Journal of Experimental Psychology:
Human Perception and Performance
2016, Vol. 42, No. 6, 847–865

0096-1523/16/$12.00 http://dx.doi.org/10.1037/xhp0000184

The Nature of Instructional Effects in Color Constancy

Ana Radonjić and David H. Brainard
University of Pennsylvania

The instructions subjects receive can have a large effect on experimentally measured color constancy, but the nature of these effects and how their existence should inform our understanding of color perception remains unclear. We used a factorial design to measure how instructional effects on constancy vary with experimental task and stimulus set. In each of 2 experiments, we employed both a classic adjustment-based asymmetric matching task and a novel color selection task. Four groups of naive subjects were instructed to make adjustments/selections based on (a) color (neutral instructions); (b) the light reaching the eye (physical spectrum instructions); (c) the actual surface reflectance of an object (objective reflectance instructions); or (d) the apparent surface reflectance of an object (apparent reflectance instructions). Across the 2 experiments we varied the naturalness of the stimuli. We find clear interactions between instructions, task, and stimuli. With simplified stimuli (Experiment 1), instructional effects were large and the data revealed 2 instruction-dependent patterns. In 1 (neutral and physical spectrum instructions) constancy was low, intersubject variability was also low, and adjustment-based and selection-based constancy were in agreement. In the other (reflectance instructions) constancy was high, intersubject variability was large, adjustment-based constancy deviated from selection-based constancy and for some subjects selection-based constancy increased across sessions. Similar patterns held for naturalistic stimuli (Experiment 2), although instructional effects were smaller. We interpret these 2 patterns as signatures of distinct task strategies—1 is perceptual, with judgments based primarily on the perceptual representation of color; the other involves explicit instruction-driven reasoning.

Keywords: color constancy, instructional effects, asymmetric matching, color selection, task strategy

Understanding how the visual system extracts information about the color of the objects in the environment is a fundamental open question in vision science. The problem arises because of the inherent ambiguity of the signal that reaches the photoreceptors: the spectrum of light reflected from objects to the eye depends not only on object surface reflectance, which is the physical correlate of object color, but also on incident illumination. Any reflected light spectrum can result from a myriad of different surface-and-illuminant combinations. By the same token, a fixed object will reflect different spectra when viewed under different illuminations. Disentangling the intrinsic surface reflectance component from the transient and variable illumination component is computationally challenging. Despite the challenge, the visual system provides a fairly constant perceptual representation of object color, and we rely on this representation to guide action (e.g., selecting fresh and avoiding spoiled food). The underlying mechanisms that support such *color constancy*, however, are not fully understood.

Endeavors to understand constancy are complicated by the finding that the instructions subjects receive can modulate the degree of experimentally measured constancy. Early influential studies on this topic, conducted by Arend and his collaborators (Arend & Goldstein, 1987; Arend, Reeves, Schirillo, & Goldstein, 1991; Arend & Reeves, 1986) demonstrated instructional effects in the context of an asymmetric matching task. In their experiments, subjects viewed a pair of stimulus configurations, each a simulation of illuminated papers presented on a computer screen. The simulated papers across the two configurations were identical, but their simulated illumination differed. Subjects were asked to adjust a test patch in one configuration to match a corresponding patch (the standard) in the other, and the data were analyzed in terms of how much constancy the matches revealed. Across conditions, the task instructions were varied and this resulted in different degrees of measured constancy. When instructed to adjust the test to "match the hue and saturation (and/or brightness) of the test patch to those of the standard patch [. . .] while disregarding, as much as possible, other areas in the screen" subject matches indicated low constancy. Constancy increased considerably, however, when the same subjects were asked to "make the test patch look as if it were cut from the same piece of paper." Similar instructional effects were measured for full color stimuli (Arend et al., 1991; Arend & Reeves, 1986) and for the achromatic stimuli (Arend & Goldstein, 1987).

Several characteristics of the experimental design of these early studies are worth noting. First, the studies employed relatively simple stimuli. These were simulation of flat matte surfaces ren-

This article was published Online First January 4, 2016.

Ana Radonjić and David H. Brainard, Department of Psychology, University of Pennsylvania.

This work has been supported by NIH RO1 EY10016 (DHB), Core Grant P30 EY001583, and a 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. Supporting material for this publication is available online at http://color.psych.upenn.edu/supplements/instructionaleffects/

Correspondence concerning this article should be addressed to Ana Radonjić, Department of Psychology, University of Pennsylvania; 3401 Walnut Street, Wing C, Suite 327C, Philadelphia, PA 19104. E-mail: radonjic@sas.upenn.edu

dered under spatially diffuse illumination; they consisted either of pairs of surfaces (disk-and-annulus configurations) or larger sets of overlapping rectangular surfaces (Mondrian configurations). Second, the effects of instructions were measured within-subjects. Third, the majority of subjects were experienced (the authors and the members of their labs)—aware of the computational problem of color constancy and familiar with how changes in illumination affect the light reflected from a fixed object.

A number of subsequent studies replicated the instructional effects reported by Arend and colleagues using asymmetric matching and similar stimulus configurations. This was done both for successive (Troost & de Weert, 1991) and simultaneous (Bauml, 1999; Cornelissen & Brenner, 1995; Troost & de Weert, 1991) matching and both with within-subjects (Cornelissen & Brenner, 1995; Troost & de Weert, 1991) and between-subjects designs (Bauml, 1999). One of those studies found instructional effects only for experienced subjects (Cornelissen & Brenner, 1995), but others reported instructional effects for naive subjects as well (Bauml, 1999; Troost & de Weert, 1991).

Large instructional effect on color constancy were also revealed in the a study that used a modified version of the asymmetric matching task in which, rather than adjusting the test surface, the subjects rated (on a scale of 0% to 100%) the extent to which the test appeared either (a) the same hue and saturation as the target surface, or (b) as if it were made of the same piece of paper as the target (Reeves, Amano, & Foster, 2008). In this study, which used both within- and between-subjects designs, the subjects also completed a task in which they provided a yes/no judgment on whether the two Mondrian configurations were made out of the same material; this type of judgment was highly correlated with the surface-based (paper) matching ratings. In a related study, Van Es, Vladusich, and Cornelissen (2007) showed that across a simulated change in illumination subjects are able to make fairly accurate judgments of both (a) the local properties of the test surface (did the test patch change in hue/saturation/brightness), which was interpreted to indicate low color constancy, as well as (b) global properties of the scene (did the test patch change in the manner consistent with the overall change in illumination), which was interpreted to indicate higher constancy.

Some studies which probed achromatic color perception introduced a third type of instructions: In addition to brightness and lightness (paper) matches, the subjects were asked to make *brightness contrast* matches ("make the brightness difference between the test and the surround the same as between the standard patch and the surround") and, under certain conditions, these matches differed from both brightness and lightness matches (Arend & Spehar, 1993a, 1993b; Blakeslee, Reetz, & McCourt, 2008; for another version of instructional manipulation in the lightness domain see Rudd, 2010).

A number of color and lightness constancy studies, however, varied instructions along the same lines as the early studies of Arend and colleagues and failed to find substantial effects (Delahunt & Brainard, 2004; Logvinenko & Tokunaga, 2011; Madigan & Brainard, 2014; Ripamonti et al., 2004). These studies were all conducted using naive subjects and experimental methods other than adjustment-based asymmetric matching (such as achromatic adjustment or palette matching) and, predominantly, a between-subjects design (Logvinenko & Tokunaga, 2011 study was within-subjects). In addition, these studies employed more naturalistic stimulus configurations than the studies reviewed above that reported large instructional effects. Here the stimuli were real illuminated objects or fairly realistic graphics simulations (e.g., three-dimensional scenes presented stereoscopically). This difference suggests that both the task and the class of stimuli used in the experiment may modulate instructional effects in constancy studies.

