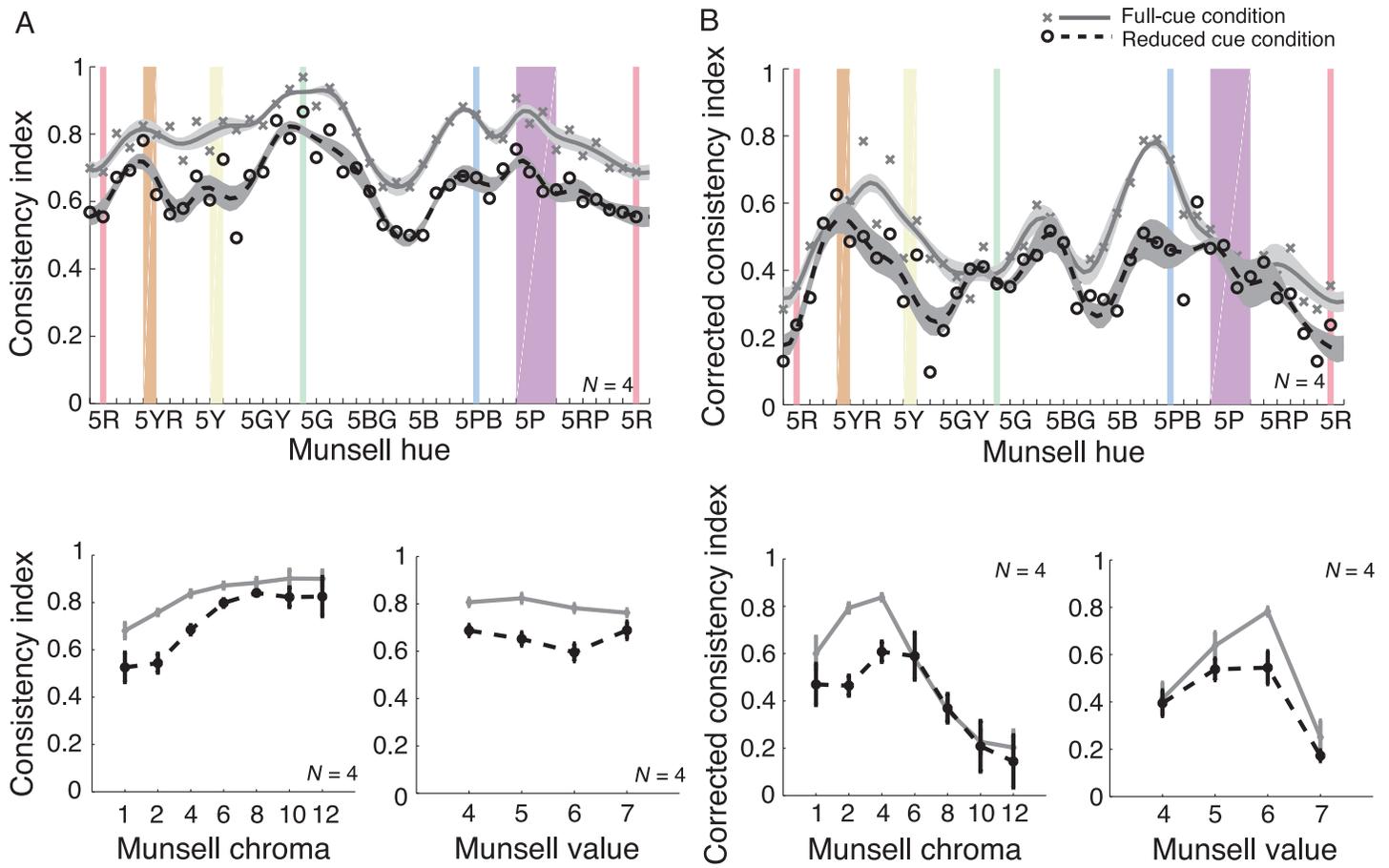
Figure 5. A: Uncorrected naming consistency indices for full-cue illuminants (gray curve) and reduced-cue illuminants (black dashed curve) are plotted as a function of Munsell hue (top panel), Munsell chroma (lower left) and Munsell value (lower right). Vertical bars in the upper panel indicate the range of prototypical hues. The smooth curves drawn through the data points were derived by averaging over two adjacent hues and interpolating between these averages. Shaded areas around the curves in the upper panel, and error bars in the lower panels indicate the standard errors of the means. B: The corrected naming consistency indices are plotted for Munsell hue (top), chroma (lower left) and value (lower right). Details as in A.

shows the comparison between the corrected naming consistency index across observers and the corrected index across illuminants. The observer consistency index varied across hue in a manner that was rather similar to the variation in the illuminant consistency index. However, peaks in the observer consistency index coincided with the prototypes to a larger extent than the peaks in the illuminant consistency index. The largest correspondence between observer and illuminant consistency was for the orange, green and blue hues. The biggest discrepancy was around red and yellow where illuminant consistency was much higher than between observers consistency.

The relationship between observer and illuminant consistency is further illustrated in Figure 7B. Gray crosses denote consistency in the full-cue condition, and black open circles in the reduced-cue condition. Correlation between the corrected observer and illuminant consistency indices was .61 ($p < 0.001$) for the full-cue

condition and .73 ($p < 0.001$) for the reduced-cue condition.

Category boundaries and color constancy

Figure 8A shows naming data for all subjects in the baseline condition under full-field illumination. The best fitting category boundaries are drawn with black lines. Both the category boundaries and the convergence points varied across subjects, but the categories overlapped for the most part (Figure 8B).

