Color constancy supports cross-illumination color selection

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

Nicolas P. Cottaris Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

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

We rely on color to select objects as the targets of our actions (e.g., the freshest fish, the ripest fruit). To be useful for selection, color must provide accurate guidance about object identity across changes in illumination. Although the visual system partially stabilizes object color appearance across illumination changes, how such color constancy supports object selection is not understood. To study how constancy operates in real-life tasks, we developed a novel paradigm in which subjects selected which of two test objects presented under a test illumination appeared closer in color to a target object presented under a standard illumination. From subjects' choices, we inferred a selection-based match for the target via a variant of maximum likelihood difference scaling, and used it to quantify constancy. Selection-based constancy was good when measured using naturalistic stimuli, but was dramatically reduced when the stimuli were simplified, indicating that a naturalistic stimulus context is critical for good constancy. Overall, our results suggest that color supports accurate object selection across illumination changes when both stimuli and task match how color is used in real life. We compared our selection-based constancy results with data obtained using a classic asymmetric matching task and found that the adjustment-based matches predicted selection well for our stimuli and instructions, indicating that the appearance literature provides useful guidance for the emerging study of constancy in natural tasks.

Introduction

Color helps us identify objects and informs us about object properties; this, in turn, guides our actions. In everyday life we rely on color to make decisions, such as which fruit to pick in the market or whether the person we are talking to is blushing.

For color to provide reliable guidance about object properties, our perception of an object's color needs to remain relatively constant across different contexts in which the object is viewed. Achieving such *color constancy* presents a considerable computational challenge to the visual system, because the light reflected from an object to the eye depends not only on the object's physical characteristics (spectral reflectance), but also on the spectrum of the incident illumination and other aspects of the viewing conditions (e.g., the object's pose). These can vary widely across natural contexts in which objects are viewed. Understanding the degree to which and how the visual system compensates for variation in viewing context to stabilize object color is central to understanding color perception.

Although color constancy has been the topic of extensive research for more than a century, it has rarely been studied in the ways it is typically used in real-life situations—using both naturalistic stimuli and naturalistic tasks. Indeed, although there have been a variety of studies of object color constancy that used rich and nearly natural stimuli (Brainard, Brunt, & Speigle, 1997; Boyaci, Doerschner, & Maloney, 2004; Mizokami, Ikeda, & Shinoda, 2004; Granzier & Gegenfurtner, 2012), the majority of these studies employed tasks in which the subject adjusted the color of a test object to match either a neutral internal standard (achromatic adjustment) or a reference object presented under different illumination (asymmetric matching). In everyday life, however, we rarely adjust object color; what we typically do is use color to select objects as targets of our actions. Little is known about

Citation: Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2015). Color constancy supports cross-illumination color selection. *Journal of Vision, 15*(6):13, 1–19. http://www.journalofvision.org/content/15/6/13, doi: 10.1167/15.6.13.

doi: 10.1167/15.6.13 Received November 17, 2014; published May 18, 2015

Figure 1. Stimuli for the color selection task (Experiment 1; right-eye images). The button in the center of the right side of the cube is the target and the two buttons on the left side are the tests. In the illuminant-constant condition (left) both sides of the cube were under the standard illumination. In tthe illuminant-changed conditions (center and right), the right side of the cube was under the standard illumination, but the illumination of the left side differed and could be either yellowish (center) or bluish (right). In the illuminant-changed examples shown here, the top test is the tristimulus match for the target and the bottom test is the reflectance match. In the illuminant-constant condition (left), the bottom test is both a tristimulus match and a reflectance match to the target, while the top test differs.

how color is used to support object selection across different viewing contexts (but see Bramwell & Hurlbert, 1996; Robilotto & Zaidi, 2004, 2006; Zaidi & Bostic, 2008). Furthermore, it is not clear how the degree of naturalness of the stimuli affects constancy. Studies that measure constancy across variations in stimulus richness while controlling for the low-level properties of the stimulus are rare, and consistent conclusions have not yet emerged (e.g., Hedrich, Bloj, & Ruppertsberg, 2009; de Almeida, Fiadeiro, & Nascimento, 2010).

To study constancy using a naturalistic task, we developed a selection paradigm in which, as in many everyday situations, the subject selects objects based on their color. We employed our color-selection task using simulated naturalistic scenes produced via physically-based rendering software. The stimulus design was inspired by the color cube illusion of Lotto and Purves (1999). The target objects were embedded in a multifaceted cube suspended in midair in a room in which the illumination, coming from multiple light sources, varied spatially. The subjects were instructed to "choose the test object that is closest in color to the target" and made their selections over a series of trials in which the illumination varied across the test and target locations. To produce an impression of depth in the scene, the stimuli were presented stereoscopically.

We show that the color-based selection performance revealed by our task in conjunction with naturalistic stimuli is moderately good (Experiment 1). We quantify this by developing a selection-based constancy index. The stimulus naturalness, however, was a key factor: when stimuli were simplified and consisted of flat patches embedded in a textured color background across which a simulated illumination varied (Experiment 2), selection-based constancy was significantly reduced. This result held even though the two stimulus sets were closely matched in their colorimetric properties.

To connect our current data to the large extant literature on color constancy, we also used asymmetric matching to measure the stability of color appearance across illumination changes. We did this with the stimulus sets closely matched to the two that we used with our more natural selection task. We find that the cross-illuminant shift in object appearance measured via a classic adjustment task is consistent with our color selection results.

Our results show that a reasonable degree of color constancy operates effectively in support of object selection. That is, constancy emerges as a natural feature of the results when we measure how subjects select objects in a task that models the real-life use of object color.

Experiment 1

Methods

Stimulus

On each trial, the subjects saw a large cube in the center of the screen (Figure 1). Three visible sides of the cube were covered with a checkerboard-like pattern. The cube had three distinctive buttons on it, one on the right and two on the left. The button on the right was the target, and varied from trial to trial. The two buttons on the left were the tests, with the degree of similarity to target varying across trials. The subject's task was to "choose the test button that is closest in color to the target." In illluminant-constant trials, which served as a control, both sides of the cube were rendered under the same standard illumination (6500 K daylight). In illuminant-changed trials, the right side of

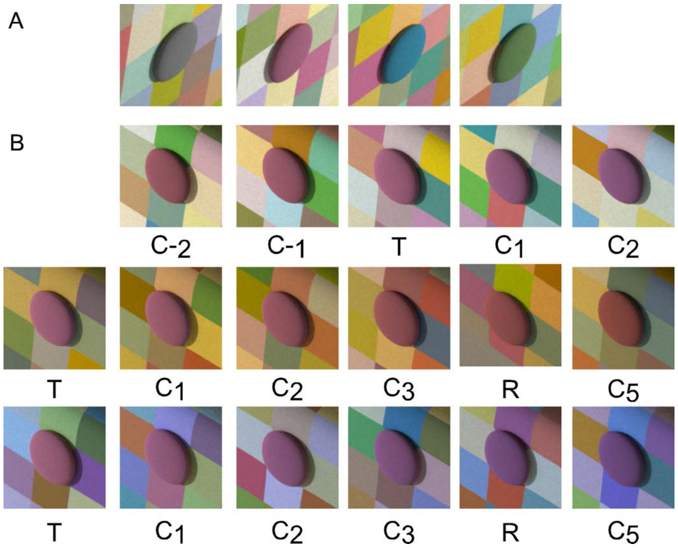

Figure 2. Targets and competitor sets (Experiment 1). (A) Four colored targets used in the experiment are shown under the standard illumination. (B) Competitor set for one of the targets in the illuminant-constant (first row) and two illumination-changed conditions (second row: yellowish; third row: bluish). The competitors always included the tristimulus match for the target (T). In the illuminant-changed conditions they also included a reflectance match (R), an overconstancy sample (C_5), and three color samples equally spaced between the tristimulus and the reflectance match in CIELAB space (C_1, C_2, C_3). Note that the actual sample denoted by C_1 is different for the yellowish and bluish competitor sets. For the illuminant-constant condition, we use the convention that samples from the yellowish competitor set are denoted by C_{-1} and C_{-2} while samples from the bluish competitor set are denoted by C_1 and C_2.

the cube was rendered under the standard illumination, but the illumination of the left side was either yellowish or bluish (4500 K and 12,000 K, respectively).

Targets

We used four colored targets (Figure 2A; see online supplement for colorimetric specification). These were chosen manually, subject to the constraints that all had the same luminance, that one target was essentially achromatic (CIELAB chroma = 0.4; CIELAB hue angle 314.01°) and that the remaining three targets were equal in CIELAB chroma ($\approx$ 25.6) but differed in CIELAB hue angle (rose, 355.31°; teal, 227.14°; green, 135.46°).

Competitor sets

On each trial, the two tests were drawn from a set of competitors, which was predefined for each target and each illuminant condition. In the illuminant-changed condition, the competitor set included six color samples (see Figure 2B):

1. A tristimulus match for the target, which had the same tristimulus values as the target under the standard illuminant (denoted as T).
2. A reflectance match for the target (denoted as R), which had the same surface reflectance as the target, but different tristimulus values when it was rendered under the test illuminant.

3–5. Three color samples that were linearly spaced between the tristimulus and the reflectance match in CIELAB space (C_1, C_2, C_3).

6. An overconstancy sample (C_5), equally distant in CIELAB from the reflectance match as competitor C_3, but displaced in the opposite direction in color space.

In the illuminant-constant condition, the competitor set included five color samples—the tristimulus match for the target and the two closest competitors (C_1 and C_2) from both the yellowish and bluish illuminant competitor sets. Across trials, each of the four targets was shown with all pairwise combinations of its competitors (15 possible combinations in the illuminant-changed conditions and 10 in the illuminant-constant condition).