That the choice of stimuli affects experimentally measured constancy is also suggested by a number of studies that used neutral (nonspecific) instructions. In our recent work, for example, we showed that when stimuli were fairly realistic simulations of illuminated objects constancy was good, but that it dramatically decreased when stimuli were reduced to square patches presented against the textured background, even though the colorimetric characteristics of the stimuli were closely matched (Radonjić, Cottaris, & Brainard, 2015b). High degrees of constancy were also found in other studies that used naturalistic stimuli (Brainard, Brunt, & Speigle, 1997; Kraft & Brainard, 1999), suggesting that "surface-based" instructions are not necessary for good constancy. It remains unclear however whether introducing such instructions would have led to even higher degrees of measured constancy in these experiments (see also Wright, 2013).

In summary, the extant literature makes clear that it is possible to find instructional effects in studies of color and lightness constancy. What is much less clear, however, is the nature of these effects and what they tell us about the human color and lightness constancy (Brainard & Radonjić, 2014; Kingdom, 2011).

Some authors have argued that different instructions prompt subjects to report about different aspects of a fixed perceptual representation, in the same way one can, for example, independently judge object's size or its orientation (Arend & Spehar, 1993a). In a similarly dualistic vein, it has also been proposed that different instruction probe different types of processes that support color constancy or different "perceptual modes" (Arend et al., 1991; Arend & Reeves, 1986; see also Rock, 1983). Others posit that the perception of object color is based on a unitary perceptual representation and that instructional effects reflect the fact that certain types of instructions drive subjects to rely on explicit reasoning (from the unitary perceptual representation) to make a prompted-for match. In this regard, some posit that hue/saturation/brightness instructions prompt subjects to reason when making (unnatural) judgments about the characteristics of the proximal stimulus (Gibson, 1950; Gilchrist, 2012; see also Koffka, 1935; MacLeod, 2012). Others, however, argue that subjects tend to engage in explicit reasoning when they are given surface-based (paper) instructions and that under these instructions, their matches are best described as *inferred* color (or lightness) judgments (Blakeslee & McCourt, 2015; Blakeslee et al., 2008). Distinguishing between the various theoretical accounts is challenging because it is not clear what experimental data would clearly support one over the others.

To provide a better understanding of instructional effects on constancy and their nature, we designed a study to measure systematically whether and how such effects depend on stimulus set and experimental task. We asked four groups of subjects, each of which received a different type of instructions, to complete two different color constancy tasks: a classic asymmetric matching task, for which instructional effects have been frequently reported, and a color selection task, recently developed in our lab.

In the color selection task, subjects are asked to select objects based on color across a change in illumination. The task is intended to probe constancy in the manner that captures the real-world use of color, where we frequently rely on it to select objects to meet specific goals. For example, we use color information to select the ripest tomatoes, rather than adjusting the tomatoes until they look ripe enough to eat. Our previous study compared constancy across color selection and the asymmetric matching for neutral instructions, and found good agreement for those instructions. This finding held for both simplified and naturalistic stimuli (Radonjić et al., 2015b).

In this study we also used two different classes of stimuli: simplified and naturalistic. These were identical to stimuli we used in our previous study (Radonjić et al., 2015b). Our simplified stimuli (Experiment 1) consisted of simulations of diffusely illuminated two-dimensional patterns of rectangular matte paper patches; they resembled the flat-matte diffusely illuminated geometric patterns used in studies which report large instructional effects. Our naturalistic stimuli (Experiment 2) consisted of realistic simulations of three-dimensional scenes and were presented stereoscopically. They were modeled after the color cube illusion of Lotto and Purves (1999) and depicted a large multifaceted cube suspended in midair in the center of a room in which the illumination varied spatially.

We used four different types of instructions, labeled as neutral, physical spectrum, objective reflectance, and apparent reflectance instructions. The neutral instructions simply ask subjects to judge color, without any further definition of the term *color*. The physical spectrum instructions are formulated to probe sensory experience and ask subjects to make judgments based on the qualities of the light reaching their eye while disregarding any signals in the image that might indicate changes in illumination. These are similar to the hue/saturation/brightness instructions used in the prior literature. The objective reflectance and the apparent reflectance instructions are both surface-based and the difference between them is subtle: the latter asks for judgment based on subjective appearance, while the former asks for judgment based on objective surface properties (e.g., "adjust the test so that *it is made* out of the same material as the target square" vs. "adjust the test so that it *looks like it is made* from the same material as the target;" see also Wagner, 2012). Data for the neutral instructions were reported in our previous paper (Radonjić et al., 2015b), and are presented again here to provide a baseline for comparison.

Experiment 1

Method

Apparatus. The stimuli were presented on a calibrated 21" CRT color monitor (ViewSonic, Model Graphic Series G225fB) driven via a dual-port video card (NVIDIA GeForce GT120) at a pixel resolution of 1,280 × 1,024 and refresh rate of 75 Hz and with 8-bit resolution for each RGB channel. The host computer was an Apple Macintosh with an Intel Xeon quad-core processor. An eye tracker (EyeLink 1000, Desktop remote model, SR Research) was used to record the position of the eye, but eye-tracking data will not be reported here.

The subject's head position was stabilized using a chin rest. Subjects viewed the stimuli monocularly using their right eye while their left eye was covered with an eye patch. The distance between the subject's eye and the center of the screen was 76 cm. The experimental programs were written in Matlab and relied on Psychtoolbox (Brainard, 1997; Pelli, 1997, http://psychtoolbox.org) and mgl (http://justingardner.net/doku.php/mgl/overview) routines.

Stimulus. The stimulus configuration consisted of five squares (each 3.5 cm a side, 2.6°) presented against a textured color background (see Figure 1). The square in the center of the screen served as the target and was surrounded by the four squares (each at 8° eccentricity, measured from the center of the target to the center of the surrounding square).

Across trials, we used four different colored targets (Figure 2A). They all had the same luminance (23.6 cd/m^2). One target ("gray") was achromatic (CIELAB chroma: 0.4, hue angle: 314.01°); the remaining three ("rose," "teal," and "green") were equal in saturation (chroma: 25.6), but were sampled from different regions of the hue circle (hue angle: 355.31°, 227.14°, 135.46°).

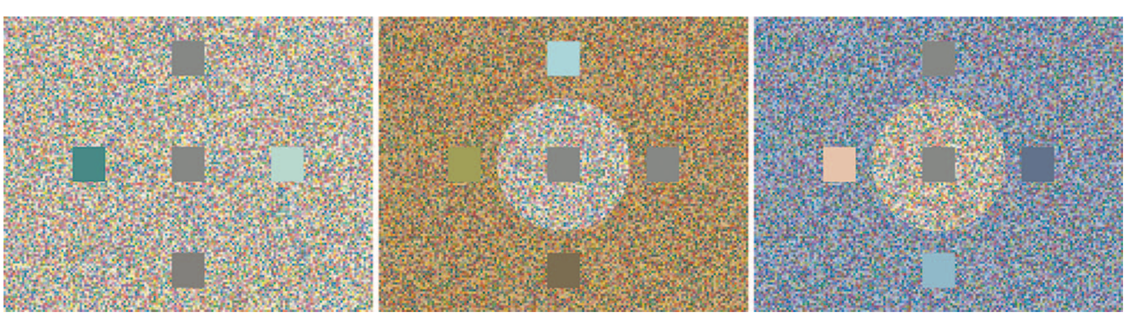

Figure 1. Experiment 1: Stimuli. Example stimuli for the color selection task. The square in the center of the configuration is the target and it is surrounded by four squares. In the illuminant-constant example (left), the top square is the tristimulus and reflectance match for the target while the bottom square is a competitor (C_{-1}). In the yellowish illuminant-change example (center) the square on the right is the tristimulus match, while the bottom square is the reflectance match. In the bluish illuminant-change example (right), the top square is the tristimulus match while the square on the right is the reflectance match. All three examples show the gray target. See the online article for the color version of this figure.

