Illumination discrimination for chromatically biased illuminations: Implications for color constancy

Stacey Aston
Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK
Current address: Department of Psychology, Durham University, Durham, UK

Ana Radonjić
Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

David H. Brainard
Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

Anya C. Hurlbert
Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK
Current address: Department of Psychology, Durham University, Durham, UK

We measured discrimination thresholds for illumination changes along different chromatic directions starting from chromatically biased reference illuminations. Participants viewed a Mondrian-papered scene illuminated by LED lamps. The scene was first illuminated by a reference illumination, followed by two comparisons. One comparison matched the reference (the target); the other (the test) varied from the reference, nominally either bluer, yellower, redder, or greener. The participant's task was to correctly select the target. A staircase procedure found thresholds for discrimination of an illumination change along each axis of chromatic change. Nine participants completed the task for five different reference illumination conditions (neutral, blue, yellow, red, and green). We find that relative discrimination thresholds for different chromatic directions of illumination change vary with the reference illumination. For the neutral reference, there is a trend for thresholds to be highest in the bluer illumination-change direction, replicating our previous reports of a "blue bias" for neutral reference illuminations. For the four chromatic references (blue, yellow, red, and green), the change in illumination toward the neutral reference is less well discriminated than changes in the other directions: a "neutral bias." The results have implications for color constancy: In considering the stability of surface appearance under changes in illumination, both the starting chromaticity of the illumination and direction of change must be considered, as well as the chromatic characteristics of the surface reflectance ensemble. They also suggest it will be worthwhile to explore whether and how the human visual system has internalized the statistics of natural illumination changes.

Introduction

Color constancy is the perceptual phenomenon by which object colors remain relatively stable despite spatial and temporal changes in the illumination spectrum, which alter the spectrum of light reflected from object surfaces to the eye. However, laboratory measurements of the magnitude of color constancy vary with the experimental conditions and indicate that it is rarely perfect (for reviews, see Brainard & Radonjić, 2014; Foster, 2011; Hurlbert, 1998; Maloney, 1999; Smithson, 2005). A variety of experimental methods have been developed to measure color constancy. These generally involve an assessment of the color appearance of individual surfaces across a change in illumination—for example, by asymmetric surface color matching (e.g., Arend & Reeves, 1986; Brainard, Brunt, & Speigle, 1997; Burnham, Evans, & Newhall, 1957) or achromatic adjustment (e.g., Brainard, 1998; Helson & Michels, 1948). Other paradigms do not directly assess surface color appearance but instead measure constancy through—for example, categorical color naming (e.g., Olkkonen, Witzel, Hansen, &

Citation: Aston, S., Radonjić, A., Brainard, D. H., & Hurlbert, A. C. (2019). Illumination discrimination for chromatically biased illuminations: Implications for color constancy. *Journal of Vision*, *19*(3):15, 1–23, https://doi.org/10.1167/19.3.15.

https://doi.org/10.1167/19.3.15 Received June 9, 2018; published March 29, 2019 ISSN 1534-7362 Copyright 2019 The Authors

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Gegenfurtner, 2010; Troost & de Weert, 1991), object selection tasks (Radonjić, Cottaris, & Brainard, 2015, 2016), or classification of the physical origin of image changes (e.g., material vs. illumination change; Craven & Foster, 1992). Results from such studies demonstrate that the overall degree of color constancy may depend on the spectral reflectance of the surface under view (Burnham et al., 1957; Helson & Michels, 1948; Ling & Hurlbert, 2008), the spectral character of the change in illumination (Brainard & Wandell, 1992; Daugirdiene, Kulikowski, Murray, & Kelly, 2016; Delahunt & Brainard, 2004; Worthey, 1985), the ensemble of surfaces in the scene (Bäuml, 1994, 1995), and manipulation of cues in the image that might mediate constancy (Kraft & Brainard, 1999; Yang & Maloney, 2001).

In this study, we approach color constancy by probing the discriminability of global illumination changes on a scene with unchanging surface reflectances. The illumination discrimination task (IDT) that we employ was introduced by Pearce, Crichton, Mackiewicz, Finlayson, and Hurlbert (2014). A number of other recent studies also employ similar methods (Alvaro, Linhares, Moreira, Lillo, & Nascimento, 2017; Radonjić, Pearce et al., 2016; Radonjić et al., 2018; Weiss, Witzel, & Gegenfurtner, 2017; see also Lucassen, Gevers, Giksenij, & Dekker, 2013). On each trial of the IDT, the participant views three successively presented scenes: a reference scene and then two comparisons. The reference scene is illuminated by a reference illumination. One of the two following comparison scenes is illuminated by the same reference illumination, while the other comparison scene is illuminated by a test illumination whose chromaticity can vary along a specified chromatic direction. The order of the comparison scenes is chosen randomly on each trial. The participant's task is to signal which comparison scene illumination is most similar to the reference illumination. Discrimination thresholds along different directions are measured by varying the size of the illumination chromaticity change along specified chromatic directions and finding the size of the illumination change that leads to a criterion percent correct.

The potential connections between illumination discrimination thresholds and color constancy are twofold. First, if we assume that the participant is able to detect a change in illumination only if that change evokes a discriminable change in the appearance of scene surfaces, then illumination discrimination thresholds measure the chromatic changes in illumination that can occur while still preserving surface color appearance; that is, the illumination thresholds indicate illumination changes across which there is essentially perfect color constancy. We note, however, that it is theoretically possible that the participant could perceive a change in the appearance of scene surfaces but no change in illumination. We think that the design of the experiment here—in which the participant is aware that the surfaces are real and unchanging—makes this possibility unlikely. It is also theoretically possible that the participant may perceive a change in illumination without perceiving a change in the appearance of scene surfaces. In this case, the illumination discrimination thresholds represent a lower bound on illumination changes across which essentially perfect color constancy holds. The relation between illumination discrimination and surface discrimination thresholds remains to be tested explicitly, as does the relation between illumination discrimination thresholds and color constancy for suprathreshold illumination changes.

The second potential connection between the IDT and color constancy relates more generally to the ability of the visual system to encode and represent the scene illumination (Logvinenko & Maloney, 2006; Logvinenko & Menshikova, 1994; Rutherford & Brainard, 2002; Zaidi, 1998). More specifically, the accuracy with which the visual system represents the illumination may directly affect the accuracy with which it represents surface reflectance, given that the two are entangled in the proximal image. Illumination discrimination may therefore provide information about the constancy of object surface color under particular directions of illumination change. In this context, the reported tendency of illumination discrimination thresholds to be highest for illumination changes in a bluer chromatic direction might be interpreted as the visual system's internalization of the statistics of natural daylight (Hernández-Andrés, Romero, Nieves, & Lee, 2001; Spitschan, Aguirre, Brainard, & Sweeney, 2016).

Using the general illumination discrimination paradigm described above, previous studies have found that illumination discrimination depends on the illumination-change direction (Pearce et al., 2014), the ensemble of surface reflectances in the scene (Radonjić, Pearce et al., 2016), whether the surfaces in the scene remain fixed across the illumination change (Radonjić et al., 2018), and on whether an observer has normal, anomalous, or dichromatic color vision (Alvaro et al., 2017; Aston, Le Couteur Bisson, Jordan, & Hurlbert, 2016).

A consistent finding that emerges across the majority of these studies is that in scenes whose average surface reflectances are neutral, thresholds for "bluer" directions of change are higher than for other directions when the results are expressed in the CIELUV color space. We refer to this feature as "the blue bias" for illumination changes relative to a neutral reference chromaticity (Pearce et al., 2014; Radonjić, Pearce et al., 2016; Radonjić et al., 2018; Weiss et al., 2017). This bias is intriguing. Since CIELUV was designed to be

approximately perceptually uniform for surface color discrimination, the bias is not readily explained in terms of differences in discrimination for different color directions. Previously, the blue bias has been interpreted as an optimization of color constancy mechanisms for natural illumination changes, an idea we will expand on later in the text.

In addition, Radonjić, Pearce et al. (2016) showed that relative illumination discrimination thresholds across different chromatic directions vary with the average image chromaticity. More specifically, they showed that varying the chromatic bias of the surfaces in the scene modulates relative illumination discrimination thresholds. In general, a chromatic bias in the average image chromaticity may arise from a chromatic bias in the surface reflectances (e.g., green foliage) or in the illumination (e.g., reddish sunset) and distinguishing between these solely based on the irradiance signal from any one scene location is theoretically impossible without prior constraints or assumptions. Indeed, this statement summarizes the computational problem at the core of color constancy (Hurlbert, 1998; Maloney, 1999). For example, "gray world" constancy algorithms assume that scenes tend to have distributions of surface reflectances that average to neutral. This assumption is deliberately violated by the stimuli in Radonjić, Pearce et al. (2016) in which the average surface reflectances in the scene are chromatically biased while the reference illumination remains neutral.

The question we address here is whether relative illumination-discrimination thresholds also vary when the gray world assumption is upheld in the stimulus scenes, and a chromatic bias is instead introduced by a shift in the chromaticity of the reference illumination. Our specific aim is to determine whether and how thresholds for discriminating illumination changes in different chromatic directions are influenced by systematic shifts in the chromaticity of the reference illumination. The data also enable us to address other key questions: (a) does the previously observed blue bias in discrimination thresholds occur for nonneutral reference chromaticities, and (b) how are illumination discrimination thresholds linked to fundamental mechanisms of chromatic discrimination?

Methods

Overview

The experiment consisted of a series of trials. On each trial, participants first viewed a Mondrian-papered scene illuminated by a reference illumination (Figure 1). Participants then saw the same scene under two successively presented comparison illuminations. One of the comparison illuminations was always the same as the reference (the target), while the other (the test) varied along one of four illumination-change directions (referred to as the bluer, yellower, redder, and greener directions). The participant's task on each trial was to indicate which of the two comparison illuminations most closely matched the reference illumination. A staircase procedure governed selection of the test illumination on each trial. Twelve staircase procedures (three for each illumination-change direction) were interleaved to determine illumination discrimination thresholds. Thresholds were measured along each illumination-change direction for five different reference illumination conditions. Each participant completed the full 5×4 (Reference Illumination $\times$ Illumination-Change Direction) design. Each reference illumination condition was completed in a separate session.

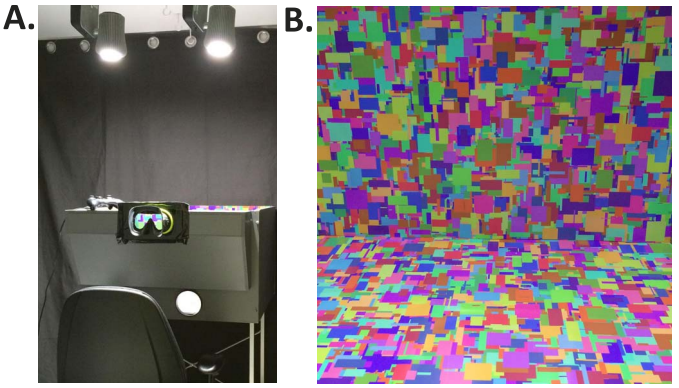

Figure 1. The experimental box. (A) Participants positioned their head against the goggles, restricting their view to inside the stimulus box. The box was illuminated by two spectrally tuneable LED luminaires, which also provided the general room illumination. (B) The participant's view of the inside of the box (shown here illuminated by an arbitrary illumination that was not used during the experiment).

Ethics

Ethical approval for the study was received from the Newcastle University Ethics Board. Written consent was received from all participants prior to participation in the study.

Participants

Nine participants were recruited (four male, five female, mean age of 23 $\pm$ 3 years). All participants had normal or corrected-to-normal visual acuity (self-reported) and no color vision deficiencies, assessed using Ishihara Color Plates. They received cash compensation for their time.