Simulated illuminants

The intensities of the standard and test illuminants were adjusted for each condition to obtain desired chromaticity and luminance values for the illumination impinging on the target and test locations. This was achieved by rendering the stimulus cube and its buttons with a perfectly reflective surface specified at all locations, extracting the average reflected spectrum from a rectangle contained within the target and test buttons, and converting this to chromaticity and luminance. There was some variability across the three locations for each condition, just as there would have been when viewing real objects under natural illumination.

To find the reflectance match for each target under a test illuminant, we used the XYZ coordinates of the target together with the spectrum of the standard illuminant (impinging at the target location) to derive a surface reflectance function associated with the target. This was done via standard colorimetric methods (Brainard & Stockman, 2010), using a three-dimensional linear model for surface reflectance to constrain the derived reflectance. The linear model was derived from analysis of the spectra of Munsell papers, using the tabulated spectral data reported by Nickerson (1957) and may be found in the first three columns of the matrix B_nickerson provided with the Psychtoolbox distribution. We then found the reflectance match

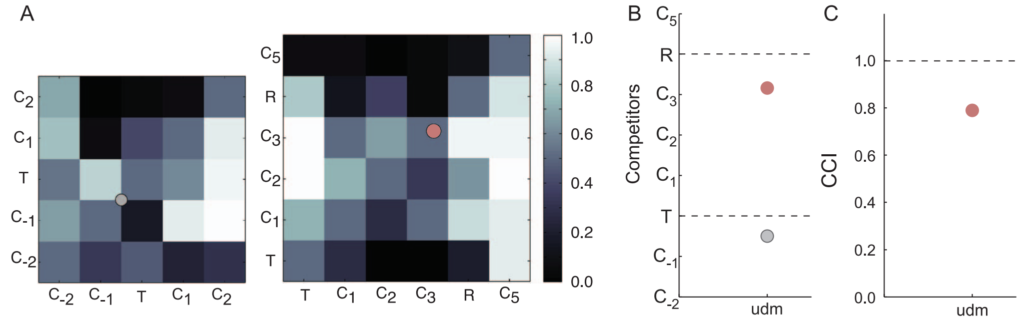

Figure 3. Using selection data to quantify constancy. (A) Choice matrices of one subject (udm) for one target (rose) in the illuminant-constant condition (left matrix) and the bluish illuminant-changed condition (right matrix). For each pairwise comparison of competitors, the matrix shows the proportion of time one competitor (vertical axis) is chosen relative to the other (horizontal axis). The circles in each matrix show the position of the subject's selection-based matches inferred from these responses. (B) Panel B plots the position of these selection-based matches relative to the competitor set. (C) Panel C plots the color constancy index derived from the selection-based match in the bluish illuminant-changed condition. Figure A2 in the Appendix illustrates the quality of the MLDS model underlying the inferred selection-based matches based on choice matrices shown in panel A.

for each target by computing the XYZ coordinates of the derived target reflectance under the test illuminant.

Asymmetric matching experiment

The stimulus configuration in the asymmetric color-matching task was nearly identical to that of the color selection task. Details on stimulus context, geometry, and experimental procedures for both experiments are provided in the Appendix.

Apparatus

The subjects viewed the stimuli stereoscopically on a stereo rig consisting of a pair of calibrated 24 in. NEC MultiSync PA241W LCD color monitors driven at a pixel resolution of 1920×1200, a refresh rate of 60 Hz, and with 8-bit resolution for each RGB channel, via a dual-port video card (NVIDIA GeForce GT120). The subject viewed the displays through two rectangular apertures (2.7×2.5 cm) in a black metal plate. The horizontal distance between the centers of the two apertures was 6.4 cm. The position of the apertures relative to the screens was such that the left screen was visible only to the left eye and the right screen was visible only to the right eye. The optical distance of each monitor to the eye was 76.4 cm. Further details about the apparatus and how it was aligned are described by Lee and Brainard (2014).

The host computer was an Apple Macintosh with an Intel Xeon quad-core processor. The experimental programs were written in MATLAB (MathWorks, Natick, MA), using routines from Psychtoolbox (Brainard, 1997; Pelli, 1997; http://psychtoolbox.org) and mgl (http://justingardner.net/doku.php/mgl/overview).

Subjects

Four subjects completed the experiment (one male, three female; ages 20–24 years). They all had normal color vision, as assessed by the Ishihara plates (Ishihara, 1977) up to one plate incorrect, normal or corrected-to-normal visual acuity (20/40 or better, as assessed by a Snellen chart), and good stereovision as assessed via the procedure described by Lee and Brainard (2014). One subject (aab) was excluded from Experiment 1 after completing the color selection task due to large proportion of missed trials (see Appendix). The research protocol has been approved by University of Pennsylvania Institutional Review Board and is in accordance with the World Medical Association Helsinki Declaration.

Analysis: From selection data to color constancy index via MLDS-based analysis

For each experimental condition, we recorded the subject's choices across a series of trials in which each target was presented with all pairwise comparisons of its competitors. The data for one target can be represented in the form of a choice matrix. For example, for one of our subjects (udm) and one target ("rose"), Figure 3A shows choice matrices for the illuminant-constant condition (left matrix) and for the bluish illuminant-changed condition (right matrix). For each pairwise comparison of competitors, the choice matrix shows the proportion of times one competitor (vertical axis) was chosen relative to the other (horizontal axis). For example, in the illuminant-constant condition, sample C_{-1} was chosen 100% of the time when paired with the sample C_2. Note that the data above and below the diagonal in the matrices shown are not independent: entries above the diagonal are obtained as 1.0 minus the corresponding entry below the diagonal. Cells along the diagonal were set to 0.5 because we did not ask subjects to select between two identical tests.

We developed an analysis method that allowed us to use the pattern of subject choices represented in the choice matrix to quantify the degree of constancy that characterizes performance in the color selection task. Our data analysis method is based on the observer model incorporated in maximum likelihood difference scaling (MLDS; Maloney & Yang, 2003; Knoblauch & Maloney, 2012). The main assumption of this model is that each stimulus (the target and the competitors) occupies a certain position in a one-dimensional perceptual space. Because the stimulus representation is subject to perceptual noise, the location of each stimulus in this space can be conceptualized as a noisy Gaussian distribution (of fixed standard deviation) centered on the "true" location. On each trial of our experiment, the representation of each stimulus (target and competitors) can be thought of as a draw from such a distribution. The subject compares the draw from the target's distribution (i.e., the representation of the target on that trial) to the draws from the distributions of the two competitors (i.e., the representations of the two competitors on that trial), and chooses the competitor whose representation on that trial is closest to that of the target.

We used numerical search to infer the location of the target and each of the competitors in the perceptual space that would best account for the subjects' choices. For each iteration, the search yielded a set of positions for the target and each of the competitors. We then simulated the pattern of subjects' choices that would correspond to these positions and compared the choice probabilities predicted in this way to those we obtained in the experiment. Across iterations, we searched for the set of target and competitor positions that maximized the log-likelihood of the measured pattern of choices. The solution yields the recovered positions of the target and each competitor along the assumed internal perceptual representation. These recovered positions corresponded to the means of the Gaussian distributions we referred to above. We can examine the quality of the solution by comparing the predicted selection probabilities, which correspond to the recovered positions, to probabilities measured for each presented pair of competitors (see Figure A2 in the Appendix).

In our implementation of the analysis, we defined the position of the first competitor as 0 (zero) and the standard deviation of the noise, which was fixed across targets and competitors, as 0.1. To simplify the calculation, we used a single Gaussian to model the aggregated effects of all noise sources. A draw from this single Gaussian was applied to the noise free difference of the distance from each competitor to the target. The scaling choices of 0 and 0.1 were arbitrary and without loss of generality: they determine the origin and scale of the recovered perceptual representations. We also constrained the positions of the competitors to be ordered in the same way in the solution as they were in the CIELAB stimulus space, and we imposed a minimum distance of 0.025 (one fourth of the standard deviation) between adjacent solution positions of the competitors. Our analysis code is provided as part of the online supplement, along with a demonstration program that produces the solution for the data of Figure 3A (right panel).

Our analysis provides the relative position of the target and competitors in a common perceptual representation, even for cases when, experimentally, the target and competitors were seen under different illumination. The position of the target in the recovered perceptual representation defines *the selection-based match*. For a given target and illuminant condition, the selection-based match should be the color sample that would be chosen as "closest to the target in color" on the majority of trials when paired with any other competitor.

When the selection-based match falls within the range of competitors, we assume that the inferred relative distances between the match and the two closest competitors will be preserved and we compute the CIELAB coordinates of the match by interpolating between the CIELAB coordinates of the two competitors adjacent to the match in the recovered representation.

When both the target and the tests are under the same illumination, the selection-based match is expected to fall at the location of the tristimulus match. The inferred selection-based match for the illuminant-constant condition is shown in both Figure 3A (left matrix) and Figure 3B as a gray circle. Both representations show that the inferred selection-based match is close to the tristimulus match. This is also true for all subjects, as shown in Figure 4A. Indeed, the average distance in CIELAB ΔE units between the selection-based match and the tristimulus match in the illuminant-constant condition is small for all targets and subjects (1.9–3.1 ΔE units; see Figure A5 in the Appendix). The fact that the inferred selection-based matches are close to the tristimulus match provides a validation of our general method.