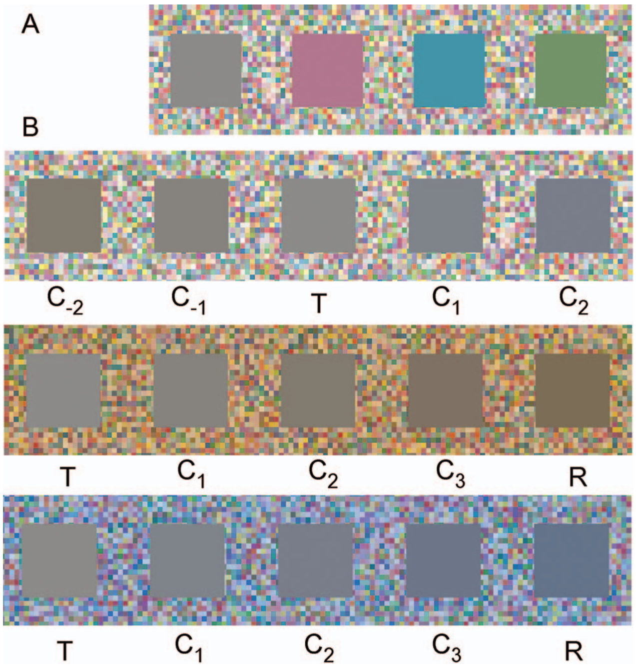

Figure 2. Experiment 1: Targets and competitors. Panel A. The four targets (shown under the standard illuminant). Panel B. Competitor set in the color selection task for one target (gray) in the illuminant-constant (top row) and the illuminant-changed conditions (middle row: yellowish; bottom row: bluish). See text for details on how the competitor sets were constructed. See the online article for the color version of this figure.

The textured background behind the squares was constructed as an array of simulated small rectangular Munsell papers (0.26 cm × 0.23 cm each, 0.17° × 0.19°). On illuminant-constant trials, the background behind the target and the surrounding squares was uniformly illuminated by the standard illuminant (6500 K CIE daylight; Figure 2, left). On illuminant-changed trials, the simulated illuminant of the background behind the surrounding squares changed to the yellowish test illuminant (4500 K CIE daylight; Figure 2, center) on one half of the trials and to the bluish test illuminant (12000 K CIE daylight; Figure 2, right) on the other half. On these trials, a small circular area of the background around the target (13.3 cm in diameter, 10°) remained under the standard illuminant.

The stimulus background was created by randomly sampling (with replacement) from a subset of ~220 Munsell paper samples (out of 462) whose surface reflectance is known (Nickerson, 1957). The subset only included the samples that we could render within the gamut of our display in each illuminant condition and whose luminance (under the standard illuminant) was at least 20 cd/m^2. We created 10 different background patterns (160 × 128 patches each). On each trial, one of these backgrounds was randomly chosen and rendered under the simulated illuminants appropriate for a given trial.

The mean xy chromaticity for backgrounds in the illuminant-constant and yellowish and bluish illuminant-changed condition was [0.33, 0.34], [0.38, 0.38] and [0.29, 0.30] while the luminance was 45.17, 29.41, and 30.32 cd/m^2, respectively. Across trials, the targets (as well as all the competitors from the color selection task; see below) were luminance decrements relative to the average of the textured backgrounds.

Color selection task. At the beginning of each trial, a black fixation cross was displayed against the textured background at the center of the screen. To initialize a trial subjects used the computer mouse to move a cursor (a small black dot) to the center of the cross and clicked the mouse. The stimulus (the target and the surrounding squares) was then displayed.

Two of the surrounding squares were distractors, whose color was highly dissimilar from the target. The remaining two squares were competitors and their degree of color similarity to the target varied across trials. The subject's task was to use the mouse to move the cursor onto the surrounding square that was closest to the target in color and click the mouse. The meaning of the word "color" was defined by experimental instructions (see below).

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

1. The tristimulus match for the target (denoted as T) which had a different simulated surface reflectance, but the same tristimulus coordinates as the target under the standard illuminant.

2. The reflectance match for the target (R), which had the same simulated surface reflectance as the target. The tristimulus coordinates of the reflectance match were different from those of the target because of the change in simulated illuminant.

3–5. Three color samples (C_1, C_2, C_3), which were equally spaced along the line in CIELAB color space that connected the tristimulus and the reflectance match. We used the XYZ coordinates of the standard illuminant ([90.38, 95.22, 103.39]) as the white point for conversion of XYZ to CIELAB values.

In illuminant-constant condition the competitor set included five color samples: the tristimulus match for the target (which, in this condition, was also the reflectance match) and the two closest competitors from the yellowish (C_{-1} and C_{-2}) and from the bluish (C_1 and C_2) competitor set.

On each trial, the squares that served as distractors were randomly drawn (without replacement) from a predefined set of distractors for each target. This set consisted of simulations of the Munsell papers used for the background checks (under the standard illuminant) that differed from the target and any of its competitors by at least 20 CIELAB ΔE units.

Illuminant-constant and illumination-changed trials were blocked. Within a block of trials, each target was presented with all pairwise combinations of its competitors. Thus, each illuminant-constant block consisted of 40 trials (1 standard illuminant × 4 targets × 10 possible competitor pairs) while each illuminant-changed block consisted of 80 trials (2 test illuminants × 4 targets × 10 possible competitor pairs; bluish and yellowish trials intermixed) presented in random order.

At the beginning of the first session all subjects completed a brief training which consisted of four illuminant-constant trials (each with a different target).

Subjects completed 20–28 illuminant-constant blocks and 30–32 illuminant-changed blocks across seven to nine 1-hr sessions. Typically, the first and the fourth session consisted of illumination-constant blocks, while the remaining sessions consisted of illuminant-changed blocks. In the last session, the subjects completed both types of trials (all remaining trials needed to finish the experiment), with all illuminant-constant blocks completed first.

Asymmetric matching task. The stimuli in the asymmetric matching task closely matched those used in the color selection task: after the subject initiated a trial (in the same manner), the stimulus configuration, consisting of five squares presented against the textured background, was displayed. As with the color selection task, the square in the center served as the target and was surrounded by four squares. One of the surrounding squares was the test square and its color was set to either white or black at the beginning of the trial. The remaining three squares were randomly chosen from the predefined set of distractors.

The subjects' task was to adjust the test square to match the target in color (as defined by the instructions). They completed the task using a controller which allowed adjustment of test's CIELAB L^*, chroma and hue. The subjects could take as much time as they needed to set the desired match. They were also allowed to select a "match impossible" option if they felt they were not able to achieve the desired match; there were only a few such trials and they were excluded from the analysis (1/72 for subjects tuj, hfe, and goh; 2/80 for mik and 2/72 for nke).

Each illuminant-constant block consisted of four trials (1 standard illuminant × 4 targets), while each illuminant-changed block consisted of eight trials (2 test illuminants × 4 targets). The subjects completed six blocks of trials in each illuminant condition in three to four 1-hr sessions (except subjects mik who completed 7 illuminant-changed blocks and mil who completed 7 illuminant-changed and four illuminant-constant blocks of trials). The sessions were blocked by illumination condition: the subjects completed all illuminant-constant blocks of trials (in one or two sessions) before moving to illuminant-changed blocks (completed in two to three sessions).