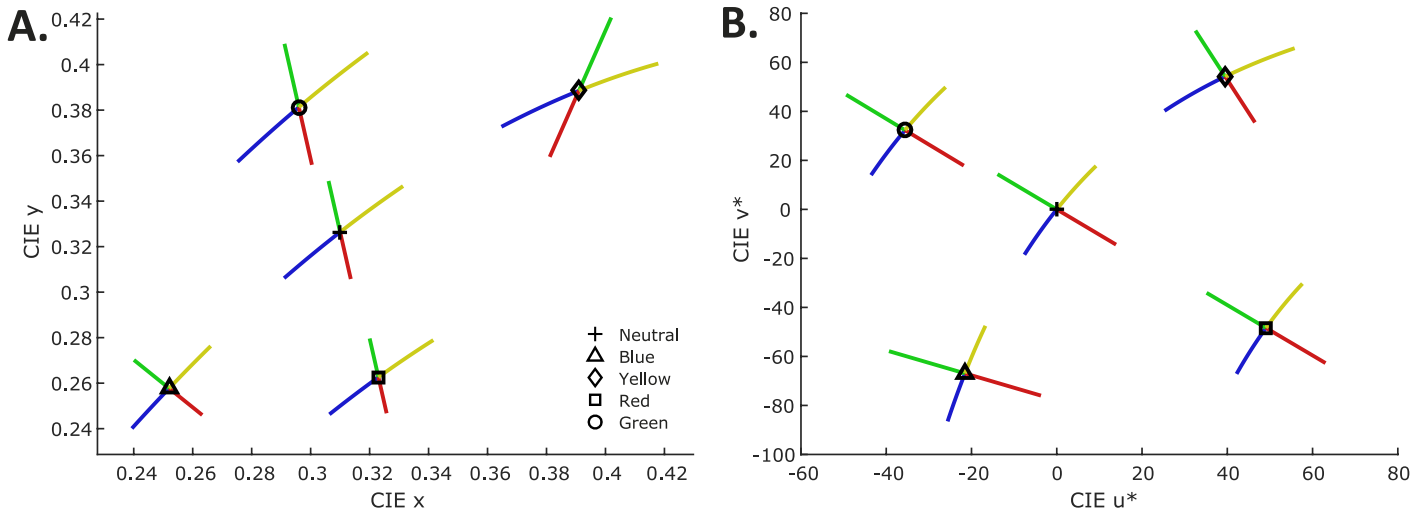

Figure 2. The illumination chromaticities. (A) Illumination chromaticities plotted in the CIE *xy* chromaticity plane. (B) Illumination chromaticities plotted in the CIELUV chromaticity plane. The black open symbol marks the chromaticity of the reference illumination in each of the five conditions. The different axes of illumination-change are shown in the color that corresponds to the respective direction of change.

The scene

The scene was viewed through a porthole in a box of dimensions: height = 45 cm, width = 77.5 cm, depth = 64.5 cm. Goggles mounted at the porthole stabilized the participant's head position. The front of the box extended out toward the participant so that the viewing distance from the participant's eyes to the back of the box was 81cm (Figure 1A). Both the scene and the participant were immersed in the illumination (the only source of light in the room). The participant's view was restricted to the inside of the box (Figure 1B). The top of the box was open to allow illumination of the scene. The bottom, rear, and sides of the box were papered with a matte-printed Mondrian. This Mondrian was designed and printed specifically for the experiment and was not used in any of the previous studies. Each patch of the Mondrian was one of 24 unique surface reflectances (Figure A1; Table A1 in Appendix A), chosen such that their average chromaticity under a hypothetical equal energy light source was approximately CIE $(x, y) = (0.33, 0.33)$. This resulted in an average reflectance that was approximately nonselective (equal at all wavelengths; solid black line in Figure A1B in Appendix A). The height and width of each patch varied in size from 2 to 42 mm (0.14 to 2.97 degrees of visual angle). The Mondrian was designed in MATLAB (MathWorks, Natick, MA) by generating 100,000 patches of varying pixel size and chromaticity and randomly placing them inside a 1001 × 1001 pixel frame that was initially set to (R, G, B) = (0, 0, 0) everywhere. After the Mondrian generation process was complete, we checked for RGB values equal to (0, 0, 0) to ensure that no black patches remained. Once the Mondrian was printed on paper, measurements of the surface spectral reflectance of each uniquely colored patch were taken, as described in Appendix A. Surface reflectance functions for each patch are available in the online supplement (see below).

The illuminations

The illuminations used in the experiment can be split into five sets corresponding to the five reference illuminations: a neutral, blue, yellow, red, and green set. In each set, the illuminations were generated such that they varied systematically away from the chromaticity of that set's reference illumination. The CIE chromaticities of the five reference illuminations were: $(x, y) = (0.31, 0.33$, neutral; D65); $(x, y) = (0.25, 0.26$, blue); $(x, y) = (0.39, 0.39$, yellow); $(x, y) = (0.32, 0.26$, red); $(x, y) = (0.30, 0.38$, green). For each reference illumination, 80 test illuminations were generated that varied away from the reference in four distinct chromatic directions: bluer, yellower, redder, and greener (20 test illuminations per direction; Figure 2).

For the neutral, blue, and yellow reference illuminations, bluer/yellower test illuminations were parameterized to fall along the Planckian locus, in order to mimic the chromaticities of daylight illuminations (defined in the CIE *xy* chromaticity plane as $y = 2.870x - 3x^2 - 0.275$; Wyszecki & Stiles, 1967). For the red and green reference illuminations, the bluer/yellower test illuminations varied along a linear translation of the Planckian locus in the CIELUV chromaticity plane.

For the neutral, blue, and yellow reference illuminations, redder/greener test illuminations were param-

eterized to fall along the constant correlated color temperature line (CCT line) through the corresponding reference illumination chromaticity. Finally, for the red and green reference illumination conditions, the redder/greener test illuminations fell along a translation of the constant CCT line at $x = 0.31$ (D65).

The members of each set of 20 test illuminations were spaced approximately one $\Delta E_{u^*v^*}$ apart from each other, with the conversions to CIELUV coordinates made using the tristimulus values of the neutral reference illumination as the fixed white point for all sets (Figure 2B). All illuminations were generated such that the luminance of a white polymer calibration tile placed flush against the back wall of the stimulus box, orthogonal to the viewing direction, was $Y = 50$ cd/m^2.

The illuminations were produced using two 10-channel (nine unique) spectrally tuneable LED lamps (HI-LED Prototype I luminaires; produced by the Catalonia Institute for Energy Research, Barcelona, Spain, as prototypes for the EU FP7-funded HI-LED project; www.hi-led.eu). The spectral power distribution of the light emitted from the lamps can be controlled in real-time by varying the pulse width of the input to each individual LED channel, allowing the overall spectral power distribution of the output mixture to be specified by a set of 10 weights.

To find a set of weights that produce a spectral power distribution with specified luminance and CIE *xy* chromaticity (as measured from the white calibration tile) we used custom MATLAB scripts to find the smoothest spectrum within the device gamut that had the desired chromaticity. The smoothness constraint improves the degree to which the spectra approximate those of natural daylights. Full details of the spectral fitting procedure are reported elsewhere (Finlayson, Mackiewicz, Hurlbert, Pearce, & Crichton, 2014; Pearce et al., 2014).

Spectral calibration

To calibrate the apparatus, we measured the spectral power distribution of each individual LED channel at maximum power. To make the measurements, a polymer white reflectance tile was placed flush against the back wall of the stimulus box with the Mondrian lining removed. A CS2000 Konica Minolta spectroradiometer (Konica Minolta, Nieuwegein, Netherlands) was used to take radiance measurements from the tile when it was illuminated by each LED channel. The measurements for each channel are then the basis functions of the illuminations. The basis functions were input to the spectral fitting code to find the weights for each primary that would achieve specified chromaticities and luminance, as described above. The spectral power distributions actually obtained were then measured directly, using the same procedure used to obtain the basis functions. They are provided in the online supplement (see below).

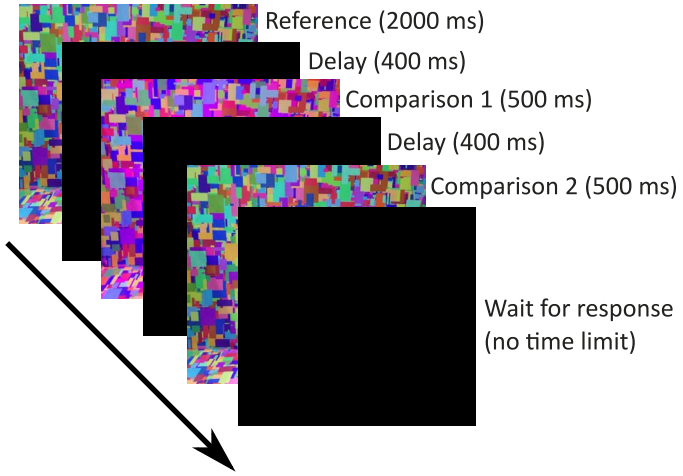

Figure 3. The IDT. On each trial, participants were first presented with the reference illumination (2000 ms), followed by two comparisons (500 ms), separated by a short dark interval (400 ms). One comparison matched the reference (the target) while the other varied from the reference (the test). Participants were instructed to indicate which comparison was most similar to the reference via a button press. Colors shown in this figure are illustrative and do not represent the actual experimental stimuli.

Procedure

The experiment consisted of five sessions, one for each reference illumination condition. At the start of the first session, each participant read the standardized instructions (see the online supplement) and was permitted to ask questions. All participants then received the same verbal instructions: "On each trial, you will see the reference illumination followed by two comparison illuminations. You will use the gaming pad to indicate which of the two comparison illuminations most closely matched the reference." These verbal instructions were repeated in the four subsequent sessions, but the participants did not read the standardized instructions a second time. On the first and all subsequent visits, participants dark adapted for 2 min before starting the task. On each trial, the reference illumination was visible for 2000 ms. Each comparison illumination was displayed for 500 ms and between each illumination there was 400 ms of darkness (Figure 3). The interval in which the comparison illumination was equal to the reference illumination was randomized across trials.

Thresholds for each direction of change were found using a one-up, three-down, transformed and weighted

staircase procedure (step sizes of one, two, or three depending on the preceeding run of trials; Kaernbach, 1991). Staircases started at a random nominal step between 10 and 20 $\Delta E_{u^*v^*}$ away from the reference and terminated after six reversals. Three interleaved staircases were completed for each direction of illumination change. Staircases for the four illumination-change directions were also interleaved, for 12 interleaved staircases per reference illumination. Participants were told that they could take as long as they needed to respond and that they could take a break at any time during the experiment (by remembering their response but not entering it until they were ready to continue). Participants were required to take a mandatory break after every 100 trials. This break could last for as long as the participant desired. Participants did not leave the dark experimental room during breaks.

Data analysis

For each reference illumination, the thresholds for each illumination-change direction were calculated by taking the mean of the last two reversals from each of the three interleaved staircases (a mean over six reversals for each illumination direction). In the calculation we used actual (not nominal) differences between the reference and test illuminations. For this purpose, we created a set of lookup tables[1] where for each nominal test illumination (Steps 1 to 20) we calculated the actual chromaticity difference relative to the reference illumination in CIELUV $\Delta E_{u^*v^*}$. We used the chromaticities of the illuminations as measured during calibration and a luminance of 50 cd/m^2. The *XYZ* tristimulus coordinates of the neutral reference illumination were used as the white point for conversion to CIELUV coordinates. We refer to this set of lookup tables as the *fixed white point lookup tables*. To compute the thresholds reported in the main text, staircase reversal points for each reference/illumination-change direction were converted to CIELUV$\Delta E_{u^*v^*}$ differences from the reference using these lookup tables and then averaged.

We also created two alternate sets of lookup tables that mapped nominal staircase steps (1 to 20) to CIELUV $\Delta E_{u^*v^*}$ values. The first alternative set of lookup tables, referred to hereafter as the *image mean lookup tables*, used the *XYZ* tristimulus values of the mean image chromaticity under each illumination rather than the *XYZ* tristimulus values of the illumination spectra to compute the CIELUV $\Delta E_{u^*v^*}$ differences. The mean image chromaticity (CIE *xy*) and luminance under each illumination were calculated based on the surface spectral reflectance of the Mondrian-papered back wall (obtained using a hyperspectral image as described in Appendix B). This was done by using the surface reflectances to calculate the spectral power distribution of the light reflected from the scene under each illumination, using wavelength-by-wavelength multiplication. For each pixel in the image, the calculated reflected spectrum at each pixel was converted to CIE *Yxy* values and these were averaged across all pixels. The mean *Yxy* values were then converted to *XYZ* for each image. The mean *XYZ* for the neutral reference scene was used as the white point for conversion from mean image *XYZ* values to CIELUV coordinates. Across illuminations, mean image luminance (Y) varied only slightly: mean luminance of 8.51 cd/m^2 with standard deviation of 0.04 cd/m^2. Converting nominal staircase steps to CIELUV $\Delta E_{u^*v^*}$ values using these tables expresses thresholds with respect to the proximal image viewed by the participants rather than with respect to the distal scene variable of illumination change.

For both the fixed white point and image mean lookup tables, the neutral reference illumination (D65) was used as the white point for calculation of CIELUV $\Delta E_{u^*v^*}$ values in all reference illumination conditions. However, it could be argued that if the participant adapts to the temporal average of the illumination during each condition of the experiment, then the reference illumination from each condition is a more appropriate white point choice for conversions. Hence, we constructed another set of lookup tables which we refer to as the *variable white point lookup tables*. These tables were constructed from the illumination spectra but using each reference illumination's *XYZ* values as the white point for conversion to CIELUV coordinates. In Appendix C, we show that our overall conclusions remain the same when we use either the image mean or variable white point lookup tables to calculate thresholds (instead of the fixed white point lookup tables that we used to produce the thresholds reported in the results).