The achromatic point does not always need to coincide with the convergence point of category boundaries in color space (Ekroll et al., 2002). However, inspection of the data in Figure 8A indicates that the boundary convergence points were generally close to the stimuli named gray.
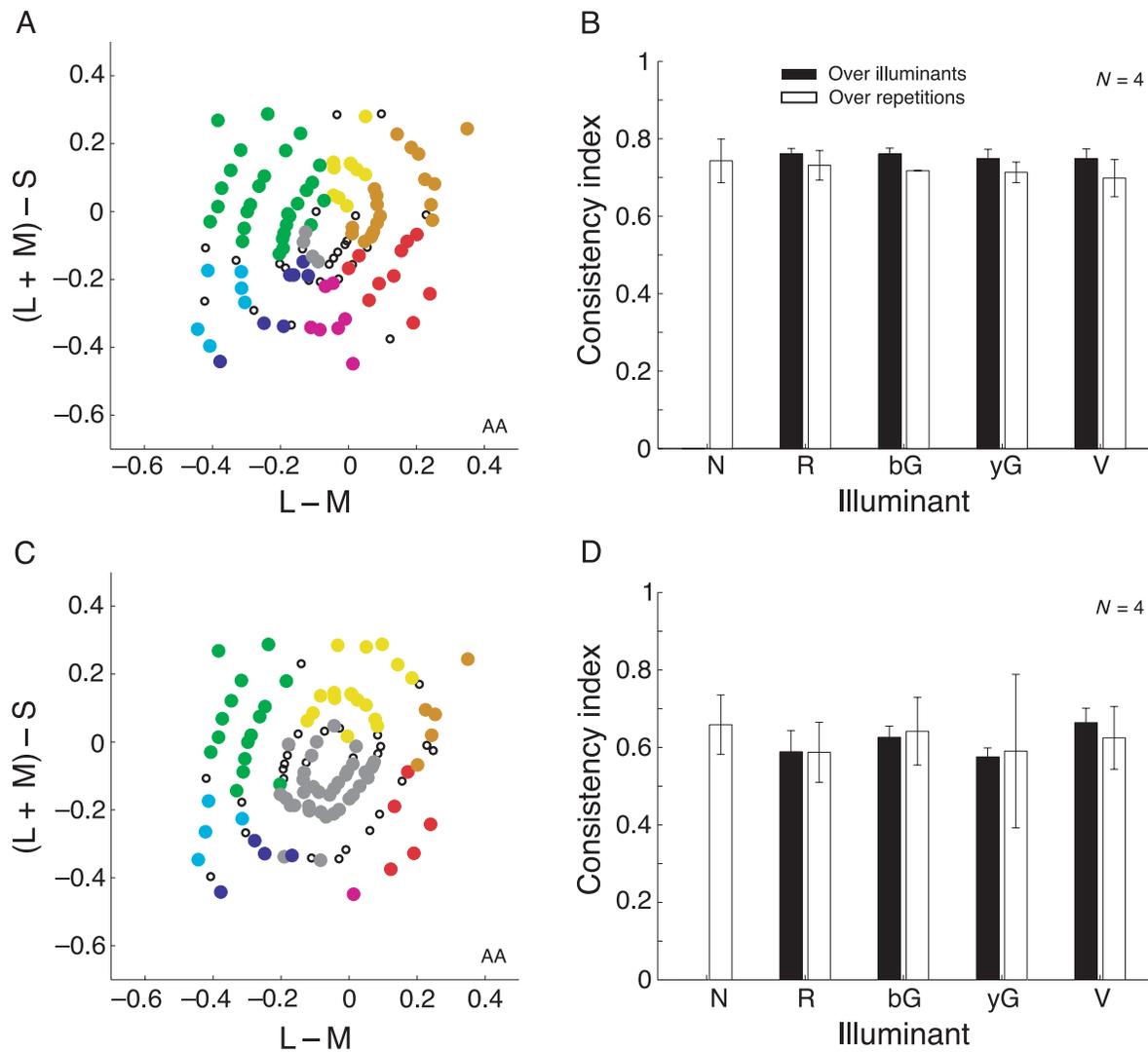
Figure 6. A: Naming consistency under the neutral illuminant over two repetitions for observer AA in the full-cue condition. Chips that were classified in the same category over repetitions are indicated with filled symbols. B: Uncorrected naming consistency across repetitions and across illuminants for the five full-field illuminants. Error bars show the standard errors of the mean. Illuminants are indicated on the abscissa as follows: N: neutral; R: red; bG: bluish-green; yG: yellowish-green, V: violet. C: Naming consistency under the neutral illuminant over two repetitions for observer AA in the reduced-cue condition. D: Uncorrected naming consistency for the reduced-cue illuminants.

Figure 9A plots the average gray category centroids (filled circles) along with the convergence points of the category boundaries (open squares) under the full-field illuminants. Crosses indicate the average chromaticity of the whole stimulus collection under each of the five illuminants. The changes in the achromatic loci closely followed the physical changes in the stimulus chromaticities, which points to high color constancy of the achromatic point. Also, the gray category centroids were slightly shifted to the right from the convergence points, but this tendency was not statistically significant (L − M coordinate: $F(1) = 0.33$ (n.s.); (L + M) − S coordinate: $F(1) = 0.09$ (n.s)).