Asymmetric matching training. Prior to the first asymmetric matching session all subjects completed a training to familiarize themselves with the matching task and learn how to use the controller. We used the same apparatus as for the experiment, but the subjects viewed the display binocularly without an eye patch. In the training, the subjects completed three to seven blocks of trials (up to 10 trials per block) across two to three sessions. On each training trial, two squares—the target and the test—were presented adjacent to one another against the illuminant-constant background and the subject made a symmetric color match. Target colors were set randomly (by drawing a random triplet of RGB values). The first training trial was completed by the experimenter, who demonstrated how to use the controller and provided step-by-step explanations while making the match. The subject made the second match with experimenter's help and then continued to make matches unassisted and without immediate feedback. At the beginning of the each following training session the experimenter reviewed the worst matches from the previous session (assessed using the CIELAB ΔE metric) and encouraged the subjects to make matches that agreed more with the target.

Subjects. Sixteen subjects (three male and 13 female, all 19–22 years of age) participated in the experiment. They all had normal color vision, as assessed by the Ishihara plates (Ishihara, 1977, up to one plate incorrect). All except one had normal or corrected to normal visual acuity of 20/40 or better (as assessed by a Snellen chart). Measured visual acuity for the remaining subject (fai) was 20/50. All experimental procedures were approved by University of Pennsylvania Institutional Review Board and were in accordance with the APA Ethical Principles and World Medical Association Helsinki Declaration.

Instructions. We used four different sets of experimental instructions (*neutral*, *physical spectrum*, *objective reflectance*, and *apparent reflectance*). A different group of four subjects was assigned to each instructional condition; each subject received only one type of instructions in all phases of the experiment. Within each instructional group, two of the four subjects completed the color selection task first, while the other two completed asymmetric matching task first.

At the beginning of the experiment all subjects except those in the neutral instructions group went through the induction procedure in which they were familiarized with the type of color judgment they were asked to make in the experiment. A full description of the induction procedure is provided in the supporting material available online (http://color.psych.upenn.edu/supplements/instructionaleffects/). Briefly, the subjects were taught the difference between *surface reflectance* and *reflected light* through a series of demonstrations in which they observed how the changes in illumination affected the light reflected from different colored papers. The induction procedure was repeated twice for each observer: (a) at their very first session (the illuminant-constant color selection session for the subjects who did the selection task first or the asymmetric matching training session for those who did asymmetric matching first); and (b) before the first illuminant-change session of the task they were to complete first.

The subjects also received task-specific instructions, which were read to them and repeated before each experimental session for the duration of the experiment (and after the induction procedure when it was performed). For each condition and task we provide instructions verbatim in the supporting material. The procedural aspects of the instructions were the same across groups, but the way the term *color* was defined differed. For each instructional group the instructions for the color selection task were as follows (ellipses are inserted in places that described procedural aspects of the experiment; see the supporting material for instructions verbatim).

Neutral. "Your task is to click on the test square that is closest to the target square in color."

Physical spectrum. "You should think about these squares as simulations of illuminated paper surfaces. In this context, your task is to click on the test square from which the light reaching your eye is most similar to the light from the target square . . . In the experiment, you may notice that on some trials there will be a change in background behind the test squares. Try to ignore that as much as possible and focus on choosing the test square that delivers that most similar light to your eye as the target square—as if you were looking at the test squares through the tube that we used when we explained to you the difference between surface reflectance and reflected light [in the induction procedure]."

Objective reflectance. "You should think about these squares as simulations of illuminated paper surfaces. In this context, your task is to click on the test square that is cut from the piece of paper most similar to the target square, that is the test that has the same reflectance properties as the target . . . In the experiment you may notice that on some trials there will be a change in background behind the test squares. Think of this as a change of illumination and focus on choosing the test square that is closest in surface reflectance to the target. That is chose the test that would be most similar to the target, if the target were under the changed illumination as well."

Apparent reflectance. "You should think about these squares as simulations of illuminated paper surfaces. In this context, your task is to click on the test square that looks like it is cut from the piece of paper most similar to the target square, that is the test that looks like it has the same reflectance properties as the target . . . In the experiment, you may notice that on some trials there will be a change in background behind the test squares. Think of this as a change of illumination and focus on choosing the test square that looks closest in surface reflectance to the target. That is chose the test that would look most similar to the target, if the target were under the changed illumination as well."

The instructions for the asymmetric matching task closely matched those used in the color selection task:

Neutral. "Your task is to adjust the test square so that it matches the target square in color."

Physical spectrum. "Your task is to adjust the test square so that the light reaching your eye from it is the same as the light reaching your eye from the target square. In the experiment, you may notice that on some trials there will be a change in background behind the four squares. Try to ignore that as much as possible and focus on adjusting the test square so that the light it delivers to your eye is the same as that from the target square—as if you were looking at the test square and the target through the tube that we used when we explained to you the difference between surface reflectance and reflected light."

Objective reflectance. "Your task is to adjust the test square so that it has the same reflectance properties as the target square . . . In the experiment you may notice that on some trials there will be a change in background behind the four squares. Think of this as a change of illumination and focus on adjusting the test square so that it matches the target in surface reflectance. That is, adjust the test so that it matches the target, if the target were under the changed illumination as well."

Apparent reflectance. "Your task is to adjust the test square so that it looks like it is cut from the same piece of paper as the target square. That is, adjust the test square so that it looks like it has the same reflectance properties as the target square . . . In the experiment, you may notice that on some trials there will be a change in background behind the four squares. Think of this as a change of illumination and focus on adjusting the test square so that it looks like it has the same surface reflectance as the target. That is adjust the test so it looks like the target, if the target were under the changed illumination as well."

In the asymmetric matching task (and training) we also used a different term to refer to CIELAB L^* dimension across instructional groups: term *intensity* was used in the neutral, term *brightness* in the physical spectrum and term *lightness* in the objective and apparent reflectance instructions. The label for the button controlling the intensity on the controller schema used in the asymmetric matching training also differed across conditions to reflect this change (see supporting material).

Postexperiment questionnaire. In the end of the study the subjects completed a short questionnaire in which they were asked to describe any strategy they might have used to complete the color selection and the asymmetric matching tasks. They were also invited to provide any additional comments they might have had about the experiment.

Supporting material. For both experiments supporting material available online (http://color.psych.upenn.edu/supplements/instructionaleffects/) provides detailed colorimetric specification of the stimuli (including CIELAB, xyY, and LMS values for each target and competitor), target reflectances and spectra of each illuminant, instructions verbatim, the postexperiment questionnaires with responses, tabulated results of statistical analyses, and individual data for each subject.

Data analysis: Color selection. We developed a method that allows us to quantify the degree of color constancy that mediates subject's performance in the color selection task. Our method relies on the observer model implemented in the maximum likelihood difference scaling (Maloney & Yang, 2003) and we have described it in detail in earlier papers (Radonjić, Cottaris, & Brainard, 2015a; Radonjić et al., 2015b). Briefly, we assume that, each stimulus (the target and the competitors) occupies a certain position in an underlying one-dimensional perceptual representation. This position is subject to perceptual noise, and on each trial it is described as a draw from a normal distribution, centered around its mean position. The subject's choice is modeled as a comparison between the current target and the competitor representations, with the subject choosing the competitor whose representation is closest to that of the target. Our analysis method takes as an input the subject's choices across a series of trials and, via a numerical search procedure, recovers the mean position of the target and each of the competitors in the underlying perceptual representation that best accounts for the subject's choices measured in the experiment.

In the recovered representation, the position of the target is the selection-based match for a given illumination condition. Conceptually, the selection-based match is equivalent to the selection-based point of subjective equality; it is the color sample that the subject would select on the majority of trials as "the closest one to the target" over any other competitor (see Radonjić et al., 2015a).