All data are presented in the form of means over participants and associated standard errors. Plotted error bars represent 1 *SE*. If the assumption of sphericity was violated when conducting an analysis of variance (ANOVA), an appropriate correction was planned depending on the value of ϵ. For $\epsilon \leq 0.75$, a Greenhouse-Geisser correction was to be performed, whereas for $\epsilon > 0.75$, we would use a Huynd-Feldt correction. Only Greenhouse-Geisser corrections were needed during the analysis. Where pairwise comparisons and simple main effects are reported, *p* values have been corrected for multiple comparisons by applying a suitable Bonferroni correction. For example, where simple main effects analyses are used to follow up the finding of a significant interaction in a two-way ANOVA with Factors A and B, the *p* values of the one-way ANOVAs for Factor A are multiplied by the number of levels for Factor B and vice-versa. Similarly,

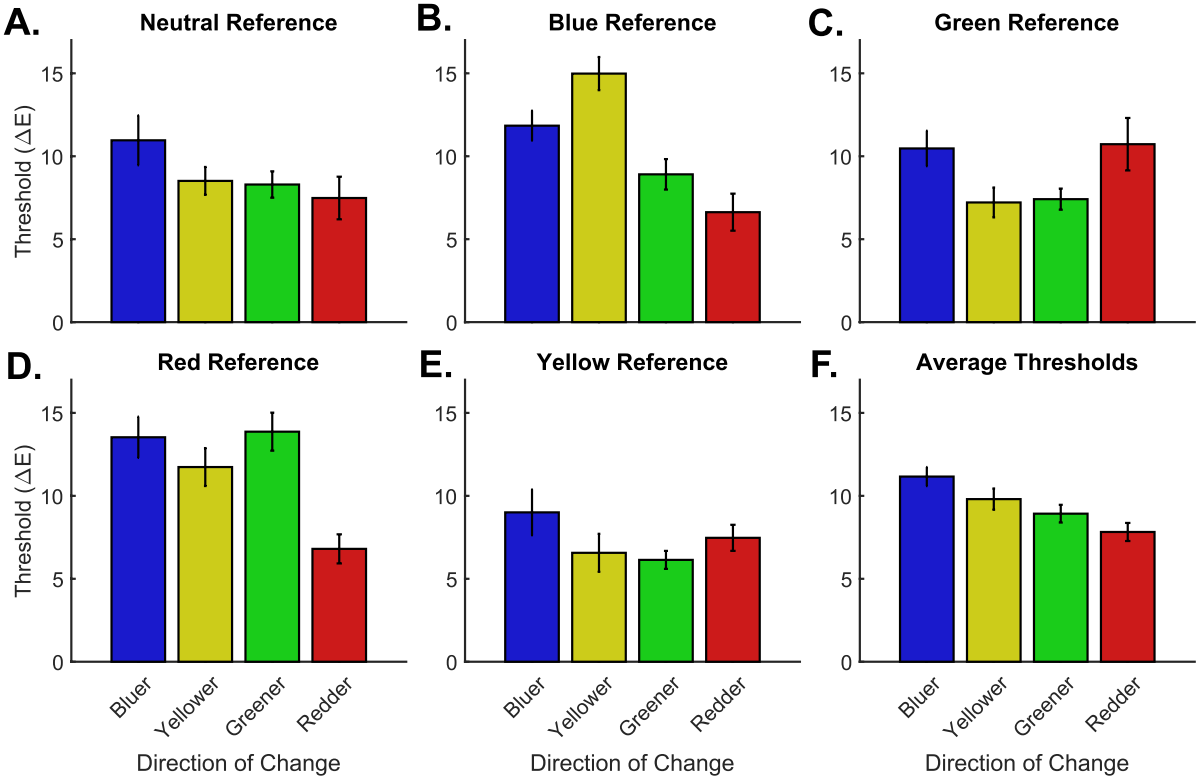

Figure 4. Illumination discrimination thresholds for the different reference illumination conditions. (A–E) Illumination discrimination thresholds for each chromatic direction of change in each of the five reference illumination conditions. (F) Thresholds for each illumination-change direction averaged over reference illumination condition.

when a one-way ANOVA whose factor has m levels is followed up with pairwise comparisons, the p values resulting from these comparisons are multiplied by $m(m - 1)/2$, the number of possible pairwise comparisons.

Online supplement

The online supplement (http://color.psych.upenn.edu/supplements/illuminationdiscriminationCB) includes: (a) three types of lookup tables specifying the difference between the reference and test illuminations in CIELUV $\Delta E_{u^*v^*}$; (b) all experimental illumination spectra; (c) surface reflectance functions for each Mondrian patch; (d) instructions verbatim; and (e) each individual participant's data.

	Illumination-change direction			
Reference	Bluer	Greener	Redder	Yellower
Neutral	10.96 (1.42)	8.30 (0.75)	7.48 (1.21)	8.51 (0.79)
Blue	11.84 (0.86)	8.91 (0.87)	6.63 (1.05)	14.98 (0.94)
Green	10.47 (1.01)	7.41 (0.60)	10.73 (1.49)	7.21 (0.85)
Red	13.53 (1.17)	13.87 (1.08)	6.80 (0.83)	11.74 (1.08)
Yellow	9.00 (1.31)	6.14 (0.51)	7.47 (0.74)	6.57 (1.08)

Table 1. Mean illumination discrimination thresholds for the different reference illumination conditions using the fixed white point lookup tables. *Note*: Values in parentheses show the standard error.

Results

Figure 4A through E and Table 1 summarize the results. The key finding is that the pattern of thresholds across illumination-change directions varies strongly as a function of reference illumination. For the neutral reference illumination, thresholds for the bluer illumination-change direction are highest. This replicates our previous findings of a blue bias for similar viewing conditions (Pearce et al., 2014; Radonjić, Pearce et al., 2016; Radonjić et al., 2018), although here the blue elevation does not rise to statistical significance (but see below). A similar pattern is seen for the yellow reference illumination, but a different one in the other reference illumination conditions. For the blue reference illumination thresholds are highest in the yellower illumination-change direction, while for the green reference illumination thresholds are highest in the redder direction. The finding that relative thresholds vary with the reference illumination chromaticity parallels our previous finding that relative illumination

discrimination thresholds depend strongly on the distribution of surface reflectance spectra in the scene (Radonjić, Pearce et al., 2016).

A 5×4 repeated measures ANOVA confirms what is visible in Figure 4—that there is a significant interaction between reference illumination and chromatic direction of illumination change, $F(12, 96) = 9.11$, $p < 0.001$. Moreover, there is a significant main effect of reference illumination, regardless of the direction of chromatic change, $F(4, 32) = 12.84$, $p < 0.001$, and a significant main effect of chromatic direction of illumination change, regardless of reference illumination condition, $F(3, 24) = 8.35$, $p = 0.001$.

Post hoc analyses also support the observation that the reference illumination affects relative thresholds in the four illumination-change directions (simple main effects analysis; Table 1: across-column comparison within each row). Illumination-change direction has a significant effect on thresholds for the blue, green, and red reference illumination conditions, $F(3, 24) = 15.45$, $p < 0.005$; $F(3, 24) = 5.37$, $p = 0.030$; and $F(3, 24) = 11.48$, $p < 0.005$, respectively, but not for the neutral and yellow reference illumination conditions, $F(3, 24) = 4.32$, $p = 0.070$ and $F(3, 24) = 2.88$, $p = 0.285$, respectively. Note that we have previously reported statistically significant main effects of the direction of illumination change for a neutral reference illumination condition (Pearce et al., 2014; Radonjić, Pearce et al., 2016), with highest thresholds in the bluer direction. Here, the main effect does not cross the threshold of statistical significance, given the Bonferroni correction for the fact that we examined such effects for five reference illumination conditions. Nonetheless, measured thresholds in the bluer direction remain the highest for the neutral reference illumination.

For the three reference illuminations with a significant main effect of illumination change, we proceeded to make post hoc pairwise comparisons between illumination-change directions. For the blue reference, thresholds are highest for the yellower direction, and significantly higher than the thresholds for the redder ($p = 0.004$) and greener directions ($p = 0.041$), but not significantly higher than bluer. Also, thresholds in the bluer direction are significantly higher than those in the redder direction ($p = 0.017$). For the green reference, no pairwise comparisons yield significant differences. For the red reference, thresholds in the redder direction are the lowest and significantly lower than thresholds in all other directions ($p = 0.006$, $p = 0.012$, and $p = 0.008$, for bluer, greener, and yellower, respectively).

To assess whether changes in the pattern of relative thresholds, for the different reference illumination conditions (described above), lead to significant differences in overall thresholds for particular illumination-change directions, we explored how thresholds across different reference illumination conditions differ within each illumination-change direction (simple main effects analysis; Table 1: across-row comparison within each column). For each illumination-change direction, except redder, these differed significantly, $F(4, 32) = 4.38$, $p = 0.024$ for bluer, $F(4, 32) = 15.15$, $p < 0.004$ for greener, and $F(4, 32) = 18.13$, $p < 0.004$ for yellower; redder $F(1.64, 13.11) = 4.32$, $p = 0.168$, with a Greenhouse-Geisser correction. For the bluer illumination-change direction, thresholds were lowest for the yellow and highest for the red reference illumination, but none of the pairwise comparisons were significant. For the greener direction, thresholds for the red reference were significantly higher than all other references except blue (neutral: $p = 0.039$, green: $p = 0.003$, yellow: $p < 0.001$). For the yellower direction, the highest threshold was for the blue reference condition and this was significantly higher than for the neutral, green, and yellow references ($p = 0.006$, $p = 0.002$, $p = 0.005$, respectively). In addition, thresholds for the yellower direction for the red reference were significantly higher than for the green and yellow references ($p = 0.011$, and $p = 0.005$, respectively).

Figure 5 replots the data in the CIELUV chromaticity plane. In this representation, an interesting pattern in the data emerges. For each of the reference illuminations that are shifted away from neutral (i.e.,

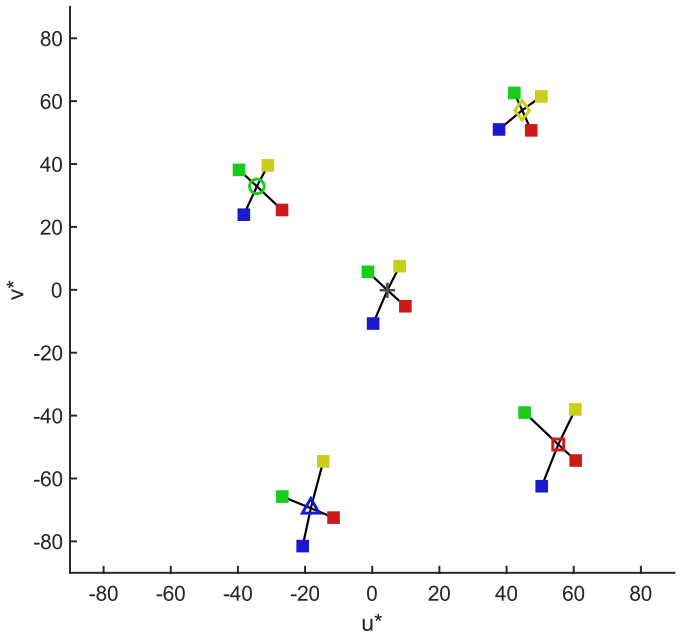

Figure 5. Mean thresholds (across participants) plotted in the CIELUV chromaticity plane. The cross and hollow markers correspond to the different reference illumination conditions (+ neutral, △ blue, ○ green, □ red, ◇ yellow). The solid square symbols represent the threshold locations for the different illumination-change directions (color-coded). The connecting black lines are drawn in to facilitate comparisons of threshold magnitudes between directions.

blue, green, red, and yellow), thresholds are highest in the illumination-change direction that is oriented toward the neutral reference. For example, for the red reference, thresholds are highest for the greener direction, while for the green reference thresholds are highest for the redder direction. Figure 6 shows the individual participants' data in the same format as Figure 5. In the blue reference condition, six out of nine participants' highest threshold is in the yellow illumination-change direction. In the other three biased reference conditions (yellow, green, and red), four out of nine participants have their highest threshold in the direction chromatically opposite to the bias (bluer, redder, and greener, respectively; individual participants' data, and alternative visualizations of the thresholds in Figures 5 and 6 as contours, are also available in the online supplement). Note that the neutral reference condition represents a typical daylight. This finding suggests that discrimination thresholds may be related to the likelihood of natural illumination changes, an explanation that we explore in the Discussion.

Lastly, Figure 4F shows the average thresholds taken across the five reference conditions. The plot shows that the average across the particular five reference conditions we chose closely resembles the results for the neutral reference, with thresholds in the bluer illumination-change direction being highest. Because there is a significant interaction between reference condition and illumination-change direction, though, we must exercise caution in interpreting this pattern. With this caveat in mind, recall from above that there was a main effect of chromatic direction of illumination change, and we proceeded to perform post hoc comparisons of the average illumination discrimination thresholds across illumination-change directions. Those in the bluer and yellower chromatic directions of change are significantly higher than redder thresholds ($p = 0.024$ and $p = 0.008$, respectively). There are no other significant differences between the average thresholds for the different illumination-change directions. However, Table 1 shows that thresholds for the bluer illumination-change direction are the highest or second highest in all reference conditions, suggesting that the blue bias we have documented for neutral reference conditions is reprised to some extent for the four chromatic reference illuminations studied here.