A similar analysis can be applied to the data from the illuminant-changed condition. In this case, the proximity of the selection-based match to the reflectance match indicates the degree of subject's color constancy. The selection-based match would fall at the location of reflectance match for a subject with perfect constancy but at the tristimulus match for a subject with no constancy. We can thus quantify constancy in

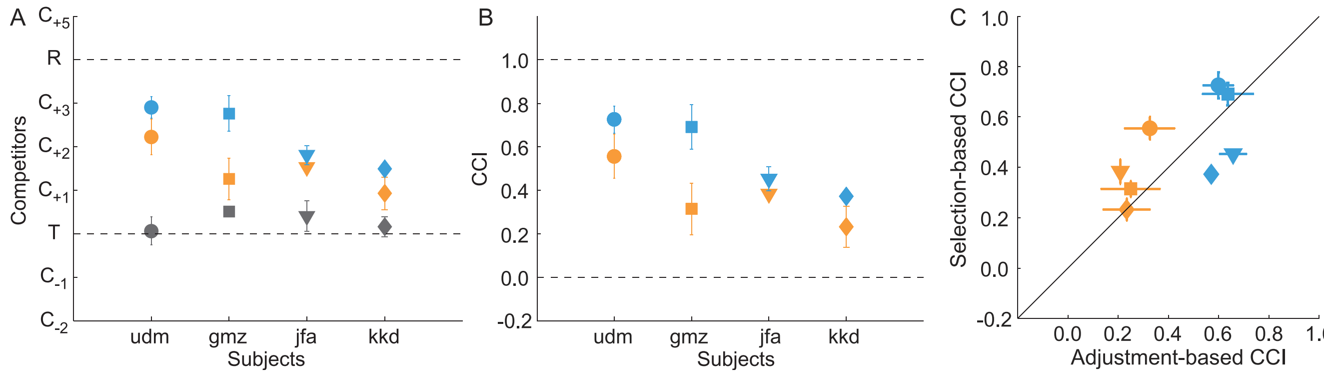

Figure 4. Results of Experiment 1. (A) Selection-based matches for four subjects. Mean selection-based match, taken across targets, is shown for each condition (gray points, illuminant-constant; orange points; illuminant-changed yellowish; blue points, illuminant-changed bluish). The y-axis is labeled according to conventions used in Figure 2. (B) Mean color constancy indices, taken across targets, for each subject. (C) Mean color constancy indices taken over targets obtained from the selection task plotted against those from the adjustment task (orange symbols plot yellowish and blue symbols plot bluish illuminant change). A different symbol is used for each of the four subjects. Error bars plot ± 1 *SEM*.

the illuminant-changed conditions by computing a color constancy index for the selection-based match (Figure 3C; see Appendix for details about constancy index computation). Our index takes on a value of 1 for a subject whose selections reveal perfect constancy (inferred match equals reflectance match) and a value of 0 for a subject whose selections reveal no constancy (inferred match equals tristimulus match).

Note that, in some cases, the inferred selection-based match falls outside of the range defined by the competitor set. This could happen, for example, if the tristimulus match is consistently chosen as closer to the target when paired with any of the other competitors. In these cases, the exact position of the selection-based match is not well-constrained by the data. Rather than leaving such a match out of the analysis altogether, we chose to assign them a nominal value that qualitatively captures the underlying pattern of choices. In particular, a selection-based match that was outside the range on the tristimulus match end of the competitor set was assigned CIELAB coordinates along the line between the tristimulus match and C_1, displaced from the tristimulus match away from C_1 by one tenth of the CIELAB distance between T and C_1. This position would correspond to a selection-based color constancy close to zero (–0.025). This happened only once in Experiment 1 (subject kkd, rose target under yellowish-illuminant change) and more frequently in Experiment 2 (subject iul, two targets; subject nau, two targets; subject sbf, one target; subject hfe, three targets; each value out of eight possibilities [4 targets $\times$ 2 test illuminants] for each subject). Following the same logic, we could assign the position to the out-of-range match that falls beyond the competitor C_6 (in Experiment 1) or C_5 (in Experiment 2); however, this never occurred for our data set. Also, there were no cases of inferred out-of-range selection-based matches for the illuminant-constant conditions.

Results

For each target and condition we inferred the selection-based match and computed constancy indices for the illuminant-changed conditions. Figure 4A plots mean position of each selection-based match, averaged over targets, for our four subjects. Figure 4B shows the corresponding constancy indices, which vary between 0.23 and 0.73, with a mean of 0.47.

Constancy was moderately high for all our subjects and did not differ significantly across the two test illuminants or different targets. A repeated measures analysis of variance (ANOVA) with illuminant (bluish vs. yellowish) and target (four levels) as within-subject factors failed to reveal significant main effect of illuminant, $F(1, 9) = 8.24$, $p = 0.07$; or target, $F(3, 9) < 1$, *ns*; or an illuminant $\times$ target interaction, $F(3, 9) = 1.05$, $p = 0.42$. There was individual variation in the overall level of constancy, with subject kkd showing the lowest constancy. In addition, two subjects showed somewhat lower constancy for the yellow than for the blue illuminant change.

Is the level of constancy we measured specific to our selection task? To address this question, we also measured constancy using adjustment-based asymmetric matching. The subject's task was to adjust a test button on the left side of the cube to match a target centered on the right side. On each trial, the position of the test button (upper or lower) was randomly chosen and set to either black or white at the beginning of the trial while the other button was

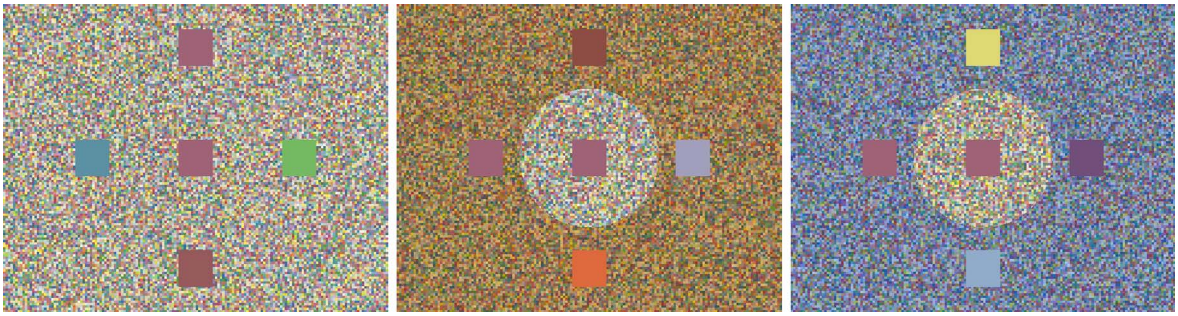

Figure 5. Stimuli for the color selection task (Experiment 2). The target square is shown in the center of the screen, surrounded by four test squares. Two of the test squares are the competitors; the other two are dissimilar distractors. In the illuminant-constant condition (left panel) the background behind the target and the test squares was under the same simulated illumination (6500 K). In illuminant-changed condition, the simulated illumination behind the test squares changed to either yellowish (4500 K; center panel) or bluish (12,000 K; right panel). In the illuminant-changed examples shown here, the left test squares are the tristimulus match; the reflectance match for the target is the top test in the center panel and right test in the right panel. In the illuminant-constant condition (left), the top test is both a tristimulus match and a reflectance match to the target.

assigned a randomly chosen Munsell paper reflectance. The stimuli were closely matched to those we employed in our selection task. For each target and condition, we computed the mean adjustment-based match for each subject (averaged across six repetitions) and used these mean matches to compute color-constancy indices in the illuminant-changed conditions (see Appendix).

For each subject and illuminant change, Figure 4C plots the mean color constancy indices obtained from the adjustment task against those obtained from color selection. The data points in the figure group close to the diagonal (the identity line) indicating reasonable agreement between the tasks. A repeated-measures ANOVA with task (selection vs. adjustment), illuminant (bluish vs. yellowish), and target (four levels) as within-subject factors failed to reveal a main effect of task, $F(1, 9) < 1$, *ns*; task × illuminant, $F(1, 9) = 4.81$, $p = 0.12$; and task × target or task × illuminant × target interaction, both $F(3, 9) < 1$, *ns*. There was a small main effect of target on constancy for the asymmetric matching task, ANOVA for asymmetric matching data: main effect of illuminant, $F(1, 9) = 93.9$, $p < 0.005$; main effect of target, $F(3, 9) = 6.24$, $p < 0.05$; and no target × illuminant interaction, $F(3, 9) < 1$, *ns*.

There was also a significant main effect of illuminant in the asymmetric matching task that is sizeable enough to be readily apparent in the data (Figure 4C). This main effect was not present for the selection task. As noted above, the illuminant × task interaction was not significant. Even had it been significant, we would be reluctant to interpret the effect as sufficient to reject the idea that selection- and adjustment-based constancy differ. First, we did not find main effects of illuminant with either task (nor an interaction) in Experiment 2, in which we studied constancy using both tasks and a different set of stimuli. Second, as we return to in the Discussion, our current selection design imposes constraints on the inferred selection-based match which may artificially reduce observed agreement between the two tasks: the level of agreement we observe is a lower bound.

To explore whether the degree of constancy we found in Experiment 1 depends on the naturalness of our stimuli, in Experiment 2 we studied color selection and its relationship to asymmetric matching using simplified stimuli.

Experiment 2

Methods

The methods were broadly the same as those of Experiment 1, except for the differences described below.

Stimuli

On each trial, the subject saw a square color target surrounded by four test squares (Figure 5). The target and the tests were presented against a textured color background, which consisted of a simulation of small, flat, matte Munsell papers. Two of the squares were dissimilar distractors whose color was chosen randomly on each trial. The remaining two squares were the competitors, sampled from a competitor set for a given target and illuminant. As in Experiment 1, in the illuminant-constant condition the simulated illumination of the background behind the target and the test squares was the same. In the illuminant-changed condition, the simulated illuminant of the background behind the tests changed to either yellowish or bluish.