When the position of the selection-based match falls within the range of the competitors, we assume that the relative distances between the match and the two adjacent competitors are preserved in the recovered representation and we use linear extrapolation to infer the CIELAB coordinates of the selection-based match. In the illuminant-constant condition, in which both the target and the competitors are presented under the same illumination, the selection-based match is expected to fall at the tristimulus match. In the illuminant-change condition, the distance between the selection-based match from the reflectance match indicates the degree of color constancy in the color selection task: the closer the selection-based match is to the reflectance match, the higher the constancy.

For each target and illuminant-change condition we quantify constancy by computing a color constancy index (*CCI*) following the formula:

$$CCI = 1 - (b/a),$$

where b denotes the Euclidian distance (in three-dimensional CIELAB space) between the selection-based match and the reflectance match and a denotes the distance between the tristimulus match and the reflectance match (Arend et al., 1991).

In some cases the recovered position of the selection-based match falls out of the range of competitors. In the illuminant-changed condition, this would occur, for example, if across all pairwise combinations of the competitors a subject always chose the sample that is closer to the reflectance match. In this case, the selection-based match is not well constrained by the data and its recovered position will be beyond the reflectance match, in the direction of overconstancy. Similarly, if a subject always chose the competitor in the pair that is closer to the tristimulus match, the selection-based match would fall out-of-range on the tristimulus match end. Rather than excluding these out-of-range matches from the analysis, we assigned them a position in color space that qualitatively captured the underlying pattern of choices. That is, we assigned them the coordinates that were outside of the range of competitors, but along the same line in CIELAB space, at 1/10 of the inter-competitor distance from R (out-of-range on the reflectance end) or T (out-of-range on the tristimulus end). These positions that corresponded to the selection-based color constancy indices of 0.975 and −0.025, respectively. Out-of-range matches occurred with some frequency in Experiment 1 (as we discuss below), but in only one instance in Experiment 2 (subject kkd; rose target in the yellowish illuminant-change condition).

Data analysis: Asymmetric matching. For the asymmetric matching task, for each subject we computed the mean match (across all repetitions) for a given target and illuminant condition. We then used this match to compute color constancy indices using the same formula we used to quantify constancy in the color selection task.

Results

For each of our 16 subjects, Figure 3A shows the mean recovered position of the selection-based match in each illuminant condition, averaged across targets. Figure 3B shows the selection-based color constancy indices for the yellowish and bluish illumination change.

The data reveal clear instructional effects. When subjects are asked to select objects based on their surface reflectance properties, constancy was higher than when they were instructed to make selections based on reflected light or when they were given neutral instructions. A three-way repeated-measure analysis of variance (ANOVA), with instructional group as between-subjects factor (four groups) and test illuminant (yellowish vs. bluish) and target reflectance (four different reflectances) as within-subject factors, revealed a significant main effect of instructions, $F(3, 12) = 11.26$; $p = .001$. We did not find a significant main effect of target or test illuminant, or any significant interaction between the factors (although there were trends for main effect of illuminant, $F(1, 12) = 4.68$, $p = .051$, and Illuminant × Instruction interaction, $F(3, 12) = 3.00$, $p = .07$). Supporting material available online provides the complete results for all statistical analyses we conducted.

The instructional effects we find are large: the mean color constancy index was 0.10 for both the neutral and the physical spectrum group, but 0.60 for the objective reflectance group and

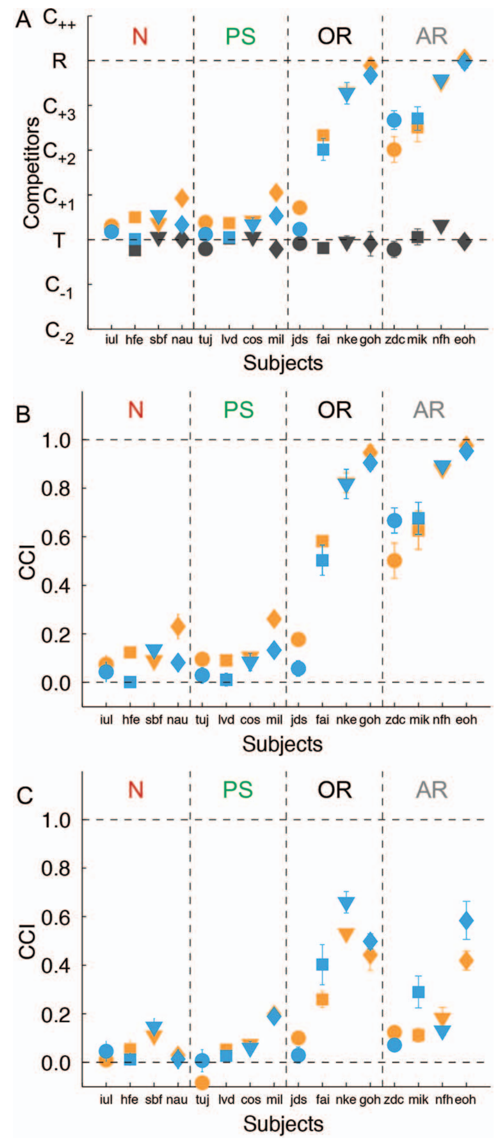

Figure 3. Experiment 1: Results. Panel A shows the mean position of the subject's selection-based match (averaged across targets) relative to the space of competitors. Panel B shows subject's mean selection-based color constancy indices. Panel C shows mean adjustment-based color constancy indices measured with asymmetric matching. Subjects that belong to different instructional groups are separated by vertical lines and letters on top of the plot indicate instructional condition (N = neutral; PS = physical spectrum; OR = objective reflectance; AR = apparent reflectance). The illuminant-constant condition is shown in gray (dark gray), yellowish illuminant-changed condition in orange (light gray) and bluish illuminant-changed condition in blue (middle gray). Error bars represent +/−1 SEM. Within each group different symbols indicate different subjects. See the online article for the color version of this figure.

0.77 for the apparent reflectance group. Figure 4A (filled bars) shows mean constancy indices for each group, averaged over subjects and test illuminants. We used a bootstrapping procedure to further explore differences in constancy between the groups. On each iteration of the bootstrapping, we sampled (randomly, with