Comparison with Radonjić, Pearce et al. (2016)

The finding that relative thresholds vary with the reference illumination chromaticity parallels our previous finding that relative illumination discrimination thresholds depend strongly on the distribution of surface reflectance spectra in the scene (Radonjić, Pearce et al., 2016). Radonjić, Pearce et al. (2016) introduced chromatic bias in the reference scene by modulating the mean surface reflectance while maintaining a neutral reference illumination chromaticity. In our experiment, the mean surface reflectance is kept constant at neutral, while the reference illumination chromaticities are varied. Under both manipulations, the mean chromaticity of the image of the reference scene is translated to a nonneutral point. We may therefore ask whether the pattern of relative thresholds for the two experiments depends in a similar way on mean image chromaticity.

To answer this question, we compared our results with those of Radonjić, Pearce et al. (2016). For this purpose, we recomputed the thresholds from Radonjić, Pearce et al. (2016, experiment 2) by using an equivalent of the image mean lookup table (see Appendix D for details). In Figure 7 we plot these thresholds together with those computed using the image mean lookup tables from the current experiment. The figure illustrates the qualitative similarity of results from the two studies in terms of magnitude and general pattern. In particular, the results from Radonjić, Pearce et al. (2016) evince the same "neutral bias"—a trend toward higher discrimination thresholds in directions opponent to the chromatic bias. When the mean reference image chromaticity is biased toward reddish-blue, sensitivity to illumination changes is lowest in the greener direction, which for this condition is the measured direction that heads toward a neutral chromaticity. Similarly, when the mean reference image chromaticity is yellowish-green, sensitivity is lowest in the bluer direction, again the measured direction that heads toward a neutral chromaticity (although thresholds in the redder direction are almost as high in this case). The general conclusion from both studies is that relative illumination discrimination thresholds vary with the mean image chromaticity, with a tendency for illumination changes that pull the image toward a neutral mean chromaticity to be less easily discriminated. It is important to keep in mind, however, that this conclusion is based on post hoc comparisons. The two studies were not designed to be closely matched and there are many differences between them, limiting the degree to which we can draw strong conclusions based on comparisons between them.

Discussion

Summary

We measured illumination discrimination in a group of observers for five different reference illumination chromaticities: neutral, as in the standard IDT used by Pearce et al. (2014) and Radonjić, Pearce et al. (2016);

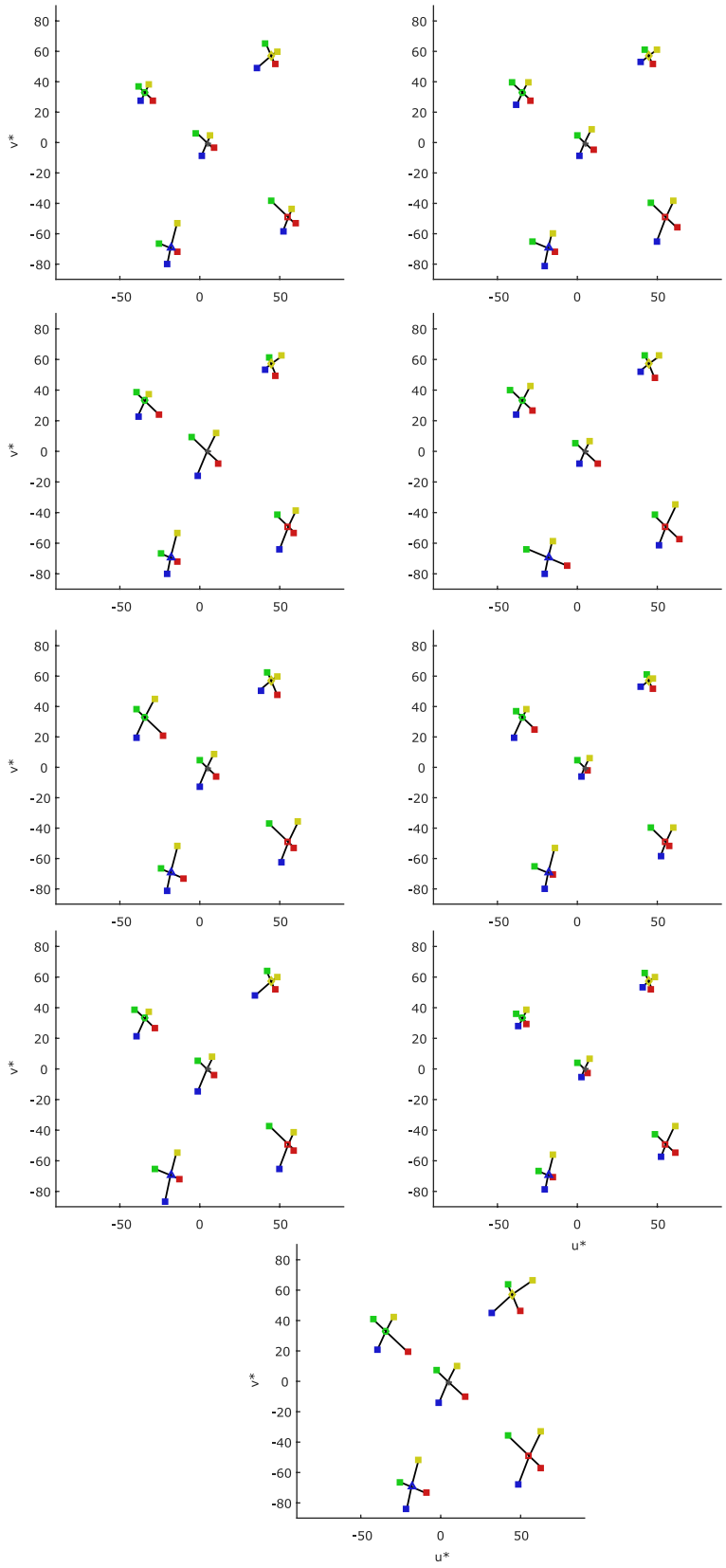

Figure 6. Thresholds for each individual participant (one participant per plot) plotted in the CIELUV chromaticity plane. For interpretation of symbols, see Figure 5.

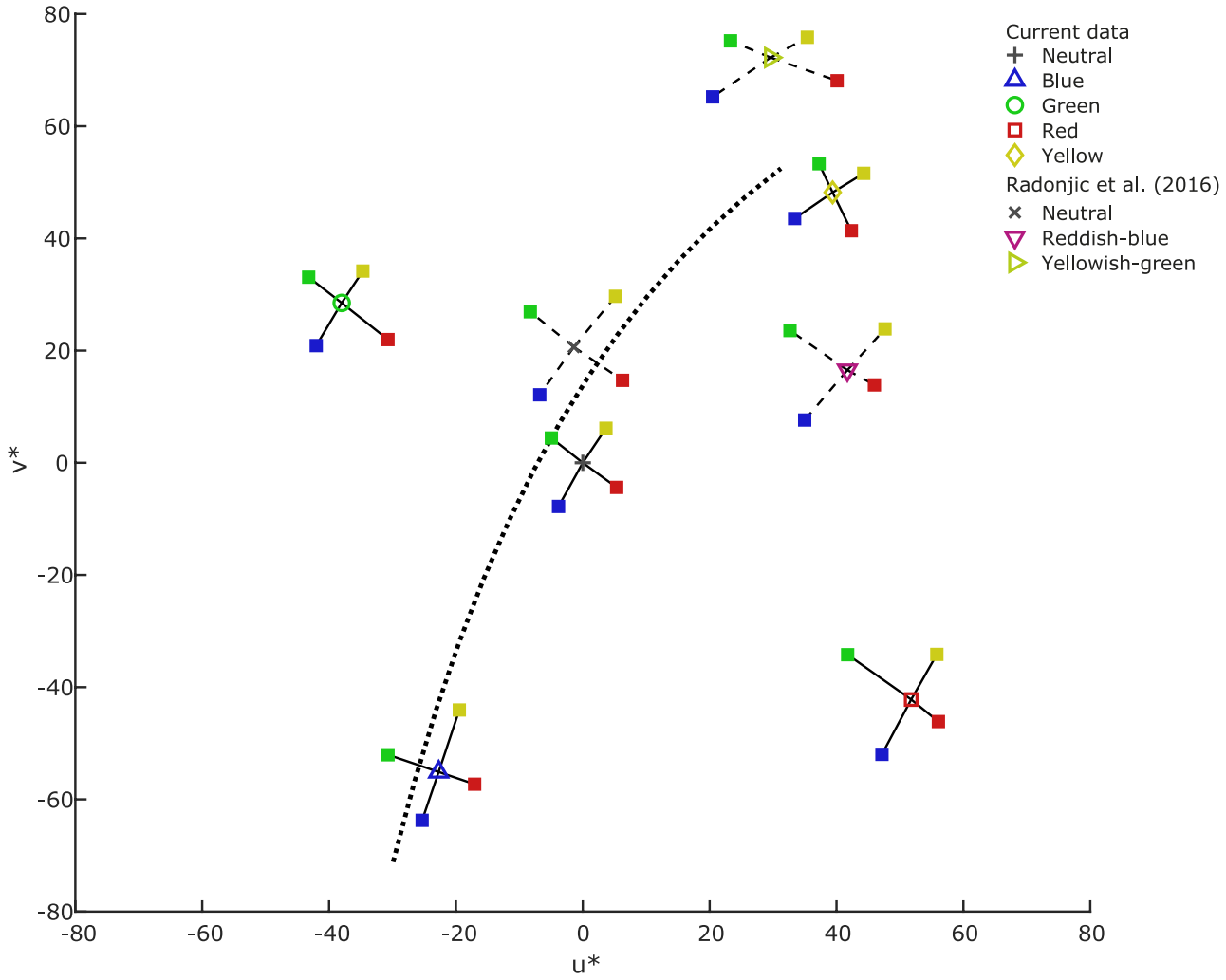

Figure 7. Mean thresholds (across participants, calculated using the image mean lookup tables) for the current study (solid lines) and the data in Radonjić, Pearce et al. (2016) (dashed lines) plotted in the CIELUV chromaticity plane. Black dotted line shows the Planckian locus (an estimate of the daylight locus). For interpretation of symbols, see Figure 5.

blue and yellow (from the daylight locus); and red and green (from an orthogonal locus). For each reference illumination, we measured thresholds in four chromatic directions: bluer, yellower, redder, and greener. The surface composition of the scene was kept constant, with the average surface spectral reflectance remaining neutral. Shifting the chromaticity of the reference illumination thus introduced a chromatic bias to the entire scene and systematically altered the adaptation point. Our general aim was to probe the effect of illumination chromaticity on relative discrimination thresholds. Specifically, we asked whether the blue bias (elevated thresholds in the bluer direction) occurs for nonneutral reference illuminations, and whether other systematic biases emerge.

We find that relative illumination discrimination thresholds depend on the reference illumination chromaticity. For the neutral reference, we find that thresholds in the bluish direction are highest, replicating the blue bias we have reported previously. Across the other reference illuminations, there is a tendency for discrimination thresholds to be higher in the direction opposite to the illumination chromaticity bias. In other words, thresholds tend to be higher for illumination changes directed toward a neutral chromaticity: a neutral bias. If we average across all five reference illumination conditions, thresholds are highest in the blue direction. But interpreting this result must be done with caution, since there are reference illuminations for which thresholds in other illumination change directions are higher.

We considered whether the neutral bias we observe might simply be a consequence of properties of the CIELUV color space for stimuli distant from the fixed neutral white point used to transform the stimulus representation to that space. We therefore reanalyzed the data using the variable white point lookup tables (described in the Methods) and showed that the same

pattern of relative thresholds occurs as for the fixed white-point illumination lookup tables (see Appendix C). Therefore, the change in threshold pattern across illumination-change directions across different reference illuminations is not due simply to the choice of white point used for the CIELUV transformations. A similar stability of results was found when we analyzed the data using the image mean lookup tables (see Figure D1, Appendix D and supplementary statistical analyses in the online supplement), again suggesting a robustness of the broad pattern of thresholds with respect to different choices of how the data are represented.