Achromatic points for the reduced-cue condition are plotted in Figure 9B. There was less agreement between the gray centroids and convergence points in this viewing condition (L − M coordinate: $F(1) = 1.35$ (n.s.); (L + M) − S coordinate: $F(1) = 5.82$, $p = 0.02$). Also, achromatic points followed the changes in stimulus chromaticities to a slightly smaller degree. This drop in color constancy was evident in the color constancy indices, which for the full-cue conditions were on average .96 and .98 for centroids and convergence points, respectively, and for the reduced-cue conditions .84 for both centroids and convergence points.
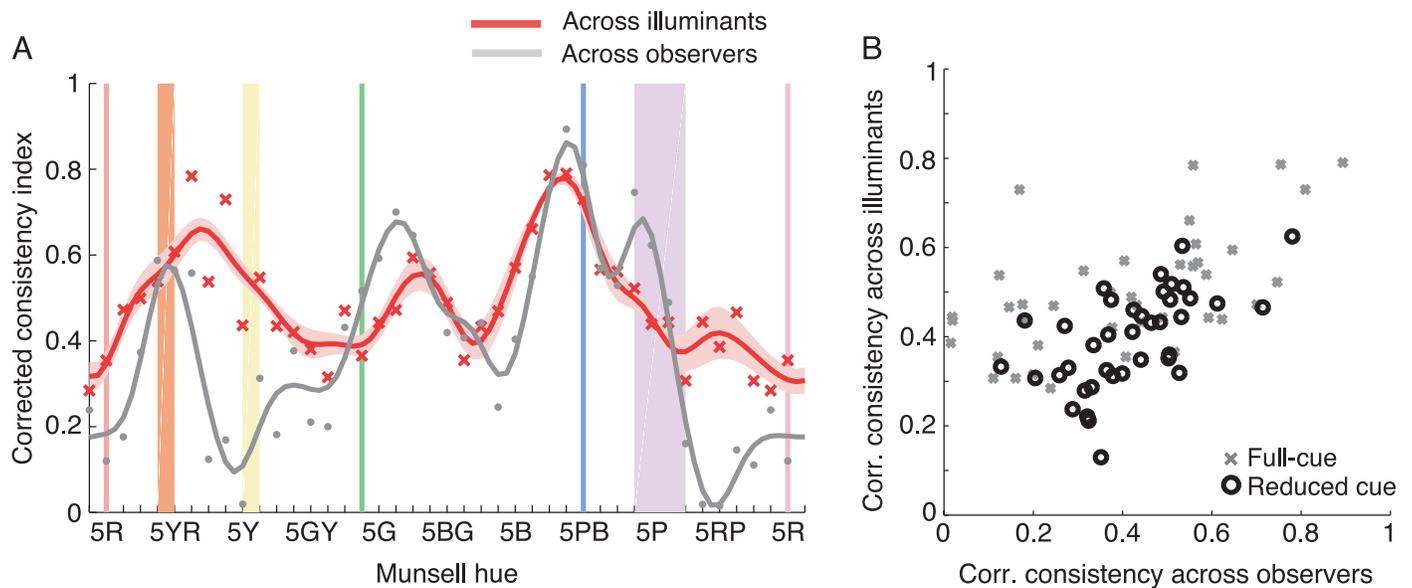
Figure 7. Comparison between naming consistency across illuminants and naming consistency across observers. A: Corrected naming consistency as a function of Munsell hue. Red crosses and the red curve plot naming consistency across illuminants. Gray circles and the gray curve plot naming consistency across observers for the neutral illuminant. Vertical bars indicate the range of prototypical hues. B: Corrected naming consistency across observers is plotted against consistency across illuminants for the full-cue (gray crosses) and reduced-cue (black circles) conditions. Each point indicates one Munsell hue, collapsed over saturation and lightness.

Average color constancy for each chromatic category based on the relative stability of category boundaries is plotted in Figure 10 with black bars, along with the constancy for the gray category centroid. The uncorrected naming consistency indices are plotted with gray bars for comparison. Color constancy indices were generally lower for the chromatic categories (on average .75) than for the gray category (.96). Conversely, naming consistency was slightly lower for the gray category than the chromatic categories (.65 vs. .76). For the gray category, the color constancy index was much higher than the naming consistency index (.96 vs. 65), but for the chromatic categories, the two measures were on average similar (.75 for both).

For the reduced-cue viewing condition, constancy was on average lower than in the full-cue condition, but the pattern of the data was the same: constancy was best for the gray category calculated from the centroid (.84). Constancy pooled over categories was on average .63 and .64, quantified as naming consistency and color constancy, respectively.

Transformations in category boundaries

We tested a simple model on the effect of illumination changes on color categories by taking the fitted boundaries in the baseline condition (neutral illuminant) and searching for the best fit between the baseline boundaries and the color naming data in each test condition (four chromatic illuminants) by either just shifting the color naming data two-dimensionally in the isoluminant plane, or shifting and scaling the naming data in two dimensions. The average errors for these two types of fit for stimuli of Munsell value 6 are plotted in Figure 11A along with a baseline error rate (the goodness-of-fit of the boundaries fitted separately in each condition), and the errors of boundary fits over the two repetitions in each illuminant condition (reliability). The average proportion of chips falling in the wrong category with the given boundaries varies on the ordinate. The two leftmost bars indicate the baseline error of the boundary fitting procedure, i.e., the rate of false classifications when fitting boundaries with grid search for each data set separately. The next two bars show the error when fitting boundaries from the first to the second measurement of a given condition. The next two sets of bars describe the average error when transforming the boundaries from the baseline condition to the data measured in the test conditions with four or two parameters, respectively. The last set of bars show the error when the boundaries from the baseline condition are merely superimposed on the test conditions without any fitting.