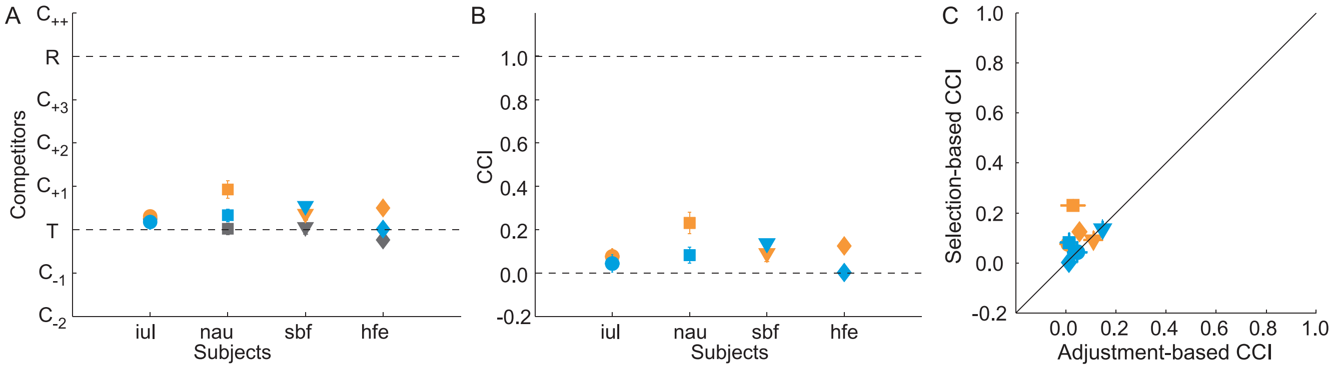

Figure 6. Results of Experiment 2. (A) Mean selection-based matches in the color selection task using simple stimuli averaged across targets. Where not visible, gray symbols lie beneath the corresponding blue symbols. (B) Mean color constancy indices computed from the selection-based matches in CIELAB space (averaged across targets). (C) Mean color constancy indices obtained from the selection task are plotted against those from the adjustment task. The figure follows the same conventions of data representation as Figure 4.

Targets

The targets were closely matched to those in Experiment 1, up to small numerical differences arising from differences in the rendering methods.

Competitor sets

The competitor sets in Experiment 2 were defined in the same way as in Experiment 1, except that they did not include an overconstancy sample in the illuminant-changed conditions (Appendix, Figure A1). In Experiment 2 we also used a predefined set of distractors, which were color samples highly dissimilar from the target. These were chosen from a subset of Munsell samples used to create textured background under the standard illuminant and included only samples whose distance in CIELAB space from any of the competitors for a given target was larger than 20 ΔE units.

Simulated illuminants

The chromaticities and luminances of the three illuminants at the test and target locations were closely matched to those of Experiment 1. Table A1 in the Appendix provides illuminant chromaticity and intensity for each condition.

Apparatus

The subjects viewed the stimuli monocularly (right eye only, with left eye occluded by an eye patch) on a calibrated 21 in. CRT color monitor (ViewSonic, Model Graphic Series G225fB) driven at a pixel resolution of 1280 × 1024, a refresh rate of 75 Hz, and with 8-bit resolution for each RGB channel via a dual-port video card (NVIDIA GeForce GT 120). An EyeLink 1000 eye tracker (desktop remote model; SR Research, Ottawa, ON, Canada) was used to record the position of the eye, although eye-tracking data are not considered here. The subject's head position was stabilized using a chin rest. The distance between the subject's eye and the center of the screen was 76 cm.

Subjects

Four different subjects participated in Experiment 2 (one male, three female; ages 20–21 years). All had normal color vision and normal or corrected-to-normal visual acuity.

Online supplement

For both experiments, instructions verbatim, detailed colorimetric specification of the stimuli (including the CIELAB, xyY, and LMS values for each competitor set, derived target reflectances and the spectrum of each illuminant), and individual subjects' data are available in the online supplement (http://color.psych.upenn.edu/supplements/colorselection1/).

Results

As with our naturalistic stimuli, the pattern of subjects' choices in the illuminant-constant condition yielded selection-based matches that approximated tristimulus matches (Figure 6A): average distance in ΔE units varied from 0.8 to 1.1 across subjects (Figure A5).

The pattern of results in the illuminant-changed conditions, however, differed considerably from that found with naturalistic stimuli. When the stimuli were simplified, constancy was low when measured via our selection task for all subjects and did not vary significantly across targets or test illuminants (Figure

6B; the main effect of illuminant, $F(1, 9) = 2.13$, $p = 0.24$; the main effect of target: $F(3, 9) < 1$, *ns*; the target × illuminant interaction $F(3, 9) < 1$, *ns*).

As in Experiment 1, we also measured constancy for a nearly identical stimulus context using asymmetric matching and found similarly low constancy for all subjects across both illuminants and targets (the main effect of illuminant, $F(1, 9) < 1$, *ns*; the main effect of target, $F(3, 9) = 2.26$, $p = 0.15$; the target × illuminant interaction, $F(3, 9) = 2.13$, $p = 0.17$).

When the color stimuli are simplified, constancy as measured by color selection and asymmetric matching were in good agreement (Figure 6C). A repeated measures ANOVA with the task, illuminant, and target as within-subjects factors did not reveal a main effect of task, $F(1, 9) = 2.01$, $p = 0.25$, or any interaction for task × illuminant, $F(3, 9) = 5.6$, $p = 0.1$; task × target, $F(3, 9) < 1$, *ns*; or task × target × illuminant, $F(3, 9) = 1.2$, $p = 0.38$.

Discussion

One of the key functions of color in real-life situations is to guide object selection. To understand how color perception and color constancy support object selection, we developed a task in which, as in real life, subjects select objects based on color across an illumination change. Based on subjects' selections, we quantified constancy and expressed it in terms commensurate with those typically used in adjustment-based color appearance studies—a selection-based match (conceptually thought of as the cross-illumination point of subjective equality) and a selection-based color constancy index. Our results show that, when the stimulus context is nearly naturalistic, the selection data reveal moderately high constancy, as shown by the selection-based constancy indices. However, the degree of selection-based constancy depends strongly on the choice of stimuli: when the stimuli are simplified, selection-based constancy was greatly reduced and close to zero for most of our subjects. This result holds even though the colorimetric characteristics of the stimuli were closely matched across experiments.

Our study is not the first to use selection to study constancy. We build on a series of papers from Zaidi and collaborators, who developed a cross-illumination odd-one-out paradigm and implemented it using real objects and illuminants (Robilotto & Zaidi, 2004, 2006; Zaidi & Bostic, 2008; see also Bramwell & Hurlbert, 1996). In their task, the subjects were shown two pairs of objects, each under a different illumination, and asked to indicate "which of the four [objects] is of a different material from the other three." On each trial of their experiment, three of the objects were identical and served as the standard (equivalent to our target object). One of the objects differed and, similarly to our competitors, its degree of similarity relative to the standard varied across trials. Consistent with our findings, as well as with findings from the color appearance literature (Foster, 2011; Brainard & Radonjić, 2014), Zaidi and colleagues show that selection-based constancy is not perfect: when the illuminant changes, subjects often chose the reflectance match as the "odd object" in favor of a test that deviated from the reflectance match in the direction of the tristimulus match. An advance between our design and that of these seminal studies is in how the test objects were chosen. In our design, tests are chosen pairwise from a competitor set that spans the region of color space between the tristimulus and reflectance matches; in the studies of Zaidi and colleagues, one of the tests was always the reflectance match. This restriction reduces the power of the resulting data with respect to inferring a selection-based match.

We were also interested in understanding whether the degree of selection-based constancy depended on the naturalness of the stimulus context. A systematic characterization of how color constancy varies with the degree of stimulus naturalness is challenging, because a definition of naturalness remains elusive. For example naturalness may be related to scene articulation, which is often operationalized as the number of elements in the scene. It has been found that increasing this type of articulation sometimes leads to higher constancy and sometimes does not (Kraft, Maloney, & Brainard, 2002; Radonjić & Gilchrist, 2013). Naturalness may also be related to whether the scene is two- or three-dimensional. Here too, findings remain inconclusive: while adding depth information to two-dimensional stimulus scenes led to improved constancy in some studies (Yang & Shevell, 2002; Hedrich, Bloj, & Ruppertsberg, 2009), others report little or no effect (Kraft, Maloney, & Brainard, 2002; de Almeida, Fiadeiro, & Nascimento, 2010).

Rather than attempting to define dimensions along which naturalness varies, we chose to study two configurations that we judged differed considerably in how natural they appeared: (a) a three-dimensional scene rendered using physically based methods and viewed stereoscopically (Experiment 1), and a two-dimensional scene in which the targets and tests were displayed as square patches presented against a textured background (Experiment 2). To experience the substantial difference in naturalness of the two sets of stimuli, it is sufficient to compare Figures 1 and 5, for which the colorimetric properties are closely matched. Perceiving stimuli as objects embedded in the scene in which the illumination spatially varies feels natural and effortless for Figure 1, but seems artificial for Figure 5. Although our approach does not allow us to isolate factors that contribute to stimulus naturalness, it did allow us to demonstrate a dramatic difference in constancy. Thus, some factor that is captured by the stimuli used in Experiment 1 but not by those used in Experiment 2 is important for human color constancy. This factor is not the choice of illuminant or surface spectra used in the two experiments, as these were well matched in our study.

To connect our selection-based measures to the classic color appearance literature in which constancy is typically studied using adjustment methods, we compared selection-based constancy with constancy measured in equivalent stimulus contexts using asymmetric matching. In both experiments, the constancy indices obtained via selection and matching were generally in good agreement. This suggests that as we build an understanding of how color is used for selection, the extant constancy literature, which is based largely on adjustment procedures, can provide useful guidance. We note that although we found no main effects of task on constancy, we did find a difference in main effect of illuminant across the two tasks in Experiment 1, albeit without a significant interaction between task and illuminant-change direction. Although it is possible that this trend toward a significant interaction reflects a real difference in the way the two tasks are performed, it is premature to draw that conclusion; a key reason for this is a limitation of our current selection method. The position of the inferred selection-based match in our experiments was constrained to the line in color space determined by the locations of the predefined competitor set. In the asymmetric matching task, however, subjects were free to set their matches anywhere in the full three dimensions of color space. Thus, the reasonable agreement we find between the two tasks represents a lower bound on what we might find if we could relax the constraint on selection. We are currently developing procedures that will allow us to do so by incorporating an adaptive psychophysical method introduced recently by Jogan and Stocker (2014).