Relation between illumination discrimination and chromatic discrimination

We do not currently have models that can predict illumination discrimination thresholds across illumination change directions and reference illuminations. One approach to developing such models is to link illumination discrimination thresholds to chromatic discrimination thresholds measured for spatially simple stimuli around varying loci of chromatic adaptation. However, there is relatively little direct evidence for asymmetries in chromatic discrimination contours, which would explain the relative illumination discrimination thresholds observed here and previously (Pearce et al., 2014; Radonjić, Pearce et al., 2016; Radonjić et al., 2018). Given that the daylight locus lies close to the S-cone–isolating axis in the chromaticity plane, an asymmetry in thresholds is consistent with known fundamental differences in the neural processing of S-cone–contrast increments versus decrements (Dacey, Crook, & Packer, 2014; Neitz & Neitz, 2016; Smithson, 2014) and there is at least one behavioral report that at suprathreshold contrast levels, sensitivity to increments along the S-cone axis (blue) in DKL space (Derrington, Krauskopf, & Lennie, 1984) is lower than decrements (yellow; Vingrys & Mahon, 1998). A recent report (Weiss et al., 2017) suggests that a blue-yellow asymmetry is found for chromatic detection thresholds around a neutral reference chromaticity, which may relate to the blue bias seen in illumination discrimination thresholds. Yet most studies of chromatic discrimination focus on asymmetries in processing between color-opponent channels rather than within. Also, those studies that have explored the effects of varying loci of adaptation on chromatic discrimination contours do not report asymmetries resembling a neutral bias (Krauskopf & Gegenfurtner, 1992; Loomis & Berger, 1979). For example, Krauskopf and Gegenfurtner (1992) found that detection thresholds for very small targets against large uniform chromatic backgrounds are constant and symmetric in the red and green directions for varying background chromaticities, while thresholds increase symmetrically in the blue and yellow directions for increasing levels of background S-cone stimulation. Another asymmetry that has been observed in chromatic discrimination experiments is between changes in hue and changes in saturation, with higher thresholds reported for the latter (Danilova & Mollon, 2016; Elliott, Werner, & Webster, 2012; Krauskopf & Gegenfurtner, 1992).

Although asymmetries have been observed in chromatic discrimination thresholds, as outlined above, linking these to the asymmetries we observe in illumination discrimination thresholds awaits a more general effort to understand how thresholds measured for single surfaces relates to those measured for more complex spatial configurations, such as those used in the illumination discrimination paradigm. Indeed, making such links will require development of models that predict how the discriminability of individual stimulus patch changes is related to overall image discriminability. This question might, for example, be approached through summation experiments that gradually increase the number of distinct patches in the stimulus. With respect to this latter issue, we have shown that shuffling the surface reflectances of the stimulus scene across intervals of the IDT, while keeping the mean reflectance essentially constant, increases illumination discrimination thresholds but does not eliminate the ability of participants to make illumination discriminations or alter the pattern of relative thresholds (Radonjić et al., 2018). This result shows that a simple model that predicts the discriminability of complex scenes on the basis of their spatial mean will not be sufficient (but see Rinner & Gegenfurtner, 2002).

Another approach to understanding the mechanisms mediating illumination discrimination is to consider how the information available at various stages of visual processing shapes performance. This general approach has been fruitful for understanding aspects of color and pattern discrimination (Chaparro, Stromeyer, Huang, Kronauer, & Eskew, 1993; Geisler, 1989; Sekiguchi, Williams, & Brainard, 1993). We have begun work to analyze illumination discrimination in this manner. We have developed a quantitative computational-observer analysis of the data of Radonjić, Pearce et al. (2016). The analysis suggests that the variation in relative thresholds for the different scenes cannot be explained simply in terms of changes in the information content of the stimuli assessed at the level of the cone excitations (Ding et al., 2018). The computational observer predicts that the pattern of thresholds observed for the neutral scene would be maintained in the yellowish-green and reddish-blue scenes, given only the information available in the excitations of the cone photoreceptors. This is contrary

to the conjecture made by Radonjić, Pearce et al. (2016) that changes in relative thresholds for scenes with biased surface reflectances are due to changes in signal size.

Relation between illumination discrimination and color constancy

As we noted in the Introduction, performance on the IDT may be related to color constancy. The most straightforward link depends on the assumption that participants are able to discern an illumination change only if they perceive a change in the color appearance of at least one surface in the scene. Under this assumption, being below threshold on the IDT means subjects see no reliable change in surface color appearance and are thus exhibiting essentially perfect color constancy. Put another way: given that there are factors in visual processing that prevent discrimination of sufficiently small changes in a chromatic stimulus, whatever their cause, the IDT measures which directions and magnitudes of illumination change fall outside those discrimination limits, for particular scenes. Under this view, the larger thresholds in the bluer direction found for the neutral reference illumination imply better color constancy for small illumination changes toward blue. This observation leads to an interesting follow up question. Specifically, does the relative size of illumination discrimination thresholds across different illumination change directions and reference illuminations predict the stability of object color for corresponding suprathreshold illumination changes? That is, for large illumination shifts under which constancy is imperfect, is the size of the shift in surface color appearance predicted by the size of the illumination discrimination threshold? Weiss et al. (2017) referred to unpublished data that made such a comparison, and described a positive but nonsignificant correlation between a measure of the degree of color constancy (achromatic adjustment) and the size of illumination discrimination thresholds, consistent with the hypothesis outlined above. We note, however, that Weiss et al. (2017) interpreted the data with respect to a different hypothesis about the link between the two types of measures, one that led them to predict that a link would lead to negative rather than positive correlation.

Other results reported by Weiss et al. (2017) support the notion that color constancy is better for illumination changes in a bluer direction from neutral. Achromatic adjustments, taken as estimates of the illumination chromaticity, are more accurate for bluish illuminations and particularly for a blue hue which lies on the daylight locus. Naming of the background color in a simulated scene (a measure of perceived illumination color) also reveals a blue bias: The background was often named as achromatic under bluish illumination, but not under other chromatic illuminations.

A full exploration of the links between illumination discrimination and color constancy will require experiments that assess illumination discrimination thresholds while also collecting a measure of the degree of surface appearance shifts under carefully matched stimulus conditions. Indeed, given the variation of illumination discrimination thresholds with reference illumination and the distribution of surface reflectances in the scene, it is critical that comparisons across the two types of measurements carefully account for these factors. In addition, measuring individual participant achromatic points and including these in analyzing different patterns of illumination discrimination thresholds might yield regularities that are not otherwise apparent.

Are color constancy mechanisms optimized for the statistics of the natural environment?

A second interesting question is whether the observed pattern of illumination discrimination thresholds is related to the statistics of natural daylight. It is well known that some illuminations are more common in the natural environment than others (Hernández-Andrés et al., 2001; Spitschan et al., 2016), and given that the computational problem of parsing the retinal image into separate contributions of illumination and surface reflectance is underdetermined (Hurlbert, 1998; Maloney, 1999), the visual system may well use prior information about likely illuminations in its processing of color. Here we speculate about this possibility.

Exactly how the visual system incorporates prior knowledge into the processing of color information requires further study. One hypothesis is that the visual system has evolved to become less sensitive to illumination changes that are more likely in the natural environment, an idea motivated by the potential links between IDT and color constancy outlined above. This explanation does not require that the observer recovers an explicit estimate of the illumination chromaticity.

A second hypothesis is that the visual system stores a representation of the statistics of natural daylights, either innately or learned during development. Such a representation may then influence explicit or implicit estimates of the illumination. This type of explanation may be modeled in a Bayesian framework (Brainard & Freeman, 1997; Knill & Pouget, 2004; Mamassian, Landy, & Maloney, 2003; Pouget, Beck, Ma, & Latham, 2013), where the sensory evidence for the illumination color may be combined with prior expectations of illuminations. Experimentally, there is evidence that the role of priors may be manipulated by

varying the memory retention interval in simple color discrimination tasks (Olkkonen, McCarthy, & Allred, 2014). Similar biases can be demonstrated for illumination discrimination (Aston, Olkkonen, & Hurlbert, 2017). So one possibility is that the requirement to hold the reference *illumination* in memory for longer than the two comparison illuminations produces a differential shift in the representation of the three illuminations, with the representation of the reference more shifted toward the prior mean. This type of effect could produce the neutral bias that we report. There are also other possibilities. For example, whether Bayesian estimators produce biases toward or away from the mean of the prior distribution depends on the precise form of the prior and the likelihood functions driving the estimators: It is not possible to intuit simply that biases will always be toward the prior mean (Wei & Stocker, 2015). One approach to testing the role of memory in the IDT would be to vary the retention intervals or have observers judge simultaneously presented scenes. In any case, there is sufficient regularity in the statistical distribution of natural daylight that it will be worthwhile to develop quantitative models of the illumination estimation process and ask whether these models can explain the biases we observe.

Conclusion

We find that for gray world scenes under chromatically biased illuminations, relative illumination discrimination thresholds strongly depend on the reference illumination chromaticity, with a tendency for discrimination thresholds to be higher in the opposite direction to the chromatic bias. In other words, thresholds tend to be higher for illumination changes directed toward a neutral chromaticity: a neutral bias. Comparison with previous results for illumination discrimination on scenes with chromatically biased surface reflectances (Radonjić, Pearce et al., 2016) suggests that more generally, changes in illumination that move the mean image chromaticity toward neutral are less easily discriminated. The results have potential implications for color constancy: In considering the stability of surface appearance under changes in illumination, both the starting chromaticity of the illumination and direction of change must be considered, as well as the chromatic characteristics of the surface reflectance ensemble. The results are also consistent with the assumption that the visual system embeds prior expectations of the likelihood of natural illumination chromaticity and their changes, and adjusts its sensitivity accordingly.

Keywords: illumination discrimination, color constancy, illumination priors, spectrally tuneable LED lamps

Acknowledgments

We would like to thank Jim Groombridge for help with setting up the experiment and data collection, May Noonan for help with data collection, and Brad Pearce for contributing code to control the spectrally tuneable luminaires. We thank the Wellcome Trust (102562/Z/13/Z to SA) for funding this work.

Commercial relationships: none.
Corresponding author: Stacey Aston.
Email: stacey.j.aston@durham.ac.uk.
Address: Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK.

Footnote

[1] In vision science, the term "lookup table" is often associated with the gamma correction step of computerized display control. Here, however, we use the term in its generic sense to refer to tables of values that give the correspondence of nominal to measured differences (in CIELUV) between the reference and test illuminations. For example, the table shows that the fifth nominal step in the bluer staircase for the neutral reference condition corresponds to a measured difference of 5.13 $\Delta E_{u^*v^*}$ between the test and reference illumination (using the fixed white point lookup tables).

References

Alvaro, L., Linhares, J. M. M., Moreira, H., Lillo, J., & Nascimento, S. M. C. (2017). Robust colour constancy in red-green dichromats. *PLoS One*, *12*(6), 1–17.

Arend, L., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America, A: Optics and Image Science*, *3*(10), 1743–1751, https://doi.org/10.1364/JOSAA.3.001743.

Aston, S., Le Couteur Bisson, T., Jordan, G., & Hurlbert, A. C. (2016). Better colour constancy or worse discrimination? Illumination discrimination in colour anomalous observers. In *Proceedings of the 39th European Conference on Visual Perception, Vol 45, (ECVP)* (pp. 207–207).

Aston, S., Olkkonen, M., & Hurlbert, A. (2017). Memory bias for illumination colour. *Journal of Vision*, *17*(10):130, https://doi.org/10.1167/17.10.130. [Abstract]

Bäuml, K.-H. (1994). Color appearance: Effects of illuminant changes under different surface collections. *Journal of the Optical Society of America A*, *11*(2), 531–542, https://doi.org/10.1364/JOSAA.11.000531.

Bäuml, K.-H. (1995). Illuminant changes under different surface collections: examining some principles of color appearance. *Journal of the Optical Society of America A*, *12*(2), 261–271, https://doi.org/10.1364/JOSAA.12.000261.

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*, *15*(2), 307–325, https://doi.org/10.1364/JOSAA.15.000307.

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*(9), 2091–2110, https://doi.org/10.1364/JOSAA.14.002091.

Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical Society of America, A: Optics, Image Science, and Vision*, *14*(7), 1393–1411, https://doi.org/10.1364/JOSAA.14.001393.

Brainard, D. H., Pelli, D. G., & Robson, T. (2002, January 15). Display characterization. In J. Hornak (Ed.), *Encyclopedia of Imaging Science and Technology* (pp. 172–188). Oxford, UK: Wiley. https://doi.org/doi:10.1002/0471443395.img011

Brainard, D. H., & Radonjić, A. (2014). Color constancy. *The New Visual Neurosciences*, *1*, 545–556.

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*, *9*(9), 1433–1448, https://doi.org/10.1364/Josaa.9.001433.

Burnham, R. W., Evans, R. M., & Newhall, S. M. (1957). Prediction of color appearance with different adaptation illuminations. *Journal of the Optical Society of America*, *47*(1), 35–42, https://doi.org/10.1364/JOSA.47.000035.

Chaparro, A., Stromeyer, C. F., III, Huang, E. P., Kronauer, R. E., & Eskew, R. T. J. (1993, January 28). Colour is what the eye sees best. *Nature*, *361*(6410), 348–350. https://doi.org/10.1038/361348a0

Craven, B. J., & Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research*, *32*(7), 1359–1366, https://doi.org/10.1016/0042-6989(92)90228-B.

Dacey, D. M., Crook, J. D., & Packer, O. S. (2014). Distinct synaptic mechanisms create parallel S-ON and S-OFF color opponent pathways in the primate retina. *Visual Neuroscience*, *31*(2), 139–151, https://doi.org/10.1017/S0952523813000230.