Fitting errors for both the shift and the shift & scale models were close to reliability, indicating that both models captured important features of the data. Fitting errors were overall slightly larger for the reduced-cue condition, but the models seemed to work as well as for the full-cue conditions. Shifting and scaling the naming data brought a slight improvement over just shifting the
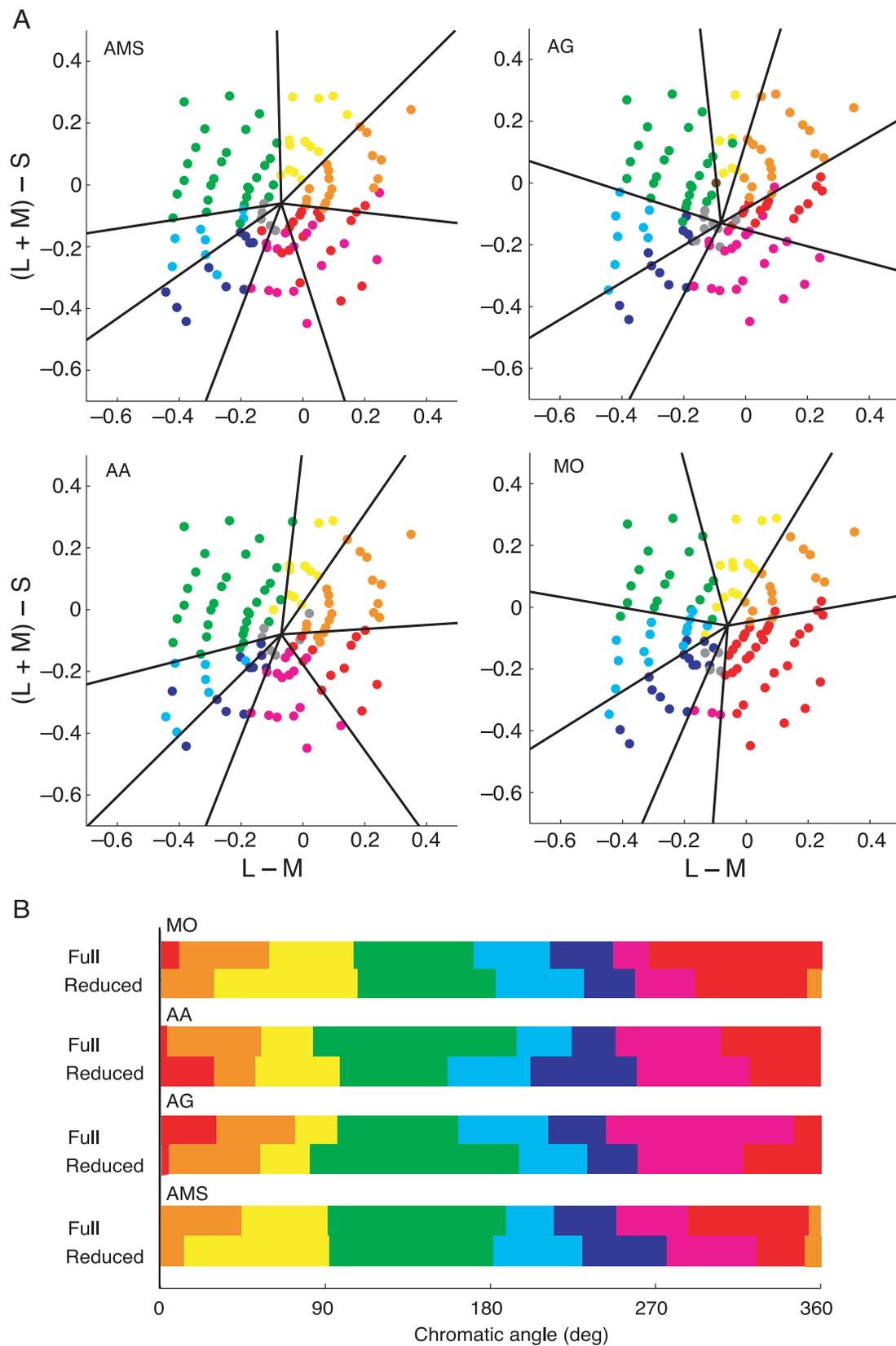
Figure 8. A: Chips with Munsell value 6 under the neutral illuminant are plotted in the isoluminant plane of DKL color space. Each quadrant is naming data for one subject in the baseline condition. Color categories are indicated with symbol colors. Black lines are best fitting category boundaries. B: Categories under neutral illumination are plotted as the color angles in the isoluminant plane between 0 and 360 degrees for all observers, and for the full-cue and reduced-cue viewing conditions.
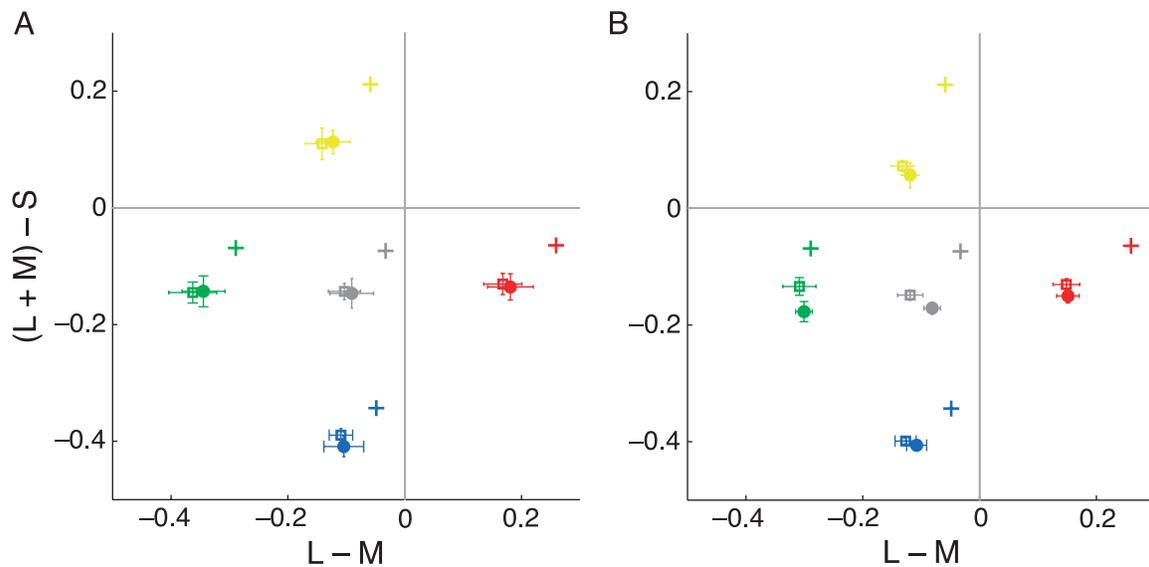
Figure 9. Achromatic points averaged over stimulus value in the full-cue (A) and reduced-cue (B) viewing conditions. Filled circles denote gray category centroids under the neutral and four chromatic illuminants; open squares denote the convergence points of the fitted boundaries under the five illuminants. Error bars are standard errors of the mean. Crosses denote the average chromaticity of the whole stimulus collection under each of the illuminants. Illuminant chromaticities are indicated by symbol colors.