It is known that varying experimental instructions can substantially modulate the degree of constancy revealed by an asymmetric matching task (Arend & Reeves, 1986; Troost & de Weert, 1991; Bauml, 1999). In the work reported in this paper, we used what we call neutral instructions ("select the test that was closest in color to the target"; "adjust the test to match the target in color"). In our view, neutral instructions direct the subject toward relying on his or her perceptual representation of the stimuli and we believe they provide an appropriate anchor point for studies of constancy. How using different instructions (e.g., "choose/adjust the test to match the surface reflectance of the target, as if the target was under the changed illumination as well") might modulate the results remains an interesting question, which we plan to address in future work.

By developing a selection task for quantifying constancy, we have taken an important step toward studying how color constancy emerges in real-life situations. Our selection task is fairly natural. For this reason, it is intuitive for subjects who successfully engage in it after hearing only brief instructions. This provides an important advantage over adjustment methods, in which extensive training is required before subjects learn how to use the controls to obtain satisfactory matches. In addition, we have found it challenging to coach subjects toward using a consistent (across subjects) criterion of when a match is close enough. This is not an issue for our selection method. Further, employing selection provides experimental control over the time course of stimulus presentation. This is difficult to accomplish with adjustment, where how long the subject takes is under his or her control. Finally, the basic color selection task we introduce here has the feature that it can be incorporated as a component of more complex and realistic tasks that require object choice that is guided by color (Radonjić, Cottaris, & Brainard, 2015).

Keywords: color perception, color constancy, color selection, naturalistic stimuli

Acknowledgments

This work has been supported by NIH RO1 EY10016 (DHB), Core Grant P30 EY001583, and University of Pennsylvania Research Foundation Grant. We thank Kira DiClemente and James Allen for their help in data collection and Christopher Broussard for technical assistance.

Commercial relationships: none.
Corresponding author: Ana Radonjić.
Email: radonjic@sas.upenn.edu.
Address: Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA.

References

Arend, L. E., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A, 3,* 1743–1751.

Arend, L. E., Reeves, A., Schirillo, J., & Goldstein, R. (1991). Simultaneous color constancy: Papers with diverse Munsell values. *Journal of the Optical Society of America A, 8,* 661–672.

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

Boyaci, H., Doerschner, K., & Maloney, L. T. (2004). Perceived surface color in binocularly viewed scenes with two light sources differing in chromaticity. *Journal of Vision, 4*(9):1, 664–679, http://www.journalofvision.org/content/4/9/1, doi:10.1167/4.9.1. [PubMed] [Article]

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436.

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, 14,* 2091–2110.

Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. *Encyclopedia of Imaging Science and Technology* (pp. 72–188). Hoboken, NJ: Wiley.

Brainard, D. H., & Radonjić, A. (2014). Color constancy. In J. S. Werner & L. M. Chalupa (Eds.), *The new visual neurosciences* (pp. 545–556). Cambridge, MA: MIT Press.

Brainard, D. H., & Stockman, A. (2010). Colorimetry. In M. Bass et al. (Eds.), *Handbook of optics, 3rd ed., Vol. III: Vision and vision optics* (pp. 10.11–10.56). New York: McGraw Hill.

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

Chatterjee, A., Widick, P., Sternschein, R., Smith, W. B., & Bromberger, B. (2010). The assessment of art attributes. *Empirical Studies of the Arts, 28,* 207–222.

Commission Internationale de l' Éclairage (CIE). (2007). *Fundamental chromaticity diagram with physiological axes, Parts 1 and 2* (Technical Report 170-1). Vienna, Austria: Central Bureau of the Commission Internationale de l' Éclairage.

de Almeida, V. M. N., Fiadeiro, P. T., & Nascimento, S. M. (2010). Effect of scene dimensionality on color constancy with real three-dimensional scenes and objects. *Perception, 39,* 770–779.

Foster, D. H. (2011). Color constancy. *Vision Research, 51,* 674–700.

Granzier, J. J. M., & Gegenfurtner, K. R. (2012). Effects of memory colour on colour constancy for unknown coloured objects. *i-Perception, 3,* 190–215.

Griffin, L. D. (1999). Partitive mixing of images: A tool for investigating pictorial perception. *Journal of the Optical Society of America A, 16,* 2825–2835.

Heasly, B. S., Cottaris, N. P., Lichtman, D. P., Xiao, B., & Brainard, D. H. (2014). RenderToolbox3: MATLAB tools that facilitate physically based stimulus rendering for vision research. *Journal of Vision, 14*(2), 1–22, http://www.journalofvision.org/content/14/2/6, doi:10.1167/14.2.6. [PubMed] [Article]

Hedrich, M., Bloj, M., & Ruppertsberg, A. I. (2009). Color constancy improves for real 3D objects. *Journal of Vision, 9*(4):16, 1–16, http://www.journalofvision.org/content/9/4/16, doi:10.1167/9.4.16. [PubMed] [Article]

Ishihara, S. (1977). *Tests for colour-blindness.* Tokyo: Kanehara Shuppen Company, Ltd.

Jogan, M., & Stocker, A. A. (2014). A new two-alternative forced choice method for the unbiased characterization of perceptual bias and discriminability. *Journal of Vision, 14*(3):20, 1–8, http://www.journalofvision.org/content/14/3/20, doi:10.1167/14.3.20. [PubMed] [Article]

Kirby, J. R., Moore, P. J., & Schofield, N. J. (1988). Verbal and visual learning styles. *Contemporary Educational Psychology, 13,* 169–184.

Knoblauch, K., & Maloney, L. T. (2012). *Use R! series: Modeling psychophysical data in R.* New York: Springer.

Kraft, J. M., Maloney, S. I., & Brainard, D. H. (2002). Surface-illuminant ambiguity and color constancy: Effects of scene complexity and depth cues. *Perception, 31,* 247–263.

Lee, T. Y., & Brainard, D. H. (2014). The effect of photometric and geometric context on photometric and geometric lightness effects. *Journal of Vision, 14*(1):24, 1–24, http://www.journalofvision.org/content/14/1/24, doi:10.1167/14.1.24. [PubMed] [Article]

Lotto, R. B., & Purves, D. (1999). The effects of color on brightness. *Nature Neuroscience, 2,* 1010–1014.

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision, 3*(8):5, 573–585, http://www.journalofvision.org/content/3/8/5, doi:10.1167/3.8.5. [PubMed] [Article]

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64,* 17–24.

McCamy, C. S., Marcus, H., & Davidson, J. G. (1976). A color-rendition chart. *Journal of Applied Photographic Engineering, 2,* 95–99.

Mizokami, Y., Ikeda, M., & Shinoda, H. (2004). Color constancy in a photograph perceived as a three-dimensional space. *Optical Review, 11,* 288–296. doi:10.1007/s10043-004-0288-8.

Nickerson, D. (1957). Spectrophotometric data for a collection of Munsell samples. Washington, DC: U.S. Department of Agriculture.

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

Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2015). Color constancy in a naturalistic goal-directed task. Manuscript submitted for publication.

Radonjić, A., & Gilchrist, A. L. (2013). Depth effect on lightness revisited: the role of articulation, proximity and fields of illumination. *i-Perception, 4*, 437–455.

Robilotto, R., & Zaidi, Q. (2004). Limits of lightness identification for real objects under natural viewing conditions. *Journal of Vision, 4*(9):9, 779–797, http://www.journalofvision.org/content/4/9/9, doi:10.1167/4.9.9. [PubMed] [Article]

Robilotto, R., & Zaidi, Q. (2006). Lightness identification of patterned three-dimensional, real objects. *Journal of Vision, 6*(1):3, 18–36, http://www.journalofvision.org/content/6/1/3, doi:10.1167/6.1.3. [PubMed] [Article]

Stockman, A., & Sharpe, L. T. (2000). Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype. *Vision Research, 40,* 1711–1737.

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

Xiao, B., & Brainard, D. H. (2008). Surface gloss and color perception of 3D objects. *Visual Neuroscience, 25,* 371–385.

Yang, J. N., & Shevell, S. K. (2002). Stereo disparity improves color constancy. *Vision Research, 42,* 1979–1989.

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

Appendix: Supplementary methods and results

Color selection: Stimulus and task

Experiment 1

At the beginning of each trial the subjects saw a cube in the center of the screen. The cube had three distinct buttons: the target, centered in the right side, and two test buttons on the left (Figure 1). In illuminant-constant trials, two sides of the cube were under the same standard illumination (6500 K daylight). In illuminant-changed trials, only the right side of the cube was under the standard illumination, while the left side was illuminated either by the yellowish or the bluish test illuminant (CIE daylights with correlated color temperatures of 4500 K and 12,000 K, respectively; half of the trials for each changed illuminant). Table A1 provides xy chromaticities and intensity (cd/m^2) for each location and illumination condition.

For Experiment 1, we report values for illuminant spectra extracted for the target and each test location, as described in Methods. For Experiment 2, we report values for simulated illuminants, which were homogenous across locations.

On each trial the two test buttons were drawn from the set of competitors for the target and illumination condition. The subject's task was to choose the test that was the closest in color to the target. Subjects indicated their choice by using a joystick (moving it up to select the top or down to select the bottom button). After they made their choice, a black square briefly flashed in the center of the chosen button after which the next trial began.