Danilova, M. V., & Mollon, J. D. (2016). Superior discrimination for hue than for saturation and an explanation in terms of correlated neural noise. *Proceedings of the Royal Society B*: *Biological Sciences*, *283*(1831).

Daugirdiene, A., Kulikowski, J. J., Murray, I. J., & Kelly, J. M. F. (2016). Test illuminant location with respect to the Planckian locus affects chromaticity shifts of real Munsell chips. *Journal of the Optical Society of America A*, *33*(3), A77–A84, https://doi.org/10.1364/JOSAA.33.000A77.

Delahunt, P. B., & Brainard, D. H. (2004). Does human color constancy incorporate the statistical regularity of natural daylight? *Journal of Vision*, *4*(2):1, 57–81, https://doi.org/10.1167/4.2.1. [PubMed] [Article]

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

Ding, X., Radonjić, A., Cottaris, N. P., Jiang, H., Wandell, B. A., & Brainard, D. (2018). Computational-observer analysis of illumination discrimination. *bioRxiv* 302315, https://doi.org/10.1101/302315.

Elliott, S. L., Werner, J. S., & Webster, M. A. (2012). Individual and age-related variation in chromatic contrast adaptation. *Journal of Vision*, *12*(8):11, 1–21, https://doi.org/10.1167/12.8.11. [PubMed] [Article]

Finlayson, G., Mackiewicz, M., Hurlbert, A., Pearce, B., & Crichton, S. (2014). On calculating metamer sets for spectrally tunable LED illuminators. *Journal of the Optical Society of America, A: Optics, Image Science, and Vision*, *31*(7), 1577–1587, https://doi.org/10.1364/JOSAA.31.001577.

Foster, D. H. (2011). Color constancy. *Vision Research*, *51*(7), 674–700, https://doi.org/10.1016/j.visres.2010.09.006.

Geisler, W. S. (1989). Sequential ideal-observer analysis of visual discriminations. *Psychological Review*, *96*(2), 267–314.

Helson, H., & Michels, W. C. (1948). The effect of chromatic adaptation on achromaticity. *Journal of the Optical Society of America*, *38*(12), 1025–1032, https://doi.org/10.1364/JOSA.38.001025.

Hernández-Andrés, J., Romero, J., Nieves, J. L., & Lee, R. L. (2001). Color and spectral analysis of daylight in southern Europe. *Journal of the Optical Society of America, A: Optics, Image Science, and Vision*, *18*(6), 1325–1335. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11393625

Hurlbert, A. C. (1998). Computational models of color constancy. In V. Walsh & J. Kulikowski (Eds.), *Perceptual constancy: Why things look as they do* (1st ed., pp. 283–232). Cambridge, UK: Cambridge University Press.

International Commission on Illumination (CIE). (2004). *Colorimetry* (3rd ed.). Paris: Bureau Central de la CIE.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, *49*(3), 227–229. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2011460

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719, https://doi.org/10.1016/j.tins.2004.10.007.

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*(1), 307–312. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=15135&tool=pmcentrez&rendertype=abstract

Krauskopf, J., & Gegenfurtner, K. R. (1992). Color discrimination and adaptation. *Vision Research*, *32*(11), 2165–2175, https://doi.org/https://doi.org/10.1016/0042-6989(92)90077-V.

Ling, Y., & Hurlbert, A. (2008). Role of color memory in successive color constancy. *Journal of the Optical Society of America A*, *25*(6), 1215, https://doi.org/10.1364/JOSAA.25.001215.

Logvinenko, A. D., & Maloney, L. T. (2006). The proximity structure of achromatic surface colors and the impossibility of asymmetric lightness matching. *Perception & Psychophysics*, *68*(1), 76–83, https://doi.org/10.3758/BF03193657.

Logvinenko, A., & Menshikova, G. (1994). Trade-off between achromatic colour and perceived illumination as revealed by the use of pseudoscopic inversion of apparent depth. *Perception*, *23*(9), 1007–1023, https://doi.org/10.1068/p231007.

Loomis, J. M., & Berger, T. (1979). Effects of chromatic adaptation on color discrimination and color appearance. *Vision Research*, *19*(8), 891–901, https://doi.org/https://doi.org/10.1016/0042-6989(79)90023-3.

Lucassen, M. P., Gevers, T., Gijsenij, A., & Dekker, N. (2013). Effects of chromatic image statistics on illumination induced color differences. *Journal of the Optical Society of America A*, *30*(9), 1871–1884.

Maloney, L. T. (1999). Physics-based approaches to modeling surface color perception. In K. R. Gegenfurtner & L. T. Sharpe (Eds.), *Color vision: From genes to perception* (pp. 387–422). Cambridge, UK: Cambridge University Press.

Mamassian, P., Landy, M., & Maloney, L. T. (2003). Bayesian modelling of visual perception. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabalistic models of the brain: Perception and neural function* (pp. 13–36). Cambridge, MA: MIT Press.

Neitz, J., & Neitz, M. (2016). Evolution of the circuitry for conscious color vision in primates. *Eye*, *31*, 286, http://dx.doi.org/10.1038/eye.2016.257.

Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision*, *14*(11):5, 1–15, https://doi.org/10.1167/14.11.5. [PubMed] [Article]

Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of Vision*, *10*(9):16, 1–22, https://doi.org/10.1167/10.9.16. [PubMed] [Article]

Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., & Hurlbert, A. (2014). Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. *PLoS One*, *9*(2), e87989, https://doi.org/10.1371/journal.pone.0087989.

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: Knowns and unknowns. *Nature Neuroscience*, *16*(9), https://doi.org/10.1038/nn.3495.

Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2015). Color constancy in a naturalistic goal-directed task. *Journal of Vision*, *15*(13):3, 1–21, https://doi.org/10.1167/15.13.3. [PubMed] [Article]

Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2016). Color constancy supports cross-illumination color selection. *Journal of Vision*, *15*(6):13, 1–19, https://doi.org/10.1167/15.6.13. [PubMed] [Article]

Radonjić, A., Ding, X., Krieger, A., Aston, S., Hurlbert, A. C., & Brainard, D. H. (2018). Illumination discrimination in the absence of a fixed surface-reflectance layout. *Journal of Vision*, *18*(5):11, 1–27, https://doi.org/10.1167/18.5.11. [PubMed] [Article]

Radonjić, A., Pearce, B., Aston, S., Krieger, A., Cottaris, N. P., Brainard, D. H., & Hurlbert, A. C. (2016). Illumination discrimination in real and

simulated scenes. *Journal of Vision*, *16*(11):2, 1–18, https://doi.org/10.1167/16.11.2. [PubMed] [Article]

Rinner, O., & Gegenfurtner, K. R. (2002). Cone contributions to colour constancy. *Perception*, *31*, 733–746, https://doi.org/10.1068/p3352.

Rutherford, M. D., & Brainard, D. H. (2002). Lightness constancy: A direct test of the illumination-estimation hypothesis. *Psychological Science*, *13*(2), 142–149, https://doi.org/10.1111/1467-9280.00426.

Sekiguchi, N., Williams, D. R., & Brainard, D. H. (1993). Efficiency in detection of isoluminant and isochromatic interference fringes. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *10*(10), 2118–2133.

Smithson, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *360*(1458), 1329–1346, https://doi.org/10.1098/rstb.2005.1633.

Smithson, H. E. (2014). S-cone psychophysics. *Visual Neuroscience*, *31*(2), 211–225, https://doi.org/10.1017/S0952523814000030.

Spitschan, M., Aguirre, G. K., Brainard, D. H., & Sweeney, A. M. (2016). Variation of outdoor illumination as a function of solar elevation and light pollution. *Nature Publishing Group*, 1–13, https://doi.org/10.1038/srep26756.

Troost, J. M., & de Weert, C. M. M. (1991). Naming versus matching in color constancy. *Perception & Psychophysicsn*, *50*(6), 591–602.

Vingrys, A. J., & Mahon, L. E. (1998). Color and luminance detection and discrimination asymmetries and interactions. *Vision Research*, *38*(8), 1085–1095, https://doi.org/https://doi.org/10.1016/S0042-6989(97)00250-2.

Wei, X.-X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain "anti-Bayesian" percepts. *Nature Neuroscience*, *18*, 1509, http://dx.doi.org/10.1038/nn.4105.

Weiss, D., Witzel, C., & Gegenfurtner, K. (2017). Determinants of colour constancy and the blue bias. *I-Perception*, *8*(6), https://doi.org/10.1177/2041669517739635.

Worthey, J. A. (1985). Limitations of color constancy. *Journal of the Optical Society of America A*, *2*(7), 1014–1026.

Wyszecki, G., & Stiles, W. S. (1967). *Color science*. Oxford, UK: Wiley.

Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vision Research*, *41*(20), 2581–2600, https://doi.org/10.1016/S0042-6989(01)00143-2.

Zaidi, Q. (1998). Identification of illuminant and object colors: Heuristic-based algorithms. *Journal of the Optical Society of America, A: Optics, Image Science, and Vision*, *15*(7), 1767–1776, https://doi.org/10.1364/JOSAA.15.001767.

Appendix A

Mondrian specifications

The spectral reflectance distribution of each unique patch used to make up the Mondrian lining was found by first taking a radiance measurement from one sample of each unique patch with a CS2000 Konica Minolta spectroradiometer (Konica Minolta, Nieuwegein, Netherlands). The location of each patch is shown in Figure A1A. These measurements were then subject to pointwise division by a measurement of the incident illumination spectrum taken from a calibration tile

Patch	CIE *x*	CIE *y*
1	0.35	0.49
2	0.37	0.26
3	0.24	0.28
4	0.21	0.14
5	0.37	0.23
6	0.28	0.21
7	0.49	0.37
8	0.46	0.38
9	0.45	0.44
10	0.46	0.29
11	0.20	0.18
12	0.30	0.48
13	0.26	0.39
14	0.39	0.49
15	0.27	0.17
16	0.42	0.28
17	0.42	0.40
18	0.23	0.23
19	0.34	0.52
20	0.25	0.18
21	0.25	0.28
22	0.30	0.44
23	0.37	0.45
24	0.28	0.39

Table A1. CIE *xy* chromaticities of each unique patch in the Mondrian under a hypothetical equal energy white light. *Note*: The mean *xy* chromaticity was [0.33, 0.33] when computed across patches and [0.34, 0.33] when computed across the entire hyperspectral image.

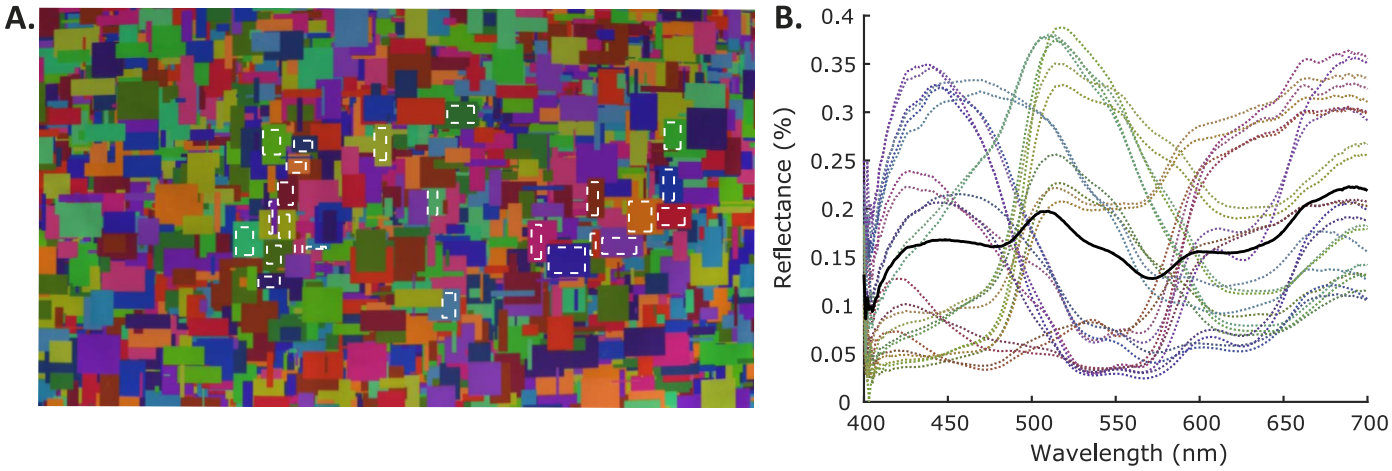

Figure A1. Parameters of the Mondrian. (A) The locations of the 24 measured patches on the back wall of the stimulus box. (B) The surface reflectance of each of the 24 unique patches used to make up the Mondrian lining of the box. Black solid line is the average reflectance of these 24 patches.

(placed in the center of the back wall) to find the spectral reflectance distributions (Figure A1B). Computed surface reflectance functions for each patch are provided in the online supplement. We calculated the tristimulus values and chromaticity that each patch would have under a hypothetical equal energy white light (using the CIE 2004 color matching functions).