data, but this difference was small compared to the overall improvement from the null model to the shift model.

The data from Figure 11A are broken down per category for the full-cue conditions in Figure 11B and for the reduced-cue conditions in Figure 11C. There were

large differences in the overall goodness-of-fits between categories, especially in the reduced-cue conditions. The baseline fitting errors for the green category, for instance, were minimal, whereas for the turquoise between 30 and 50 percent. However, relative to the baseline error, fitting errors for the models were similar across categories.

Discussion

We used color naming to characterize color constancy under five illuminants in a full-cue and a reduced-cue viewing condition. We quantified color constancy with a naming consistency index that does not depend on any color space metric, but also fitted boundaries to the naming data and calculated a color constancy index for the achromatic point as well as for the chromatic categories. The latter analyses do depend on the color space where the data are represented, and it was interesting to see whether these two types of analyses would bear similar outcomes. As a third type of analysis, we tested two simple linear models to see how categories changed under illuminant changes.
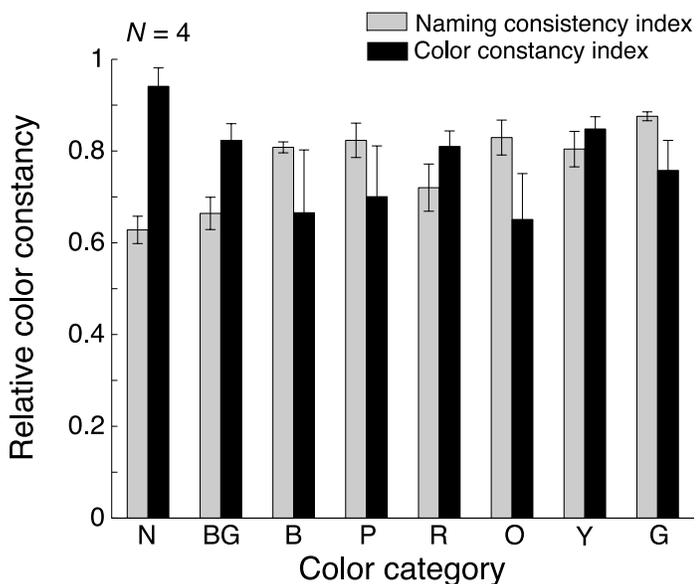

Figure 10. Comparison between the color constancy index and the uncorrected naming consistency index. Black bars plot color constancy calculated from the gray category centroids (Equation 1) and from the chromatic category boundaries (Equation 2). Gray bars plot naming consistency across illuminants. Color categories are indicated on the x-axis as follows: N: neutral; BG: turquoise; B: blue; P: purple; R: red; O: orange; Y: yellow; G: green.