In each experimental block, each target was presented with all pairwise combinations of competitors from the set in random order. Therefore, each illuminant-constant block consisted of 40 trials (1 standard illuminant $\times$ 4 targets $\times$ 10 competitor pairs) and each illuminant-changed block consisted of 120 illuminant-changed trials (2 test illuminants $\times$ 4 targets $\times$ 15 competitor pairs).

Experiment 2

At the beginning of each trial, the subjects saw a textured background illuminated by the standard illuminant. A fixation cross with a small dot in the middle was placed in the center of the screen. To start a trial, the subjects used the computer mouse to move a cursor (a small black dot) onto the cross to cover the small dot and then clicked the mouse. The computer then displayed five squares on the screen (each 3.5 cm on each side; 2.6°). The square in the center of the screen was the target color for the trial and was surrounded by four test squares (at 8° eccentricity, measured from the center of the target to the center of the test square).

On illuminant-constant trials, the target and the test squares were presented against the textured background uniformly illuminated by the standard illuminant (Figure 5, left panel). On illuminant-changed trials, the background behind the test squares changed to simulate a change in illumination. On half of these trials, the background behind the test squares was

Condition	x	y	Y (cd/m^2)
Experiment 1			
Illuminant-constant condition (target, 6500 K; tests, 6500 K)			
Top test	0.310	0.326	95.51
Bottom test	0.310	0.326	96.29
Target	0.310	0.326	95.39
Illuminant-changed condition: (target, 6500 K; tests, 4500 K)			
Top test	0.354	0.365	63.07
Bottom test	0.354	0.365	63.92
Target	0.313	0.329	94.69
Illuminant-changed condition: (target, 6500 K; tests, 12,000 K)			
Top test	0.270	0.280	65.52
Bottom test	0.271	0.281	66.57
Target	0.307	0.323	95.19
Experiment 2			
6500 K	0.313	0.330	95.22
4500 K	0.362	0.371	61.66
12,000 K	0.270	0.281	64.25

Table A1. Chromaticity (xy) and intensity (Y in cd/m^2) of standard and test illuminants. For Experiment 1, we report values for illuminant spectra extracted for the target and each test location, as described in Methods. For Experiment 2, we report values for simulated illuminants, which were homogeneous across locations.

under the yellowish test illuminant (Figure 5, center panel) and on the other half under the bluish test illuminant (Figure 5, right panel). On these trials, a circular area of the background around the target (13.3 cm in diameter; 10°) always remained under the standard illuminant. The chromaticities and intensities of the three illuminants at the test and target locations were closely matched to those of Experiment 1 and are also provided in Table A1.

On each trial, two of the test squares were distractors, drawn randomly (without replacement on each trial; with replacement across trials) from the distractor set for a given target color. The remaining two squares were the drawn from the competitor set (Figure A1). On each trial, the textured background for the target and the test squares was randomly chosen from a precomputed set of 10 different backgrounds for a given condition (see below for more detail on these backgrounds). As in Experiment 1, the subject's task was to choose the test square closest in color to the target. Subjects indicated their choice by moving the cursor onto the chosen test square and clicking the mouse.

Within an experimental block, each target was presented with all pairwise combinations of competitors for a total of 40 trials (1 standard illuminant $\times$ 4 targets $\times$ 10 competitor pairs) in illuminant-constant and 80 (2 test illuminants $\times$ 4 targets $\times$ 10 competitor pairs) in illuminant-changed blocks. Instructions for both experiments are provided in the online supplement (http://color.psych.upenn.edu/supplements/colorselection1/).

Color selection: Stimulus specification

Experiment 1

The geometry of the stimulus was created in the open-source modeling program, Blender (http://www.blender.org/), and rendered using the Mitsuba renderer (https://www.mitsuba-renderer.org/), using a path-tracer integrator and low discrepancy sampler (sample count 320). Rendering was facilitated using Render-Toolbox3 (Heasly, Cottaris, Lichtman, Xiao, & Brainard, 2014; https://github.com/DavidBrainard/RenderToolbox3/wiki). This allowed us to specify the spectral power distribution of each scene illuminant and the reflectance function of each surface. Spectral sampling was at 10 nm steps between 400 nm and 700 nm.

Rendering produced a 31-plane hyperspectral image of each stimulus scene. We then converted each hyperspectral image into a three-plane LMS representation by computing at each pixel the excitations that would be produced in the human L-, M-, and S-cones (via Stockman–Sharpe cone fundamentals; Stockman & Sharpe, 2000; Commission Internationale de l' Éclairage [CIE], 2007). These LMS images were converted to RGB images for presentation using monitor calibration data and standard methods (Brainard, Pelli, & Robson, 2002). Monitor calibrations measurements were made using a PR-670 SpectraScan

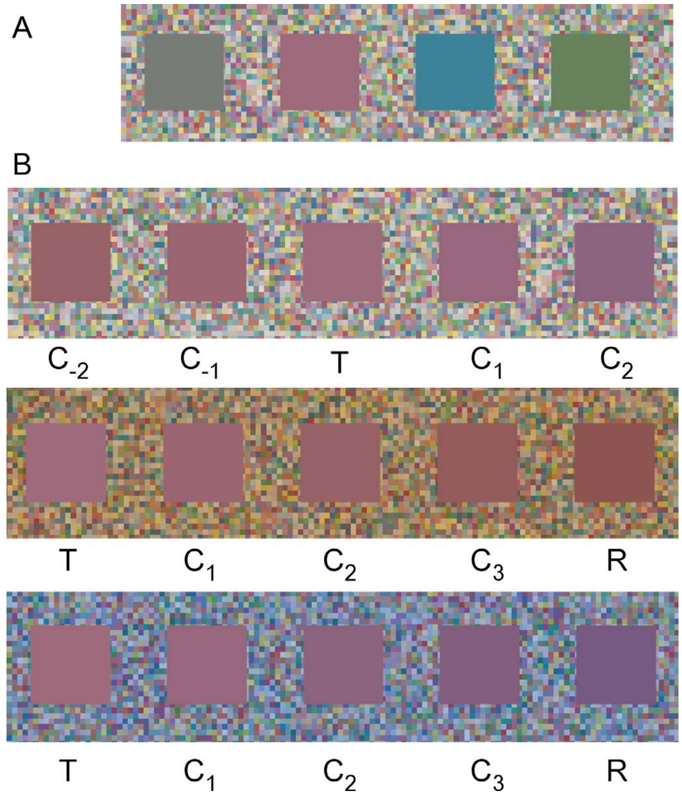

Figure A1. Targets and competitors for Experiment 2. (A) Four square colored targets shown under the standard illumination. (B) Competitor sets. In the illuminant-changed conditions (middle row: yellowish illuminant; bottom row: bluish illuminant) the competitor set includes the tristimulus match (T) and the reflectance match (R) for the target, as well as three color samples equally spaced between these in CIELAB space (C_1, C_2, C_3). In the illuminant-constant condition (top row), the competitor set includes the tristimulus match and two closest samples from the yellowish (C_{-1}, C_{-2}) and the bluish competitor set (C_1, C_2).

spectroradiometer (PhotoResearch, Chatsworth, CA) and included characterization of spectral power distribution of the monitor's primaries as well as the gamma function of each monitor channel.

We performed all data analysis using specified stimulus values. The CIE XYZ tristimulus coordinates and CIELAB values for our stimuli were obtained from the underlying hyperspectral image data. Note, however, that the CIE XYZ color-matching functions are not an exact linear transformation of the Stockman-Sharpe cone fundamentals and that the XYZ values we use differ slightly from those that would be obtained from measurement of the displayed stimuli. They match the XYZ values of the spectral stimuli for which the displayed stimuli are metamers, when the metamerism is computed with respect to the Stockman-Sharpe fundamentals.

Color selection: Stimulus context

Experiment 1

The stimulus scene consisted of a large cube located at the center of a room (Figure 1). Three sides of the cube were visible and were covered with a checkerboard-like pattern. On each trial within a block, the reflectance of the checks was randomized: each check was assigned a reflectance from a subset of 25 different samples randomly drawn (with replacement) from a subset of 203 samples chosen from a set of 462 Munsell papers whose reflectance is known (Nickerson, 1957). The subset included only the samples whose luminance under the standard illuminant was at least 20 cd/m^2. To keep the number of stimuli we needed to render manageable, however, the same arrangement of checks for each trial type was used across blocks and subjects.

The floor, the side, the back wall, and the other (not visible) sides of the cube were assigned the reflectance of the light gray square of the Macbeth color checker chart (MCC; sample from row 4, column 2, reflectances obtained at http://www.babelcolor.com and provided as part of the RenderToolbox3 distribution; McCamy, Marcus, & Davidson, 1976). The ceiling was assigned the reflectance of the white square of the MCC (sample from row 4, column 1). The front wall was a nonreflective surface, positioned approximately halfway between the cube and the eye. It contained a large window-like aperture through which the inside of the room and the cube were visible. The size of the aperture was $35.5° \times 26.6°$ of visual angle. The size of a rectangle enclosing the rendered cube was approximately $17.5° \times 21°$; the size of the target and test buttons was approximately $1.7° \times 2°$.

Two point light sources illuminated the scene—one from the left and one from the right of the cube. Each, therefore, primarily illuminated one side of the cube. Their intensities were adjusted as described above.

Each stimulus scene was rendered from two eye (camera) positions, both at 76.4 cm distance from the center of the cube, but horizontally displaced by 6.4 cm. As described in Lee and Brainard (2014), stereo pair images were rendered using the parallel ("off-axis") method. The Blender file describing the stimulus scene geometry is provided in the online supplement.