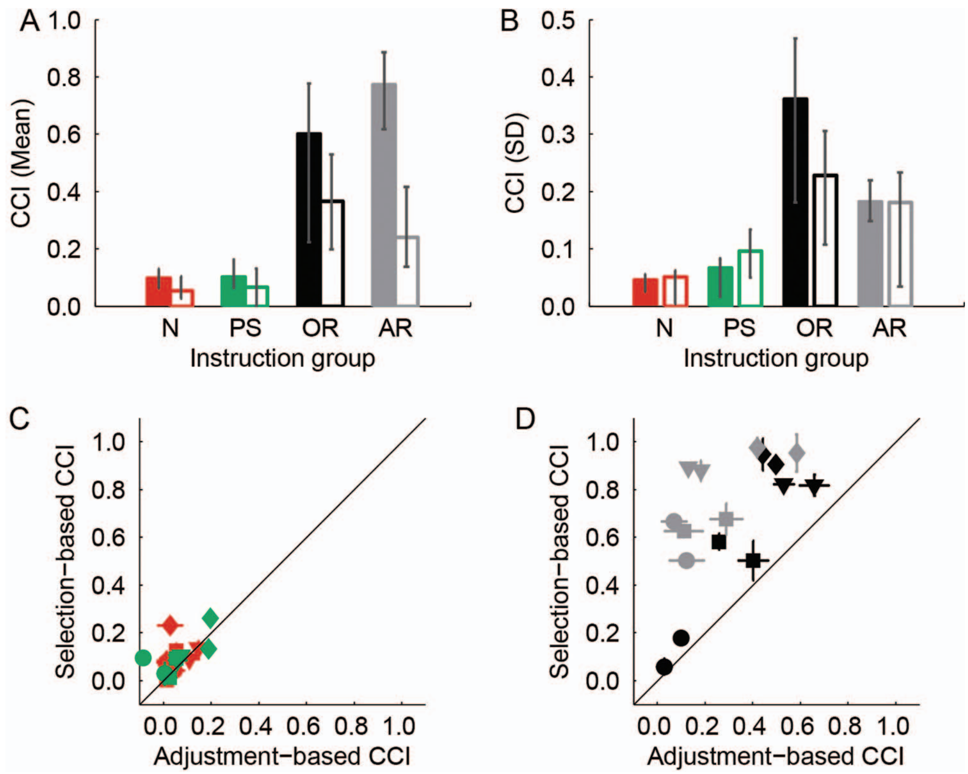

Figure 4. Experiment 1: Comparison of selection-based and adjustment-based constancy across instructional groups. Panel A shows mean constancy indices for each group for color selection (filled bars) and asymmetric matching (open bars). Panel B shows the standard deviations of the constancy indices for each group. In both panels, error bars are bootstrapped 90% confidence intervals. Panel C plots mean (over targets) selection-based indices against corresponding mean adjustment-based indices for subjects in the neutral (red or dark gray) and physical spectrum (green or light gray) instructions groups. Panel D shows the corresponding plot for objective reflectance (black) and apparent reflectance (gray) groups. Error bars represent +/−1 SEM. Symbols for individual subjects are the same as in Figure 3. In panels C and D there are two points for each subject as the means for each illuminant-changed condition are plotted separately. See the online article for the color version of this figure.

replacement) a new set of four subjects from the four actual subjects in a given instructional group. We then computed the mean constancy indices for each group over 2,000 iterations of the resampling along with the corresponding 90% confidence intervals (Efron & Tibshirani, 1993, pp. 178–201). The bootstrapped means were essentially identical to the group means (+/−0.005); the bootstrapped 90% confidence intervals are plotted as error bars in Figure 4A. Comparing the confidence intervals across groups suggests that: (a) constancy was lower for the neutral group than for either the objective or apparent reflectance groups, (b) constancy was also lower for the physical spectrum group than either the objective or apparent reflectance groups, and (c) the neutral and physical spectrum groups did not differ from each other in constancy, nor did the objective and apparent reflectance groups.

The same general trends are revealed by the constancy indices obtained from the asymmetric matches (Figure 3C), but the instructional effects we measured were smaller. We found significant differences in constancy across groups (main effect of instructions $F(3, 12) = 3.74$, $p < .05$): matches of subjects who received neutral or physical spectrum instructions indicated considerably lower degree of constancy (mean *CCI* of 0.06 and 0.07) than matches of subjects who received objective or apparent reflectance instructions (mean *CCI* of 0.24 and 0.37). Although the overall degree of constancy varied across targets (main effect of target, $F(3, 36) = 3.10$, $p < .05$; Illuminant × Target interaction, $F(3, 36) = 4.31$; $p < .05$), these variations were independent of the instructional manipulation (we did not find a significant Instructions × Target, Instructions × Illuminant, or Instructions × Target × Illuminant interaction; values for all statistical tests are provided in the supporting material available online). Figure 4A shows mean constancy indices across groups for the asymmetric matching task (open bars) with error-bars representing bootstrapped 90% confidence intervals. Comparing the confidence intervals across groups shows the same pattern of differences as we found for the color selection task: Constancy did not differ for the neutral and the physical spectrum groups or for the two reflectance groups, but was higher for the two reflectance groups than for the neutral/physical spectrum groups.

We analyzed the variability in constancy indices across instructional groups, using the same bootstrapping method we used to analyze the means. Figure 4B shows the standard deviations of constancy indices for the color selection task (filled bars) and the asymmetric matching task (open bars), with error bars representing

90% bootstrapped confidence intervals. For both tasks, the overall variability was lower for the neutral and the physical spectrum groups than for the objective and the apparent reflectance groups. For the color selection task, this difference exceeded the measurement variability as indicated by the confidence intervals. For the asymmetric matching task the measurement variability was larger and confidence intervals across instructional groups overlapped.

To understand better the nature of instructional effects, we compared subjects' performance across the two tasks. Figure 4 (panels C and D) plots mean adjustment-based constancy indices measured in the asymmetric matching task against the selection-based constancy indices measured in the color selection task for each subject and test illuminant. For clarity, neutral, and physical spectrum instructions groups are plotted in Figure 4C while objective reflectance and apparent reflectance groups are plotted in Figure 4D. The diagonal in the panels indicates the identity line: The closer the matches are to the diagonal, the better the agreement between two constancy measures.

The figure illustrates three main differences between the neutral and physical spectrum instructions groups on the one hand and the objective and apparent reflectance instructions groups on the other. For the neutral and physical spectrum instructions groups (a) overall constancy is low in both tasks, (b) variability between subjects is also low, and (c) the selection-based and adjustment-based asymmetric matches are in good agreement: Subjects' matches group close to the diagonal. In contrast, in the objective and apparent reflectance groups (a) overall constancy is higher in both tasks, (b) variability between subjects is fairly high, and (c) the selection-based matches systematically deviate from the adjustment-based asymmetric matches: All matches in the Figure 4D lie above the identity line, indicating that selection-based constancy is higher than constancy measured with the asymmetric matching task.

To contrast the patterns of matches across instructional groups we also compared the degree of constancy across tasks (averaged across targets and illuminants) via a repeated measure ANOVA (with instructional group as a between-subjects factor and task as a within-subject factor). For all instructional groups, overall constancy was higher in the color selection than for the asymmetric matching task (main effect of task: $F(1, 12) = 56.61$, $p < .001$). This difference was small for the neutral and physical spectrum groups (0.05 and 0.04, respectively, relative to typical *CCI* range of variation of 0 to 1) but quite pronounced for the objective and apparent reflectance groups (0.24 and 0.53) leading to a significant Task × Instructions interaction, $F(3, 12) = 16.94$; $p < .001$. Consistent with our main findings, there was also a main effect of instructions, $F(3, 12) = 7.66$, $p < .01$.

We also examined the difference in performance across instructional groups by examining the pattern of subject matches across sessions. We did this for the color selection data, where subjects completed between four and five sessions; exploring such effects for our asymmetric matching data set was not practical because subjects completed only two illuminant-changed sessions for this task.

Figure 5A plots constancy indices as a function of session for two subjects. For subject iul (neutral instructions) constancy does not vary systematically across sessions (left panel). This is typical of all of the subjects in the neutral and physical spectrum groups. In contrast, for subject zdc (apparent reflectance instructions) constancy increases systematically with session (right panel); two other subjects (also from the apparent reflectance group) showed a similar pattern.