Naming consistency

Observers named stimuli very consistently across five moderately different illuminants. Naming consistency was especially high for hues around green, blue and purple, and relatively low for reddish hues. This sort of inhomogeneity
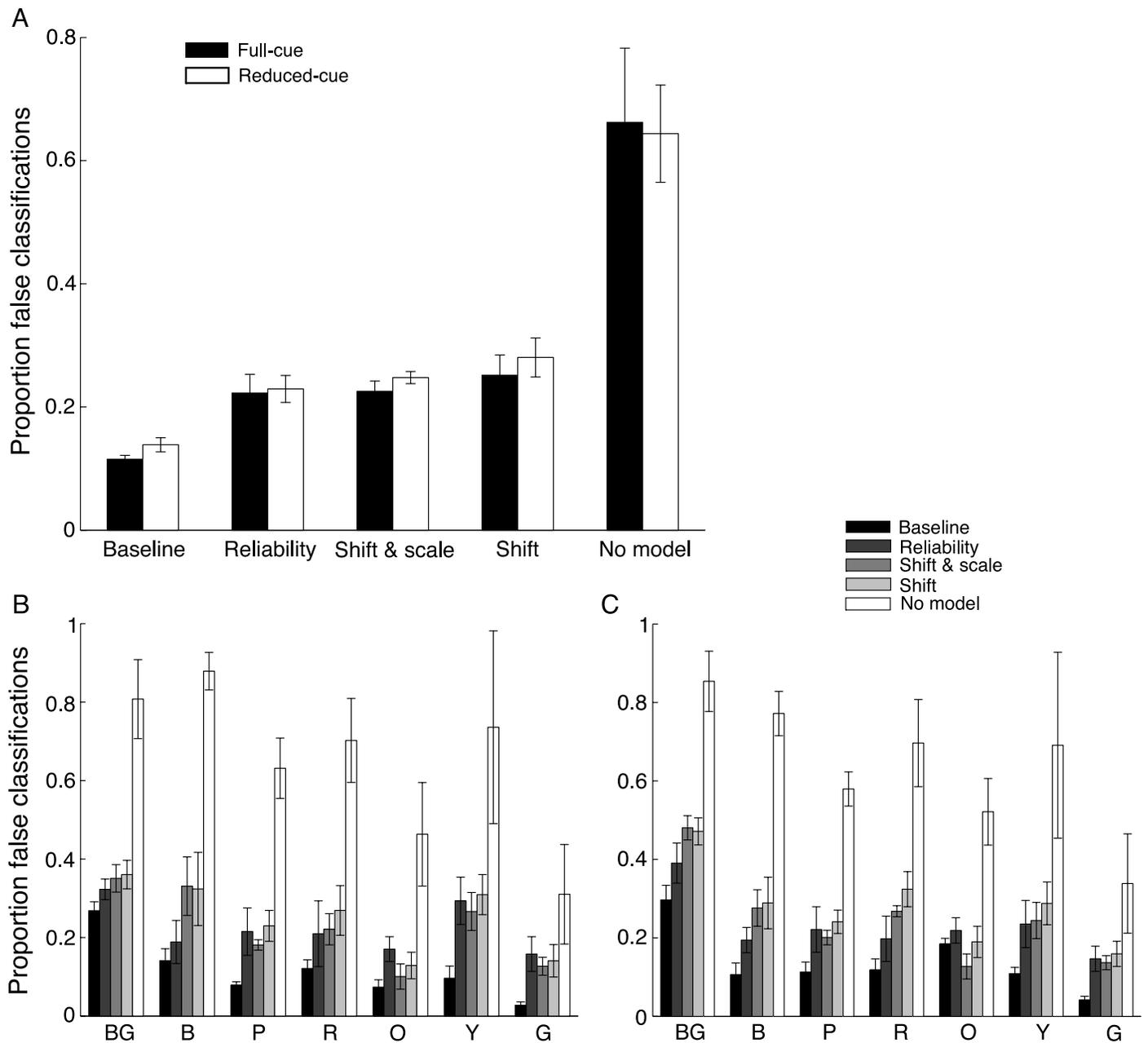
Figure 11. Boundary fit errors. A: Errors pooled over categories for the full-cue (black bars) and reduced-cue (white bars) viewing conditions for value level 6. B: Errors shown by category for the full-cue conditions. C: Errors shown by category for the reduced-cue conditions. Error bars show the standard errors of the means over illuminants (N = 5). Baseline: average errors of the boundary fits in each illuminant condition separately. Reliability: average errors from fitting boundaries over the two repetitions in each illuminant condition. Shift & scale: fitting baseline boundaries to the test conditions with two shift and two scale parameters. Shift: fitting baseline boundaries to the test conditions with two shift parameters. No model: superimposing baseline boundaries on the test data without transformations. Color categories are indicated on the x-axis as follows: BG: turquoise; B: blue; P: purple; R: red; O: orange; Y: yellow; G: green.

across color space might be explained by the fact that the employed illuminant changes were not large enough to push all stimuli out of their category, causing some baseline consistency even in the absence of color constancy. This sort of baseline consistency would be expected to be more pronounced for large color categories, such as green, and would be confounded with color constancy if not accounted for. We simulated this effect and used the

simulated index to derive a corrected naming consistency index. As expected, some stimuli were rather consistent even without color constancy; the average baseline consistency was around .5 compared to the average measured index of .8. The correction accounted for some of the peaks in naming consistency, such as for the green region, but for some other regions, such as for the orange and blue, consistency remained high even after the correction.

In other words, consistency for these regions was not only determined by category size or location in color space.

Observers were generally as consistent when categorizing stimulus colors over time as categorizing them across illuminants, and the pattern of inconsistencies across color space was similar in both cases. This implies, in effect, very high color constancy; as observers were not perfectly consistent in their naming performance over time, some inconsistency in naming across illuminants would be expected. We were also interested in comparing naming consistency across illuminants to naming consistency across different observers. For some hues, such as blue and orange that observers named most consistently across illuminants, inter-observer agreement was also high. Overall, the pattern of consistency as a function of hue was similar for the two indices except for hues around yellow and red, where illuminant consistency was much higher than observer consistency. This implies a possible relationship between categorical color constancy and inter-observer agreement on color categories. A similar link between stability of surface colors and color categories has been suggested on computational grounds by Philipona and O'Regan ([2006]). Philipona and O'Regan showed that some surfaces have more "singular" reflection properties than others, meaning that these surfaces cause more stable cone activity under illuminant changes, and further, that these singularities can predict the loci of focal hues. However, their theoretically derived singularity index did not correlate ($r = -0.04$, n.s.) to our empirical measure of the stability of chromaticities. Apparently, the regions of most stability as measured with color naming are at least partly different from the regions in color space that cause the most stable responses in the cone photoreceptors.