Experiment 2

The target and test stimuli were presented against a textured color background, consisting of a simulation of illuminated small paper squares. Each background consisted of 160×128 squares (0.25×0.23 cm each, subtending $0.17° \times 0.19°$ of visual angle), randomly sampled with replacement from a subset of about 220 different Munsell samples. This subset included Munsell papers (a) that could be rendered within the gamut of our display under all three simulated illuminants, and (b) whose luminance under the standard illuminant was at least 20 cd/m^2. We created 10 different background texture patterns and rendered each of them under the three simulated illuminants. As noted above, one of these backgrounds, rendered under the appropriate illuminant, was chosen at random for each trial. The mean xy chromaticity across backgrounds for the standard, bluish and yellowish test illuminant [0.33 0.34]; [0.38 0.38]; [0.29, 0.30] while the mean luminance were 45.17, 29.41, and 30.32 cd/m^2, respectively. Therefore, across all trials, the target and the competitors were always decrements relative to the background mean.

Asymmetric matching

The stimulus configurations were nearly identical to those we used for the color selection task.

Experiment 1

The subject saw a cube in the center of the screen with a target button on the right and two buttons on the left side. On each trial one of the buttons on the left was either black or white, and this was the test for the trial. The reflectance of the other button was chosen randomly from the subset of Munsell paper reflectances used for the background checks. As in the color selection task, in the illuminant-constant trials both sides of the cube were rendered under the same simulated illumination (6500 K). On illuminant-changed trials only the right side of the cube was under the standard illuminant, while the left side was under the yellowish test illuminant (4500 K) in one half and the bluish test illuminant (12,000 K) in another half of the trials. The illuminant intensities were the same as in the selection task (Table A1).

The reflectances of checks covering the sides of the cube were randomized following the same procedure as for the color selection task. A different arrangement of checks was rendered on each trial and each block but was the same across subjects.

Experiment 2

The subject saw the target square in the center of the screen surrounded by four colored squares. One of the surrounding squares was the test, set to either white or black at the beginning of the trial. The color of the remaining three squares was randomly chosen from the set of distractors for the target. On illuminant-constant trials, the target and the test were under the same simulated illuminant. On illuminant-changed trials, the background behind the target was always rendered under the standard illuminant, while the background of the test squares was under the bluish simulated illuminant on one half and under the yellowish under the other half of the trials.

In both experiments, each illuminant-constant block consisted of four trials (1 standard illuminant $\times$ 4 targets) and each illuminant-changed block consisted of eight trials (2 test illuminants $\times$ 4 targets).

In both experiments, subjects were asked use a game controller to adjust the test so that it matched the target in color. Controls allowed adjustment of test square CIELAB L*, chroma, and hue.

Asymmetric matching training

Before the first color-adjustment session, the subjects were trained how to use the controller to achieve a desired match. The training consisted of two to three one-hour sessions during which the subjects completed total of four to seven blocks of trials (up to 10 trials in a block) in which they adjusted a test to match a randomly chosen target color. In these training blocks, the target and the test were presented adjacent to one another against the illuminant-constant background. Images of training stimulus examples and training instructions are available in the online supplement. For both experiments, this training was done using the single screen apparatus of Experiment 2, viewed binocularly.

In the first training session, the experimenter made the first match, followed by the step-by-step explanation, and then helped the subject make the second match. The subject then continued to make matches unassisted. At the beginning of each subsequent training session, subjects were shown examples of some of their worst matches (assessed using the CIELAB ΔE metric) and the experimenter indicated that they should try to make their adjustments agree more with the targets. Each subject participated in either two or three training sessions before moving on to experimental sessions.

Implementation of stimulus adjustment in Experiment 1

It was not feasible to fully rerender the stimulus from the scene description after each button press in Experiment 1. Indeed, rendering each scene takes tens of minutes on our current hardware. To vary the color of the test button during the adjustment procedure, we used a variant of the partitive mixture image synthesis method introduced by Griffin (1999). A MATLAB class that supports this partitive-mixing method is provided as part of the open-source BrainardLab-Toolbox (http://github.com/DavidBrainard/BrainardLabToolbox; class PartitiveImageSynthesizer). Our method is described in detail in an earlier paper from our lab (Xiao & Brainard, 2008). Briefly, for each trial we rendered two sets of eight basis images of the stimulus scene (one set for each eye). Images within a basis set were identical except for the rendered surface reflectance of the test button. The rendered reflectances were chosen so that the average monitor RGB values of the test button included each member of the set: {[1 1 1]; [0 0 1]; [0 1 0]; [1 0 0]; [0.2 0.2 0.2]; [0 0 0.2]; [0 0.2 0]; [0.2 0 0]}. By recombining the basis images appropriately, we could quickly produce images with any desired XYZ values within the monitor gamut, while maintaining the high-fidelity of the overall rendering. Note that some of the reflectances used to render these basis images were not physically realizable.

Experimental procedures

In both experiments, two of the subjects completed the color selection task first while the other two completed the asymmetric matching task first. At the beginning of the first color selection session all subjects completed a brief training consisting of four illuminant-constant trials, each with a different target color. In the color selection part of Experiment 1, each subject completed 30 blocks of illuminant-constant trials and 30 blocks of illuminant-changed trials in the course of six sessions, each lasting about an hour. In Experiment 2 subjects completed 20 illuminant-constant and 30 (31 for subject sbf) illuminant-changed blocks of trials in seven 1-hr sessions. The sessions were blocked: the first and the fourth session consisted only of illuminant-constant trials; the second, third, fifth, and sixth sessions consisted only of illuminant-changed trials. In Experiment 2, in the final (seventh) session, the subjects completed the remaining blocks of trials from both conditions needed to finish the experiment (all illuminant-constant trials were completed first).

In the color adjustment part of the study, each subject completed six blocks of trials for each illuminant condition in three to five 1-hr sessions. The sessions were blocked by illuminant condition with the first session always being the illuminant-constant condition. In Experiment 1, the illuminant-constant session and illuminant-changed sessions generally alternated. In Experiment 2, the first illumination-constant session was followed by two illuminant-changed sessions.

At the end of the study all subjects completed a questionnaire in which they described any strategy they might have adopted for judging similarity of color in the color selection and the color adjustment tasks. They were also allowed to write down any additional comments they might have had about the experiment. All subjects in Experiment 1 also completed three additional questionnaires: the aesthetic experience questionnaire (adapted from Chatterjee, Widick, Sternschein, Smith, & Bromberger, 2010), the vividness of visual imagery test (Marks, 1973), and the visualizer-verbalizer questionnaire (Kirby, Moore, & Schofield, 1988). We do not present data from the postexperimental questionnaires here.

Excluded subject

One subject (subject aab, male, age 18) was excluded from the Experiment 1 after completing the color selection task and thus, did not participate in the color adjustment part of the study. This is because a large percent of his trials (19.27%) were completed in less than 300 ms, which we considered to be the minimal time needed to view and process the stimulus. The number of such trials for all other subjects was either zero or negligible (subject gmz, 3 out of 4,800 trials; these trials were excluded from the analysis).

We analyzed the data for subject aab after excluding all trials shorter than 300 ms and found that his data follow the same general trends we identified for other subjects. His individual data is available in the online supplement.

Examining the quality of the MLDS solutions

To examine the quality of the solution, we can compute the predicted selection probabilities corresponding to the recovered positions of the target and competitors and plot them against the selection probabilities measured in the experiment. Figure A2 (left panel) shows examples of such plots for the data shown in Figure 3A of the paper. To indicate how typical these fits are relative to all other fits in our data set, in the right panel we plot the log-likelihood of the shown solution (red line) relative to the distribution of all solution log-likelihoods for this experimental condition (across all targets and subjects). Larger log likelihoods should generally correspond to better correspondence between predicted and measured probabilities. For each experiment and subject, the online supplement shows plots of the quality of MLDS model solutions for each target and condition.

Examining the bias and reliability of the MLDS method

Our study used fewer comparison stimuli than some earlier reports that employed the general MLDS method (Knoblauch & Maloney, 2012; Maloney &

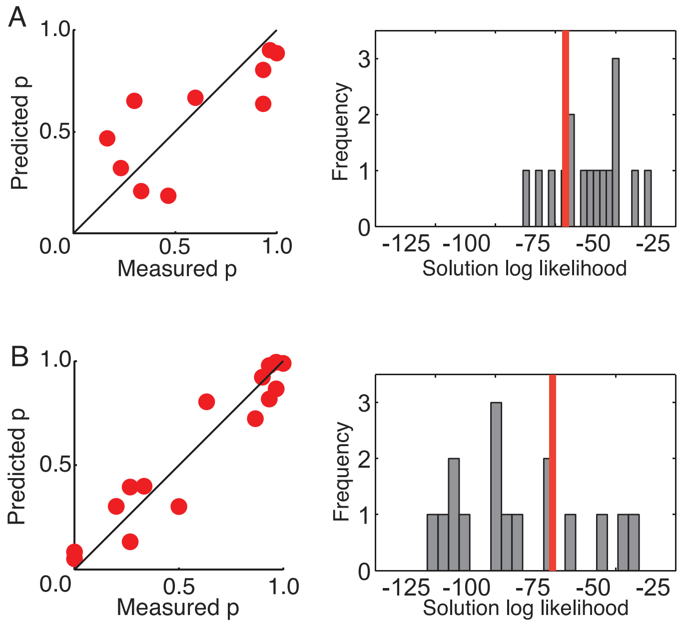

Figure A2. Quality of the MLDS model. For each pair of competitors, the left panel plots measured probabilities of the first competitor in a pair being chosen (over the second) against those predicted based on the results of our MLDS-based analysis. (A) Panel A plots the illuminant-constant and (B) panel B plots the bluish illuminant-changed condition for the pattern of responses shown in Figure 3A (subject udm; target: rose). The right panels indicate log-likelihood of the solution shown on the left (red line) relative to the distribution of all solution log-likelihoods (across all targets and subjects) for the corresponding experimental condition.