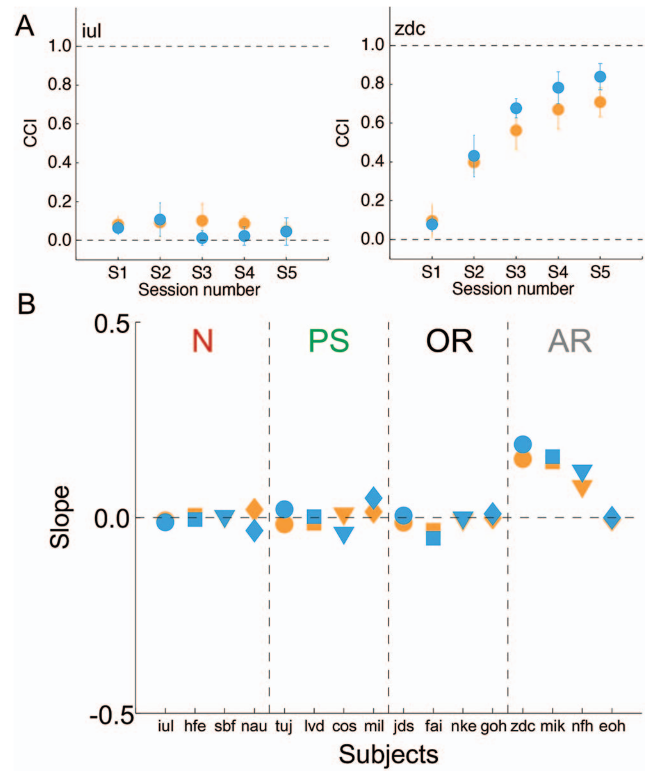

Figure 5. Experiment 1: Selection-based constancy indices across sessions. Panel A shows examples of two distinct patterns of responses in the color-selection task. Mean selection-based indices (averaged over targets) as a function of session are shown for subjects iul (neutral instructions; left) and zdc (apparent reflectance; right). Error bars represent +/−1 SEM. Panel B shows slope for linear fits to each subject's data across sessions. The yellowish illuminant-changed condition is shown in orange (light gray); the bluish illuminant-changed condition in blue (middle gray). See the online article for the color version of this figure.

To quantify the stability of constancy for each subject, we fitted a line to the subject's selection-based constancy indices as a function of session. The slope of this line then characterizes any systematic linear change. For each subject and illuminant condition these slopes are shown in Figure 5B. The slopes are close to zero for most subjects, indicating stable constancy across sessions. The slopes are high for three of the subjects in the apparent reflectance group. Interestingly, for two of these subjects (zdc and mik) the variation in constancy across sessions spanned practically the entire available range, with low constancy in the first session, similar to the levels measured in the neutral and physical spectrum instructions groups, and excellent constancy in the last session.

Plots showing selection-based constancy indices across sessions for each subject are available in the supporting material and include the lines fitted to the data. There we also provide the plots of selection-based matches across sessions for all illuminant conditions; there are only two to three sessions to compare in the illuminant-constant condition, but these matches appear fairly stable for all subjects.

Although we find differences in constancy with instructions, it seems unlikely that these are due to individual or group differences in how precisely the subjects performed our experimental tasks. In the illuminant-constant condition, in which the selections or adjustments are made under uniform illumination, subject matches approximate the target well and their precision, measured as the match-to-target distance in CIELAB ΔE, does not significantly differ across instructional groups. For both the color-selection task and the asymmetric matching task a two-way repeated-measures ANOVA (with instructional group as a between-subjects factor and target reflectance as within-subject factor) failed to reveal a significant main effect of instructions on the match-to-target precision or Instruction × Target interaction (see supporting material), although in the asymmetric matching task the overall precision of matches differed across targets (main effect of target: $F = 6.45$, $p < .01$). Moreover, in both tasks the absolute precision of the illuminant-constant matches was good in an absolute sense: Mean ΔE across subjects was 1.1, (ranging from 0.6 to 2.1 ΔE) for the color selection and 1.6 (varying from 0.9 to 2.3 ΔE) for the asymmetric matching task.

The instructional effects we find (0.59 in the color selection task; 0.24 in the asymmetric matching task; computed as the difference between neutral/physical spectrum groups and the objective/apparent reflectance groups) are comparable with those previously reported for studies that used similarly simple chromatic stimuli and simultaneous asymmetric matching (mean 0.34, range: 0.19–0.57, computed across nine different experiments from five different studies, see also Foster, 2011). Interestingly, the overall constancy in our asymmetric matching task was somewhat lower for both instructional categories (0.06 vs. 0.30) than in the previous studies (0.22 vs. 0.57), possibly due to differences in stimuli we used (which featured a textured background and where the illumination change was presented within a continuous image, rather than across two separated images).

Experiment 2

In Experiment 2, we measured the effect of instructions for the color selection and the asymmetric matching task using more naturalistic stimuli. The basic logic of Experiment 2 was the same as for Experiment 1.

Method

Apparatus. Subjects viewed the stimuli stereoscopically on a stereo-rig consisting of a pair of calibrated 24” NEC MultiSync PA241W LCD color monitors. Each monitor was driven via a dual-port video card (NVIDIA GeForce GT120) at a pixel resolution of 1,920 × 1,200, a refresh rate of 60 Hz and with 8-bit resolution for each RGB channel. The subject viewed the stimuli through a pair of rectangular apertures (2.7 cm × 2.5 cm each) in a black metal plate. The horizontal distance between the centers of apertures was 6.4 cm; they were positioned so that left screen was visible only to the left eye while the right screen was visible only to the right eye. The optical distance of each monitor to the eye was 76.4 cm. Further detail on the apparatus is available in Lee and Brainard (2014).

Stimuli. Stimulus scenes consisted of a rendered room with a large cube in the center. Three visible sides of the cube were covered with a checkerboard like pattern (a 7 × 7 array of squares). The cube also had three distinct buttons (see Figure 6)—one on the right and two on the left side. On illuminant-constant trials, both sides of the cube were rendered under the same illumination (the standard illuminant; Figure 6, left). On illuminant-changed trials only the right side of the cube was rendered under the standard illuminant, while the illumination of the left side changed to yellowish on one half of the trials (Figure 6, center) and to bluish on the other half (Figure 6, right).

The button in the center of the right side was the target for each trial. In the color selection task the two buttons on the left side were chosen from the predefined set of competitors and the subject’s task was to indicate whether the top or the bottom button appeared closer in color to the target. The competitor sets were constructed in the same way as in Experiment 1, except they included an additional color sample (C_5; Figure 7B) in the illuminant-changed condition. This competitor was “an overconstancy match;” it was located on the same line in CIELAB and at the same distance from the target as the competitor C_3, but in the

Figure 6. Experiment 2: Stimuli. Example stimuli for the color selection task. The button in the center of the right side of the cube is the target, while the two buttons on the left side are the competitors. In the illuminant-constant trial (left), the bottom button is the tristimulus and reflectance match for the target while the top one is a competitor (C_1). In the yellowish illuminant-change trial shown here (center) the bottom button is the tristimulus match, while the top one is the reflectance match. In the bluish illuminant-change trial (right), the top button is the tristimulus match while the bottom one is the reflectance match. See the online article for the color version of this figure.

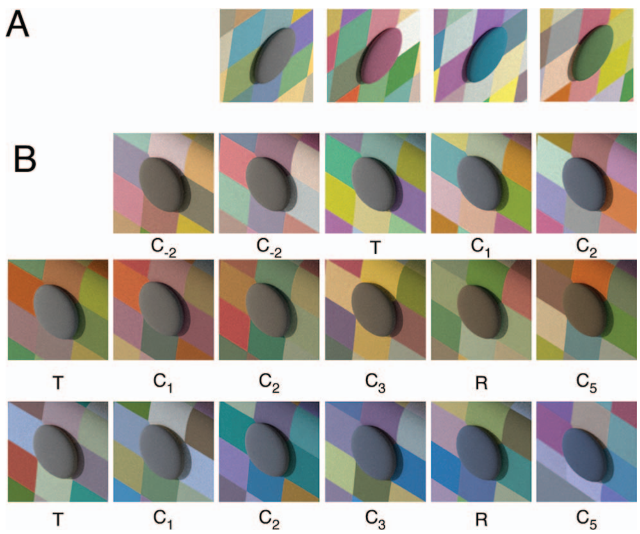

Figure 7. Experiment 2: Targets and competitors. Panel A. The four targets (rendered under the standard illuminant). Panel B. Competitor set for the color selection task for one target (gray) in the illuminant-constant (top row) and the illuminant-changed conditions (middle row: yellowish; bottom row: bluish). See text for detail on how the competitor sets were constructed. See the online article for the color version of this figure.

opposite direction in color space. As in Experiment 1, the XYZ coordinates of the standard illuminant [90.23, 95.33, 102.76], which differed slightly across experiments, were used as a white point for conversion of XYZ to CIELAB values.