Different color constancy metrics

We found rather similar magnitudes of color constancy with the naming consistency index that does not involve assumptions about color space metrics, and the color constancy indices that intimately depend on the color space in which they are calculated. There were some differences, however. For the gray category, the color constancy index approached 1, compared to the naming consistency index of .65. For the chromatic categories, the two indices came closer at a level around .75. It seems that if color constancy is only measured from the achromatic point and quantified with the traditional color constancy index, constancy for the rest of color space might be overestimated. On the other hand, the naming consistency index when calculated only for the gray category would underestimate overall constancy. Speigle and Brainard ([1999]) were able to predict the amount of constancy in chromatic settings from achromatic settings when measuring both tasks under simultaneous asymmetric viewing. In this viewing condition, where fixation was constantly altered between two fields, constancy was

around .70 for both achromatic settings and asymmetric matching. For a traditional achromatic setting task where only one stimulus had to be fixated, constancy was much higher (.94). This latter condition is comparable to our full-cue viewing condition, and consequently is in agreement with the high constancy index we found based on the gray namings.

One large difference between the traditional asymmetric matching task and the achromatic setting task is that in the former, full adaptation to any of the two illuminants is hindered due to constant changes in fixation, in contrast to the latter task where only one context is viewed. Our results together with Speigle's and Brainard's findings suggest that when measuring constancy under full adaptation to the illuminant, constancy for the achromatic point might exceed constancy for the rest of color space.

Boundary transformations

One important motivation for this study was to find out what kind of transformations the perceptual color space undergoes under illuminant changes. The shift model, where the naming data was just moved around rigidly in the isoluminant plane, worked nearly as well as the shift and scale model where the data was both moved around and scaled on the x and y axes. This indicates that any transformations were accounted for mainly by a rigid shift of the whole color naming space. It is possible that our shift and scale model did not capture all changes in the data, but the fact that fitting errors were roughly of the same magnitude as test–retest variability indicates that the fits were rather good. The boundary fits were overall a bit worse in the reduced-cue viewing conditions, indicating that performance was noisier in this viewing condition compared to the full-cue condition. However, fitting errors were again of the same magnitude as test–retest variability, indicating a good fit to the data.

Relationship to previous research

Two previous studies used a similar paradigm and stimuli to the present experiment. Troost and de Weert ([1991]) measured categorical color constancy with simulated patches embedded in a neutral or a chromatic background. Troost and de Weert also found higher naming consistency with the background that had the illuminant chromaticity, i.e. in their full-cue condition, but the consistency indices that they report are overall lower than in the present study. On average, 38% of the stimuli in Troost and de Weert's study were classified in the same category under all illuminants, compared to 50% in the present study. For the reduced-cue condition, Troost and de Weert reported 10% compared to the present 27% consistency. Moreover, their color constancy indices were systematically lower than what we found: .65 compared to

.97 under full-field illumination. A few differences in the methods might explain the higher color constancy we observed. We used surface reflectance and illuminant spectra to simulate stimuli under illuminant changes, whereas Troost and de Weert used a type of von Kries transform to derive stimulus chromaticity coordinates under each test illuminant. This method might underestimate the complexity in the transformations in stimulus chromaticities under illuminant changes (cf. Smithson, 2005). Moreover, we conducted the experiments in a viewing chamber where observers were immersed in the illumination, and even under reduced viewing conditions, the wall in the periphery was illuminated. Recent studies show that the size of the field of view is important for color constancy (Hansen et al., 2007; Murray et al., 2006).

Recently, Hansen et al. (2007) used color naming to investigate the effect of spatial and temporal context on color constancy. Hansen et al. had observers name stimuli from the DKL isoluminant plane in various viewing conditions with differing amounts of cues to the illumination. The full-cue and reduced-cue viewing conditions of the present study are similar to their conditions 1 and 6, where they observed color constancy indices of .99 and .50, respectively, compared to .97 and .84 in the present study. An important difference between their condition 6 and our reduced-cue condition is that because they did not simulate their surfaces under the test illuminants, the mean chromaticity of the stimulus collection was not biased to the illuminant. In this case the only cues to the illuminant were in the peripheral visual field. In the present study, however, observers were able to use the mean chromaticity of the stimulus collection as an additional cue, which might explain the difference in the degree of color constancy between the two studies.

Conclusions

The accuracy with which single observers name colors across illuminants in full-cue conditions seems to be only limited by their accuracy of naming the same colors over time, pointing to nearly perfect categorical color constancy. The strong correlation between naming consistency across illuminants and naming consistency across observers after correcting for category size effects suggests a relationship between color constancy and consistent communication about color.

Acknowledgments

References

Arend, L. E., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A, Optics and Image Science, 3,* 1743–1751. [PubMed]

Bäuml, K.-H. (1994). Color appearance: Effects of illuminant change under different surface collections. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 11,* 531–542. [PubMed]

Bäuml, K.-H. (1999). Simultaneous color constancy: How surface color perception varies with the illuminant. *Vision Research, 39,* 1531–1550. [PubMed]

Bäuml, K.-H. (2001). Increments and decrements in color constancy. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 18,* 2419–2429. [PubMed]

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. [PubMed]

Brainard, D. H. (1998). Color constancy in the nearly natural image: II. Achromatic loci. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 17,* 307–325. [PubMed]

Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image: I. Asymmetric matches. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 14,* 2091–2110. [PubMed]

Brainard, D. H., & Wandell, B. A. (1992). Asymmetric color matching: How color appearance depends on the illuminant. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 9,* 1433–1448. [PubMed]

Bramwell, D. I., & Hurlbert, A. C. (1996). Measurements of colour constancy by using a forced-choice matching technique. *Perception, 25,* 229–241. [PubMed]