Yang, 2003). We therefore used simulation to explore the expected precision of the selection-based matches recovered with our experimental design (e.g., five competitors and a target). We simulated subject's responses for a range of different target positions along the space delineated by five competitors and then used our MLDS-based method to recover the target position from simulated responses. In the simulation, we assumed the same standard deviation of the noise as the actual data analysis (0.1) and we tested five different competitor spacing magnitudes. The spacings we chose corresponded to those inferred via our MLDS analysis from our experimental data: five competitors were linearly spaced over ranges of 0.15, 0.5, 1, 3, or 5 units. Figure A3 (panel A) shows the average inferred match for nine different target positions and five different sizes of the competitor spacings, averaged over 1,000 simulations. In general, the average accuracy of target recovery is good, but is less precise for larger competitor spacings. The nature of this effect can be seen more clearly in panel B of the figure, which shows 55 different target locations using 100 simulations, for linear spacings of 0.15 and 5. When the spacing is large, the noise is not sufficient to produce graded responses as targets move between one competitor and the next, so the target recovery error becomes systematic and approaches one half of the spacing between targets. The CIELAB spacing between our targets is perceptually quite modest, between 3.2 and 5.4 ΔE units, which means that a "worst case" recovery error of approximately 1.6 to 2.7 ΔE units (small relative to the effects of interest in the studies reported here) is represented. In addition, for the naturalistic stimuli, 80% of the recovered selection-matches corresponded to a recov-

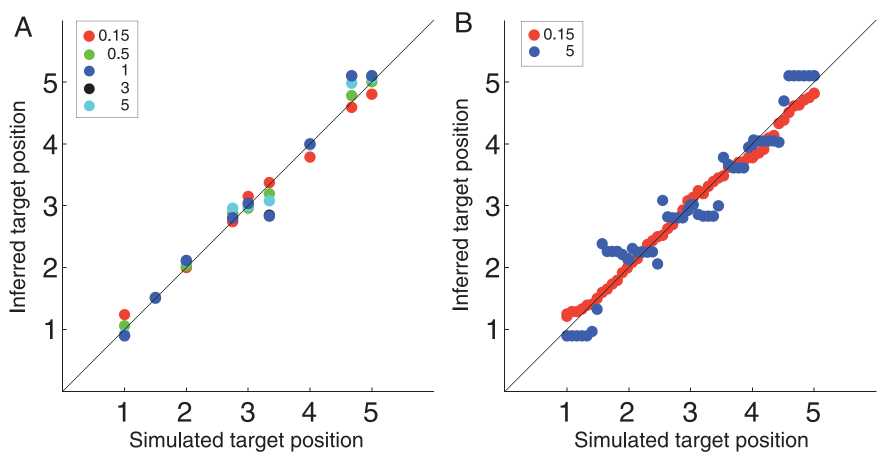

Figure A3. Recovered MLDS solution from simulated trials is accurate and unbiased. (A) For nine different target positions, the panel shows the average inferred match (over 1,000 simulations) for five different competitor spacings (each plotted in a different color; see legend). Diagonal indicates the identity line. (B) For 55 different targets, the panel shows the average inferred match over 100 simulations for two extreme competitor spacings (large in blue and small in red). On both panels, ± 1 *SEM* are shown but are too small to be clearly visible.

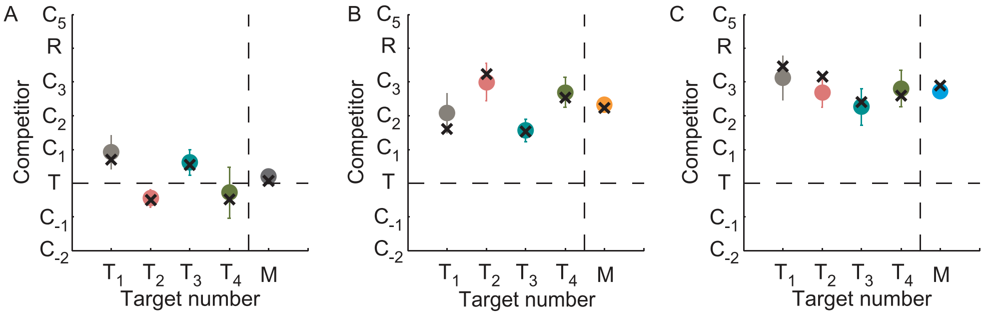

Figure A4. Bootstrapped inferred matches indicate high reliability of our MLDS procedure. Each panel shows different condition: (A) illuminant-constant, (B) yellow illuminant change, and (C) blue illuminant change. Within a condition, for each target (gray, rose, teal, and green) filled circles of the corresponding color plot mean bootstrapped inferred match (over 1,000 iterations of our resampling procedure). In each panel, the mean bootstrapped match for each condition (over targets) is shown on the far right and labeled as M. The error bars indicate bootstrapped 90% confidence intervals. Black crosses indicate the inferred match from the actual data.

ered competitor range of 1 or less, while for the simple stimuli the recovered range was typically higher, possibly because the selection-based matches tended to cluster near the tristimulus match. Indeed, the average separation between the recovered position of the tristimulus match and the closest competitor (C_1) for the simple stimuli was 0.28.

Further, to estimate confidence intervals for the selection-based matches, we used a bootstrapping procedure. From the set of one subject's responses for each competitor pair (across all trials), we resampled with replacement a new set of responses of the same size. We repeated this procedure for all competitor pairs for a given target and condition and computed the inferred match from the resampled data. Filled circles in Figure A4 show mean inferred match for each target and condition over 1,000 iterations of this resampling procedure of one subject (udm) in Experiment 1. Error bars indicate bootstrapped 90% confidence intervals; black crosses indicate the position of the inferred match from the actual data. Each panel shows different experimental condition (panel A, illuminant-constant; panel B, yellow illuminant change; panel C, blue illuminant change). For all targets, the actual inferred matches are in good agreement with the bootstrapped matches and the confidence intervals are small, indicating that our data and analysis yield stable estimates of selection-based matches. The online supplement provides a figure showing bootstrapped inferred matches for each of our subjects (see individual subject pages).

Computing constancy indices from a selection-based matches

We computed the color constancy index from inferred selection-based matches in the illuminant-changed condition following the formula: $CCI = 1-(b/a)$, where b denotes the Euclidian distance between the inferred selection-based match under the test illuminant and the reflectance match and a denotes the distance between the tristimulus and the reflectance match, both computed in three-dimensional CIELAB space. This is a three-dimensional variant of the chromaticity-based constancy index introduced by Arend et al. (1991). For the selection task, in which the inferred match is constrained to lie along the line in CIELAB space that connects the tristimulus and reflectance matches, this constancy index reduces to the ratio of the distance between inferred match and tristimulus match to the distance between reflectance match and tristimulus match. For both selection and matching, the constancy index is 1 if the match is equal to the reflectance match and 0 if it is equal to the tristimulus match.

In computing color constancy indices we used the nominal physical coordinates of the tristimulus match for a given illuminant condition (CIELAB coordinates of the competitor sample equal to the tristimulus match). An alternative option would be to use the perceived tristimulus match; that is, the subject's inferred selection-based match in the illuminant-constant condition. We used the nominal tristimulus match because, in Experiment 1, the tristimulus values of the target surface differed slightly in the illuminant-constant and the two illuminant-changed conditions; thus, the inferred match in illuminant-constant condition was made relative to the slightly different target. The difference in tristimulus coordinates of the target occurred because the illumination impinging on the target surface centered in the right side of the cube varied slightly as the illumination of the right side of the cube changed across experimental conditions, as it would in a natural scene. The effect of this illuminant

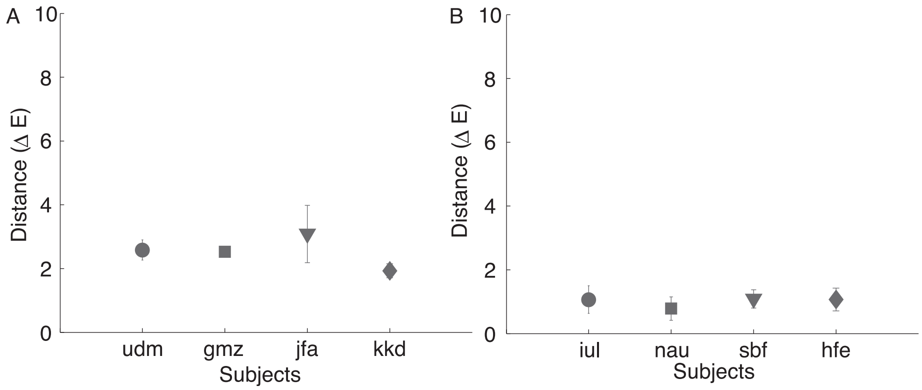

Figure A5. Distance of inferred selection-based match from the tristimulus match in the illuminant-constant condition. Average distance in ΔE units is computed over targets and plotted for each subject in Experiment 1 (A) and Experiment 2 (B). The error bars indicate ±1 *SEM*.

spillover was small (average distance between the target in the illuminant-constant and illuminant-changed condition was 0.9 CIELAB ΔE for the yellowish and 1.3 ΔE for the bluish illumination change, with the largest difference being 1.42 ΔE).

In all conversions between XYZ and CIELAB in this paper, the reference white point was taken as the XYZ coordinates of the standard illuminant, which differed slightly across the experiments (Experiment 1 white point: 90.23, 95.33, 102.76; Experiment 2 white point: 90.38, 95.22, 103.39).

In the illuminant-constant condition the distance between the inferred selection-based match and the tristimulus match is small

For each target we computed the distance in CIELAB ΔE units between the inferred selection-based match and the tristimulus match (T) in illuminant-constant condition for all subjects. The average distances (across targets) for each subject are shown in Figure A5 (panel A, Experiment 1; panel B, Experiment 2). In both experiments, these distances are small, providing a validation of our general method.