Appendix B

Hyperspectral imaging and illumination modeling

Hyperspectral images of the back wall of the stimulus box were captured using a Specim V10E camera. A surface reflectance image of the back wall was obtained by first imaging the back wall under an arbitrary white light before removing the Mondrian lining from the box and then taking an image of the white reflectance tile covering the back wall. Because the tile is smaller than the wall, the latter image was constructed by combining three images, in which the white tile was placed either flush to the right wall of the box, in the center of the box, or flush to the left wall of the box (such that all locations on the back wall were covered by the tile in at least one of the three images). Thus, we obtained a complete representation of the spatial gradients of irradiance on the back wall. Both hyperspectral images (the image of the Mondrian-papered back wall and the combined image of the white reflectance tile) were then cropped to remove any areas of the image that were above, below or to either side of the back wall of the box. The combined and cropped hyperspectral image of the white reflectance tiles was smoothed using a two-dimensional Gaussian kernel. Finally, this smoothed image of the white reflectance tile was used to estimate the illumination at each point in the scene and discount it from the image of the Mondrian-papered back wall to obtain spectral reflectance values at each pixel using point-wise division (here "pixel" refers to a point in a data cube that represents measured light spectra at different locations). To model the light reflected from each point of the Mondrian-papered back wall under each illumination, the measured spectrum of each illumination (see previous section) was combined with the measured surface reflectance at each pixel using point-wise multiplication. These images were then used to estimate the mean image tristimulus values and chromaticity under each illumination used in the experiment, in order to form the image mean lookup tables used for analysis below.

Appendix C

Analyzing thresholds using alternate lookup tables

The image mean lookup tables

If we repeat the ANOVA analysis presented in the main text on the data obtained using the image mean lookup tables, a similar pattern of results emerges (Figure C1; Table C1). We can quantify the strength of the correspondence between the two threshold sets by calculating Pearson's correlation coefficient ($r = 0.981$, $p < 0.001$; Figure C1C), although we note that the absolute size of mean thresholds is lower in 19 out of 20 cases for the image mean lookup tables than for the fixed

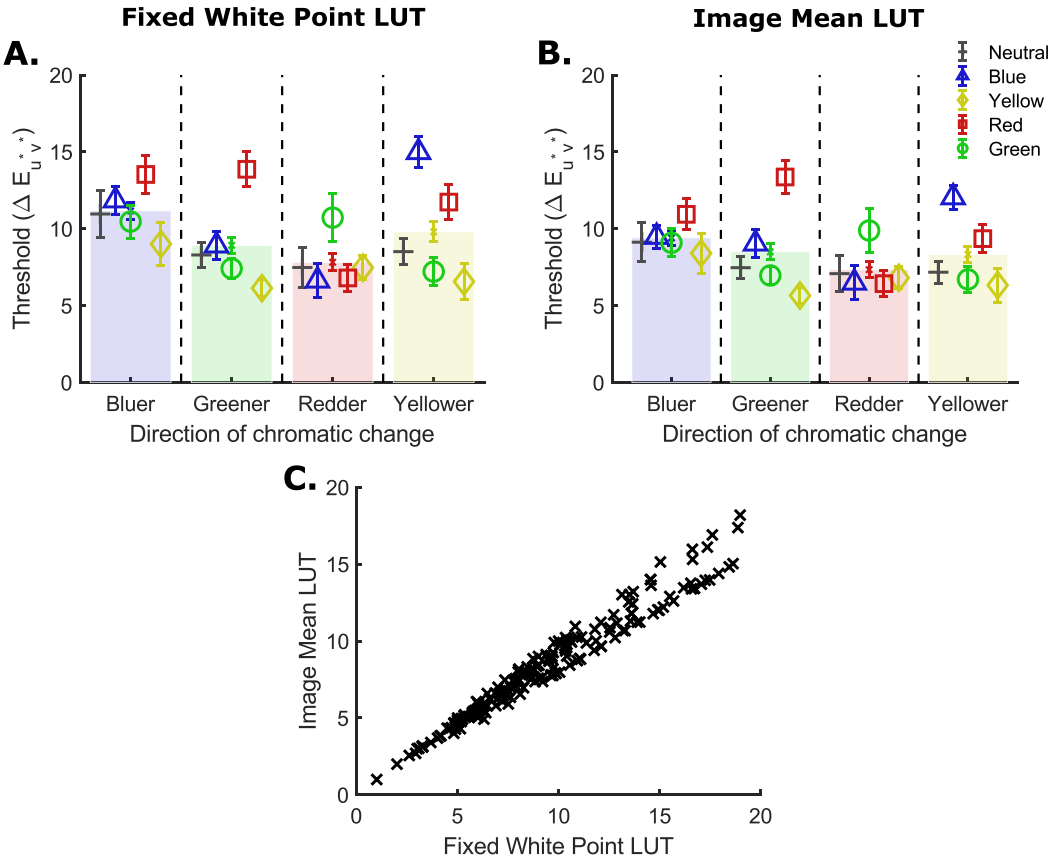

Figure C1. Thresholds calculated using the image mean lookup tables. (A) Thresholds calculated using the fixed white point lookup tables (LUTs) plotted in a compact format for comparison. (B) Thresholds across the different directions of change for the five reference illumination conditions. Transparent bars represent the main effect of chromatic direction of change. They are the thresholds for each chromatic direction averaged over reference illuminations. (C) Scatter plot of the threshold data calculated using the image mean LUTs plotted against threshold data calculated using the fixed white point LUT described in the main text. All thresholds are plotted for each participant, with 20 thresholds per participant and 180 points in total.

white point lookup tables (compare Table C1 with Table 1). We still find a significant interaction effect of reference illumination and illumination-change direction on illumination discrimination thresholds, $F(12, 96) = 7.66$, $p < 0.001$, a significant main effect of reference illumination, $F(4, 32) = 9.51$, $p < 0.001$, and a significant main effect of illumination-change direction, $F(3, 24) = 3.88$, $p = 0.022$. However, although thresholds in the bluer direction are still the highest when averaged over reference illumination conditions, no pairwise comparisons between the different directions are significant.

	Illumination-change direction			
Reference	Bluer	Greener	Redder	Yellower
Neutral	9.13 (1.18)	7.47 (0.69)	7.08 (1.12)	7.18 (0.68)
Blue	9.47 (0.69)	9.04 (0.86)	6.49 (1.06)	12.04 (0.76)
Green	9.09 (0.87)	6.96 (0.57)	9.89 (1.37)	6.71 (0.79)
Red	10.94 (0.95)	13.37 (1.02)	6.43 (0.80)	9.36 (0.88)
Yellow	8.41 (1.21)	5.66 (0.47)	6.81 (0.68)	6.33 (1.03)

Table C1. Mean illumination discrimination thresholds in the different reference illumination conditions using the image mean lookup tables. *Note*: Values in parentheses show the standard error.

Using these lookup tables, there was only a simple main effect of chromatic direction of illumination change for the blue and red reference illumination conditions, $F(3, 24) = 7.40$, $p < 0.005$, and $F(3, 24) = 11.84$, $p < 0.005$, respectively, but not in the neutral, green, and yellow reference illumination conditions, $F(3, 24) = 2.32$, $p = 0.505$; $F(3, 24) = 4.41$, $p = 0.065$; and $F(3, 24) = 2.81$, $p = 0.305$, respectively.

There was a simple main effect of reference illumination on illumination discrimination thresholds for the greener and yellower chromatic directions of change, $F(4, 32) = 15.85$, $p < 0.004$ and $F(4, 32) = 10.99$, $p < 0.004$, respectively, but not for bluer and redder, $F(4, 32) = 1.84$, $p = 0.580$ and $F(1.69, 13.57) = 3.68$, $p = 0.056$, with a Greenhouse-Geisser correction, respectively.

For completeness and comparison to the results in the main text, statistical analysis of finer-grained effects can be found in the online supplement.

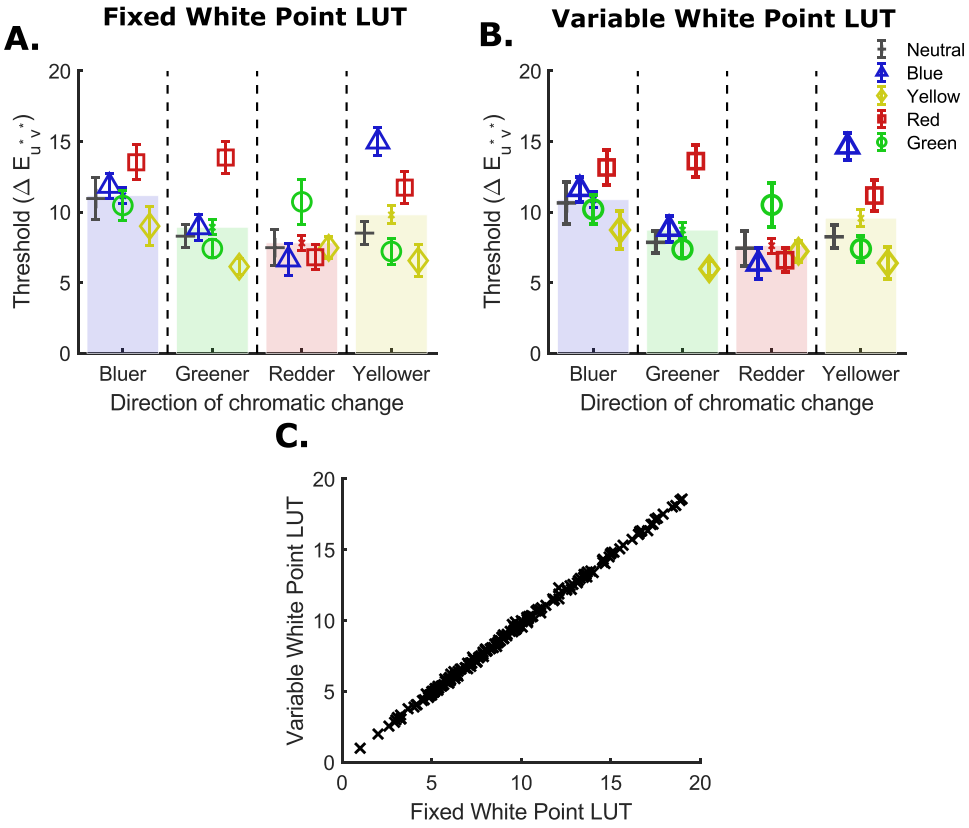

Figure C2. Thresholds calculated using the variable white point LUTs. (A) Thresholds calculated using the fixed white point LUT plotted in a compact format for comparison. (B) Thresholds calculated using the variable white point LUT. Transparent bars represent the main effect of chromatic direction of change in both (A) and (B). They are the thresholds for each chromatic direction averaged over reference illuminations. (C) Scatter plot of the threshold data calculated using the variable white point LUTs plotted against threshold data calculated using the fixed white point LUTs. All thresholds are plotted for each participant, with 20 thresholds per participant and 180 points in total.

Variable white point lookup tables

When thresholds are calculated using the reference illumination from each condition as the white point, the ordering of mean thresholds within each condition and across conditions remains the same (Figure C2; Table C2). Again, we quantified the strength of the correspondence between the two threshold sets by calculating Pearson's correlation coefficient ($r = 0.999$, $p < 0.001$; Figure C2C).

Once again we repeat the ANOVA analysis. We find a significant interaction effect of reference illumination and chromatic direction of illumination change on illumination discrimination thresholds, $F(12, 96) = 9.51$, $p < 0.001$. Moreover, there was a significant main effect of reference illumination, regardless of the direction of chromatic change, $F(4, 32) = 12.70$, $p < 0.001$, and a significant main effect of chromatic direction of illumination change, regardless of reference illumination condition, $F(3, 24) = 8.34$, $p = 0.001$. When thresholds are averaged over the different reference illumination conditions, thresholds for the bluer and yellower illumination-change directions are significantly higher than redder thresholds ($p = 0.024$ and $p = 0.007$, respectively). There were no other significant differences between the different directions of chromatic change (see online supplement).

As the interaction term was significant, we explored simple main effects. There was a simple main effect of chromatic direction of illumination change for the blue and red reference illumination conditions, $F(3, 24) = 15.80$, $p < 0.005$ and $F(3, 24) = 11.80$, $p < 0.005$, respectively, but not in the neutral, green, and yellow reference illumination conditions, $F(3, 24) = 4.20$, $p = 0.064$; $F(3, 24) = 4.57$, $p = 0.055$; and $F(3, 24) = 2.88$, $p = 0.285$, respectively.