Chichilnisky, E. J., & Wandell, B. A. (1999). Trichromatic opponent color classification. *Vision Research, 39,* 3444–3458. [PubMed]

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

Ekroll, V., Faul, F., Niederée, R., & Richter, E. (2002). The natural center of chromaticity space is not always achromatic: A new look at color induction. *Proceedings of the National Academy of Sciences of the United States of America, 99,* 13352–13356. [PubMed] [Article]

Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision, 7*(4):2, 1–15, http://journalofvision.org/7/4/2/, doi:10.1167/7.4.2. [PubMed] [Article]

Helson, H., & Michels, W. C. (1948). The effect of chromatic adaptation on achromaticity. *Journal of the Optical Society of America, 38,* 1025–1032. [PubMed]

Jacobs, G. H., & Gaylord, H. A. (1967). Effects of chromatic adaptation on color naming. *Vision Research, 7,* 645–653. [PubMed]

Jameson, D., & Hurvich, L. M. (1989). Essay concerning color constancy. *Annual Review of Psychology, 40,* 1–22. [PubMed]

Jin, E. W., & Shevell, S. K. (1996). Color memory and color constancy. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 13,* 1981–1991. [PubMed]

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,* 9085–9089. [PubMed] [Article]

Khang, B. G., & Zaidi, Q. (2002). Cues and strategies for color constancy: Perceptual scission, image junctions and transformational color matching. *Vision Research, 42,* 211–226. [PubMed]

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 307–312. [PubMed] [Article]

Krauskopf, J., Williams, D. R., & Heeley, D. W. (1982). Cardinal directions of color space. *Vision Research, 22,* 1123–1131. [PubMed]

Ling, Y., Allen-Clarke, L., Vurro, M., & Hurlbert, A. C. (2008). The effect of object familiarity and changing illumination on colour categorization. *Perception, 37,* ECVP Abstract Supplement, 149.

Linhares, J. M., Pinto, P. D., & Nascimento, S. M. (2008). The number of discernible colors in natural scenes. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 25,* 2918–2924. [PubMed]

MacLeod, D. I. A., & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America, 69,* 1183–1186. [PubMed]

Munsell, A. H. (1912). A pigment color system and notation. *American Journal of Psychology, 23,* 236–244.

Murray, I. J., Daugirdiene, A., Vaitkevicius, H., Kulikowski J. J., & Stanikunas, R. (2006). Almost complete colour constancy achieved with full-field adaptation. *Vision Research, 46,* 3067–3078. [PubMed]

Nickerson, D., & Newhall, S. M. (1943). A psychological color solid. *Journal of the Optical Society of America, 33,* 419–422.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. [PubMed]

Philipona, D. L., & O'Regan, J. K. (2006). Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Visual Neuroscience, 23,* 331–339. [PubMed]

Pointer, M. R., & Attridge, G. (1998). The number of discernible colors. *Color Research and Application, 23,* 52–54.

Rinner, O., & Gegenfutner, K. R. (2000). Time course of chromatic adaptation for color appearance and discrimination. *Vision Research, 40,* 1813–1826. [PubMed]

Smith, V. C., & Pokorny, J. (1975). Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Research, 15,* 161–171. [PubMed]

Smithson, H. (2005). Sensory, computational and cognitive components of human color constancy. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360,* 1329–1346. [PubMed] [Article]

Smithson, H., & Zaidi, Q. (2004). Colour constancy in context: Roles for local adaptation and levels of reference. *Journal of Vision, 4*(9):3, 693–710, http://journalofvision.org/4/9/3/, doi:10.1167/4.9.3. [PubMed] [Article]

Speigle, J. M., & Brainard, D. H. (1996). Is color constancy task independent? *Proceedings of the 4th IS&T/SID Color Imaging Conference* (167–172). Scottsdale, AZ: Society for Imaging Science and Technology.

Speigle, J. M., & Brainard, D. H. (1999). Predicting color from gray: The relationship between achromatic adjustment and asymmetric matching. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 16,* 2370–2376. [PubMed]

Troost, J. M., & de Weert, C. M. (1991). Naming versus matching in color constancy. *Perception & Psychophysics, 50,* 591–602. [PubMed]

Uchikawa, H., Uchikawa, K., & Boynton, R. M. (1989a). Influence of achromatic surrounds on categorical perception of surface colors. *Vision Research, 29,* 881–890. [PubMed]

Uchikawa, K., Emori, Y., Toyooka, T., & Yokoi, K. (2002). Color constancy in categorical color appearance [Abstract]. *Journal of Vision, 2*(7):548, 548a, http://journalofvision.org/2/7/548/, doi:10.1167/2.7.548.

Uchikawa, K., Uchikawa, H., & Boynton, R. M. (1989b). Partial color constancy of isolated surface colors examined by a color-naming method. *Perception, 18,* 83–91. [PubMed]

Uchikawa, K., Yokoi, K., & Yamauchi, Y. (2004). Categorical color constancy is more tolerant than apparent color constancy [Abstract]. *Journal of Vision, 4*(8):327, 327a, http://journalofvision.org/4/8/327/, doi:10.1167/4.8.327.

Werner, J. S., & Walraven, J. (1982). Effect of chromatic adaptation on the achromatic locus: The role of contrast, luminance and background color. *Vision Research, 22,* 929–943. [PubMed]

Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vision Research, 41,* 2581–2600. [PubMed]

Zaidi, Q., & Bostic, M. (2008). Color strategies for object identification. *Vision Research, 48,* 2673–2681. [PubMed]