	Illumination-change direction			
Reference	Bluer	Greener	Redder	Yellower
Neutral	10.65 (1.47)	7.86 (0.77)	7.45 (1.25)	8.25 (0.82)
Blue	11.63 (0.89)	8.81 (0.90)	6.34 (1.09)	14.64 (0.98)
Green	10.20 (1.05)	7.36 (0.63)	10.51 (1.55)	7.39 (0.91)
Red	13.17 (1.21)	13.61 (1.11)	6.59 (0.85)	11.18 (1.11)
Yellow	8.73 (1.34)	5.98 (0.53)	7.23 (0.76)	6.39 (1.11)

Table C2. Mean illumination discrimination thresholds in the different reference illumination conditions using the variable white point lookup tables. *Note*: Values in parentheses show the standard error.

There was a simple main effect of reference illumination on illumination discrimination thresholds in all chromatic directions, $F(4, 32) = 4.49$, $p = 0.020$ for bluer, $F(4, 32) = 14.54$, $p < 0.004$ for greener, and $F(4, 32) = 17.17$, $p < 0.004$ for yellower, except redder, $F(1.63, 13.07) = 4.57$, $p = 0.148$, with a Greenhouse-Geisser correction.

For completeness and comparison to the results in the main text, statistical analysis of finer-grained effects for different choices of lookup tables can be found in the online supplement.

Appendix D

Calculating image mean lookup tables for the Radonjić et al. (2016) data

In the Results section we compare our results with those from our previous study in which we measured illumination discrimination thresholds for different chromatically-biased scenes (experiment 2 in Radonjić et al., 2016). Unlike the experiment we report here, that study was conducted using simulated scenes presented stereoscopically on a pair of calibrated computer monitors. Detailed methods as well as the rationale for generalizing illumination discrimination results across real and simulated scenes are available in Radonjić et al. (2016). There we reported threshold measurements computed based on the lookup tables that quantify illumination differences based on spectra of the experimental illuminations used for rendering stimulus scenes. (These are conceptually equivalent to thresholds computed using fixed white point lookup tables). To compare the results from this earlier study to those we report here, we recomputed thresholds using image mean lookup tables. A different set of lookup tables was computed for the neutral, reddish-blue, and yellowish-green scene conditions. For each condition, we computed the mean scene CIEXYZ and CIELUV value for each stimulus stereo-image pair from the images' RGB values and monitor calibration data,

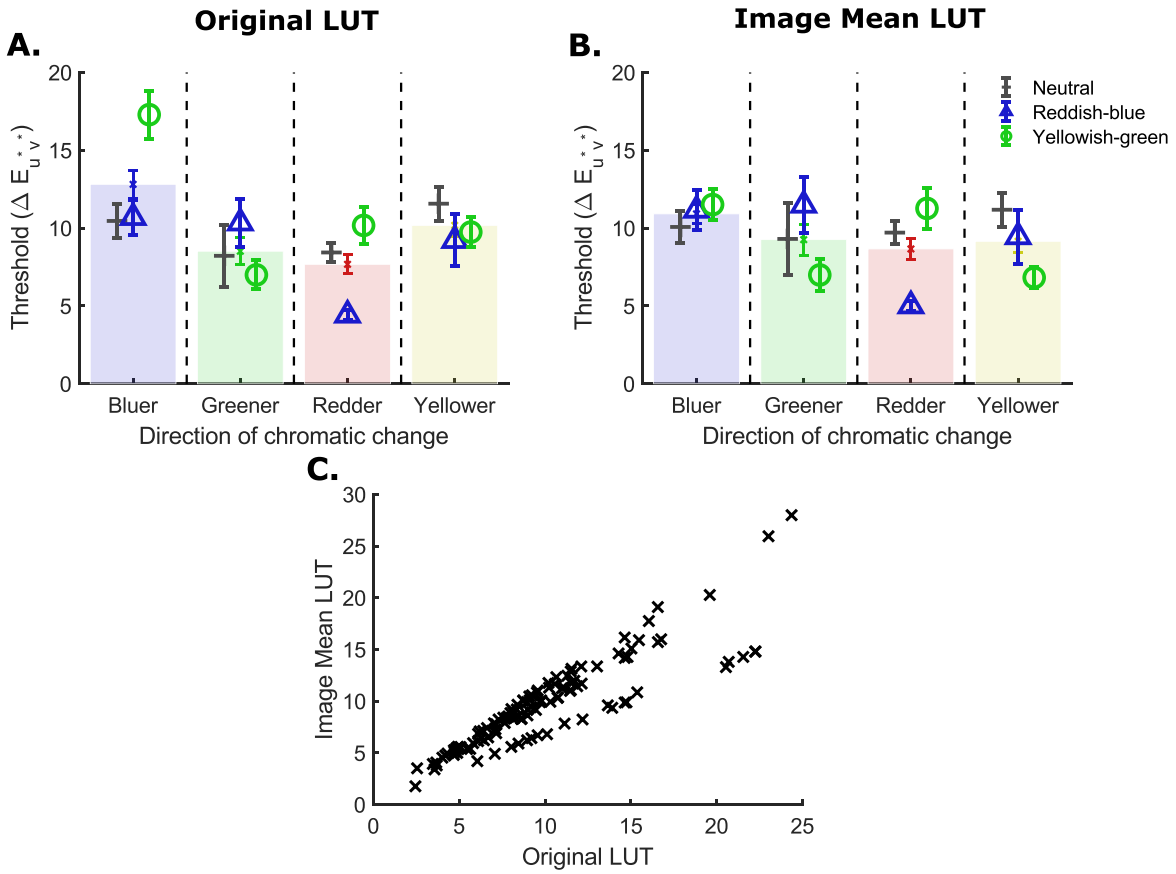

Figure D1. The effect of LUTs on illumination discrimination thresholds for the data of Radonjić et al. (2016). (A) Thresholds computed based on the original illumination-based LUTs. (B) Thresholds computed based on the image mean LUTs. Transparent bars represent the main effect of chromatic direction of change (thresholds for each chromatic direction averaged over scene conditions). (B) Scatter plot of the threshold data calculated using image mean LUTs (image mean LUT) against threshold data calculated using the original illumination-based LUTs (original LUT). Twelve thresholds are plotted for each participant (4 Directions × 3 Scene Conditions), giving 120 points in total.

using standard methods (Brainard, Pelli, & Robson, 2002). In the conversion from XYZ to LUV, we set the luminance values to be equal to the mean luminance across all images (201 pairs) for that scene condition, removing the contribution of small image-to-image variation in luminance from the computed $\Delta E_{u^*v^*}$ values. In conversion, we used the same white point chromaticity as the one we used for the image mean lookup tables generated to analyze the data for the experiment reported here (CIE $xy = [0.31, 0.33]$, with the luminance set to the mean image luminance for each scene condition). We then recomputed thresholds using the same methods as those we used in our earlier report. Note that these methods differ slightly from those used here (they are estimated from the psychometric function fit to all responses, rather than obtained by averaging points at staircase reversals). Thresholds computed with the two sets of lookup tables are in good agreement with the exception of the bluer and yellower illumination-change thresholds for the yellowish-green condition (Figure D1). As a result of interactions between the spectra of the illuminations for these directions and the chromatic-bias of the surface reflectance for this condition, thresholds for these two directions are relatively smaller when computed with the image mean lookup tables.

Appendix E

Fitting the three-component CIE daylight model to our illuminations

To assess how closely the spectral content of the illuminations used in the experiment matches that of daylight, we asked how well our illuminations could be modeled using the three-component International Commission on Illumination (CIE) daylight model (International Commission on Illumination (CIE), 2004). The CIE daylight model consists of three basis functions: S0, S1, and S2 (Figure E1A). To fit the CIE daylight model to our illuminations we first normalized the CIE daylight basis functions by their vector norm (L^2 norm). We then took each experimental illumination spectrum (Figure E1B), normalized it by its vector norm, and used multiple least squares regression to find the weightings of the three CIE daylight basis functions

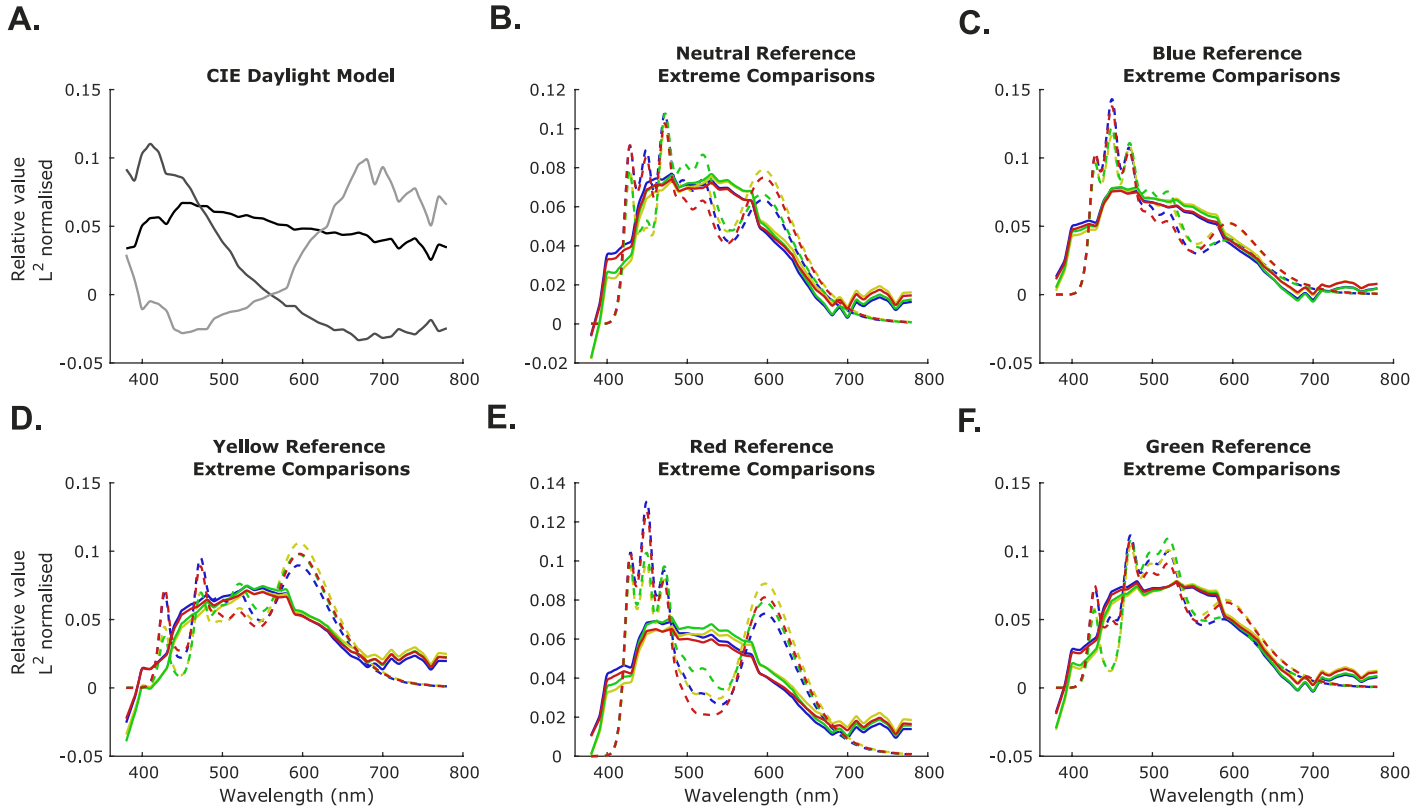

Figure E1. Fitting the three-component CIE daylight model to the experimental illuminations. (A) The L^2 normalized CIE daylight basis functions. S0 is in black, S1 is in medium gray, and S2 in light gray. (B–F) The measured spectral power distributions of the most extreme comparison illuminations (20 $\Delta E_{u^*v^*}$ away, colored according to illumination-change direction) used in the experiment for each reference illumination condition (dotted lines) compared with the recovered spectral power distributions of these illuminations obtained from fitting the three-component CIE daylight model (solid lines).

	Chromatic direction of change (%)			
Reference	Bluer	Greener	Redder	Yellower
Neutral	76.07	74.35	77.80	72.27
Blue	68.16	71.62	72.47	66.62
Green	60.21	55.07	61.21	54.76
Red	44.78	42.95	49.03	39.67
Yellow	79.72	76.37	76.13	79.68

Table E1. Table of goodness of fit values (R^2) for the CIE daylight model to each axis of chromatic change used in the experiment.

that provide the best fit to the experimental illumination. Taking a combination of the basis functions with these weightings applied gives the recovered spectrum (Figure E1C). The quality of the fit was assessed using the R^2 statistic (the coefficient of determination, the square of Pearson's correlation coefficient between the true and recovered spectra). Averaging over all illuminations, the mean goodness of fit was $R^2 = 64.95\% \pm 12.87\%$ ($M \pm SD$), indicating that the CIE daylight basis functions explain a large proportion of the variation in the experimental illuminations. In Table E1, we show the mean proportion of variance explained (mean R^2) for each chromatic axis of change used in the experiment. We see that the CIE daylight model provides a better approximation of the illuminations in the conditions where the reference illumination is parameterized to fall on the Planckian locus (neutral, blue, and yellow conditions).