In the asymmetric matching task, one of the two buttons on the left side of the cube (randomly selected on each trial) served as the test and was set to either black or white at the beginning of the trial. The other button was assigned a Munsell paper reflectance, selected randomly from the subset of Munsell papers used to tile the sides of the cube. The subject's task was to adjust the test button to match the target in color. As in Experiment 1, for both color selection and asymmetric matching tasks, the exact formulation of the task depended on the instructional group.

In both tasks, the checkerboard pattern covering the sides of the cube was created at the time of stimulus rendering by randomly sampling from a preselected set of Munsell papers (203 Munsell samples selected using the same criteria in Experiment 1). On each trial we randomly selected 25 different samples from this set and we assigned their surface reflectances to the visible checks (with replacement, but with the constraint that no two adjacent checks were the same).

The stimuli were modeled in Blender (http://www.blender.org), an open-source software package that enables three-dimensional scene modeling, and were rendered using Mitsuba (https://www.mitsuba-renderer.org/), an open-source package that uses ray-tracing techniques to provide physically accurate image synthesis. RenderToolbox3 routines (Heasly, Cottaris, Lichtman, Xiao, & Brainard, 2014, https://github.com/DavidBrainard/RenderToolbox3/wiki) were used to facilitate the rendering and to specify the reflectance of each surface and the spectral power distribution of each illuminant in the scene. Each stimulus image was rendered from two viewpoints, corresponding to the left and right eye position.

The target XYZ values were identical to those in Experiment 1. The chromaticity and luminance of the standard and the test illuminants were closely matched across experiments: A supplemental table (A1) in our previous paper (Radonjić et al., 2015b) shows illuminant chromaticity and luminance values for each illumination condition at the target and test locations for both experiments. That paper also provides additional detail about the techniques we used to match the illuminant spectra across experiments, find desired reflectance for the targets and the competitors and efficiently implement the adjustment procedure without needing to rerender the stimuli after each adjustment step.

Experimental procedures were essentially the same as in Experiment 1. In the color selection task each illuminant-constant block of trials consisted of 40 trials (1 standard illuminant × 4 targets × 10 competitor pairs) and each illuminant-changed block consisted of 120 trials (2 test illuminants × 4 targets × 15 competitor pairs; bluish and yellowish trials intermixed). Each subject completed 30 blocks of trials in each illuminant condition across 6–8 hr-long experimental sessions. The sessions were blocked by illuminant condition and followed the same sequence as in Experiment 1.

In the asymmetric matching task, subjects completed six blocks of trials in each illumination condition (four trials per block in illuminant-constant and 8 trials per block in illuminant-changed condition) across four to seven 1-hr sessions. In general, the sessions were blocked by illuminant condition. The first session always consisted of illuminant-constant blocks and was typically followed by alternating illuminant-changed and illuminant-constant sessions. For one subject (mbn) one of the illuminant-constant blocks in the first session was excluded from the analysis due to large deviations between the target and the match, suggesting difficulties in generalizing from the training to experiment. In the following session, this subject completed all remaining illuminant-constant blocks of trials, before moving to the illuminant-changed blocks and an additional block of trials was run to make up for the excluded one (deviations of all remaining matches were within a normal range).

Instructions. We used the same four types of instructions as in Experiment 1. Their formulation was essentially identical as in Experiment 1, with small modifications due to the change in stimulus and selection procedure (e.g., "*click on* the test *square*" was replaced by "*choose* the test *button*"). Rather than using a mouse, in color selection task of Experiment 2 subjects used buttons on a game controller to provide responses. Instructions verbatim for the color selection and the asymmetric matching tasks are provided in the supporting material. As in Experiment 1, four different subjects were assigned to each instructional group. Within a group, half of the subjects first completed the color selection task while the remaining half first completed the asymmetric matching task.

The induction procedure and its schedule was the same as in Experiment 1. The asymmetric matching training was the same as well (done on the same single screen viewed binocularly), but the background behind the target and the test square was changed to match the illuminant-constant background in Experiment 2. Examples of training trials and training instructions verbatim for both experiments are available in the supporting material. Two subjects in Experiment 2 (vlz and azi) took a break of 4 weeks or more between two asymmetric matching sessions. To ensure they remembered how to use the controller and set desired matches after such a long break, these subjects completed one to two training sessions before resuming the experiment.

At the end of the study each subject completed a follow-up questionnaire about the tasks (slightly modified from the questionnaire used in Experiment 1; see supporting material online) and three additional questionnaires which are not analyzed here: the aesthetic experience questionnaire (Chatterjee, Widick, Sternschein, Smith, & Bromberger, 2010), the vividness of visual imagery test (Marks, 1973), and the visualizer-verbalizer questionnaire (Kirby, Moore, & Schofield, 1988).

Subjects. Sixteen subjects (six male and 10 female, all 19–30 years of age) participated in the experiment. All subjects had normal color vision, normal or corrected to normal visual acuity (20/40 or better) and good stereovision (as assessed using a testing procedure developed in our lab, Lee & Brainard, 2014).

Two additional subjects (aab from the neutral and svc from the physical spectrum instructional condition; both male, age 18) were excluded from the study after completing the color selection task and did not participate in the asymmetric matching task. The subjects were excluded based on the preliminary data analysis which showed that they completed more than 10% of trials in less than 300 ms, which we consider to be minimal time required to view and process stimuli (aab: 19.3%; svc: 12.3%; their raw data is available in the supporting material). For the remaining subjects, there were only a few, if any, such trials (three for gmz and two for mdd, vnc, fvh, and hsc out of 4,800 trials); these individual trials were excluded from further analysis.

Results

Figure 8A shows mean selection-based matches for each illumination condition of Experiment 2. Figure 8B shows the selection-based constancy indices for the illuminant-changed conditions. When stimuli were naturalistic, the effect of instructions for the color selection task was significant (main effect of instructions: $F(3, 12) = 4.47$, $p < .05$), but weaker than that we measured for simplified stimuli. The levels of constancy across groups followed a similar pattern as for Experiment 1: Constancy was lower in the neutral and physical spectrum groups (mean *CCI*: 0.47 and 0.42, respectively) than in the objective reflectance and apparent reflectance groups (mean *CCI*: 0.52 and 0.79). Figure 9A plots group means in the selection task (filled bars) and the corresponding 90% bootstrapped confidence intervals. The figure shows that the differences between groups are considerably smaller than those we found with simple stimuli. Examining confidence intervals suggests that constancy was higher in the apparent reflectance group than in any of the other three instructional groups. Although, as in Experiment 1, constancy in the objective reflectance group was higher than in the neutral and physical spectrum groups, this difference was small relative to the confidence intervals. And in contrast to Experiment 1, constancy in the objective reflectance group was more similar to that for neutral and physical spectrum instructions groups than for the apparent reflectance group.

Instructions also had a significant effect on constancy for the asymmetric matching task (Figure 8C; main effect of instructions: $F(3, 12) = 5.23$, $p < .05$). Comparing the means and bootstrapped confidence intervals across groups (open bars, Figure 9A), however, shows that the differences between groups were small and

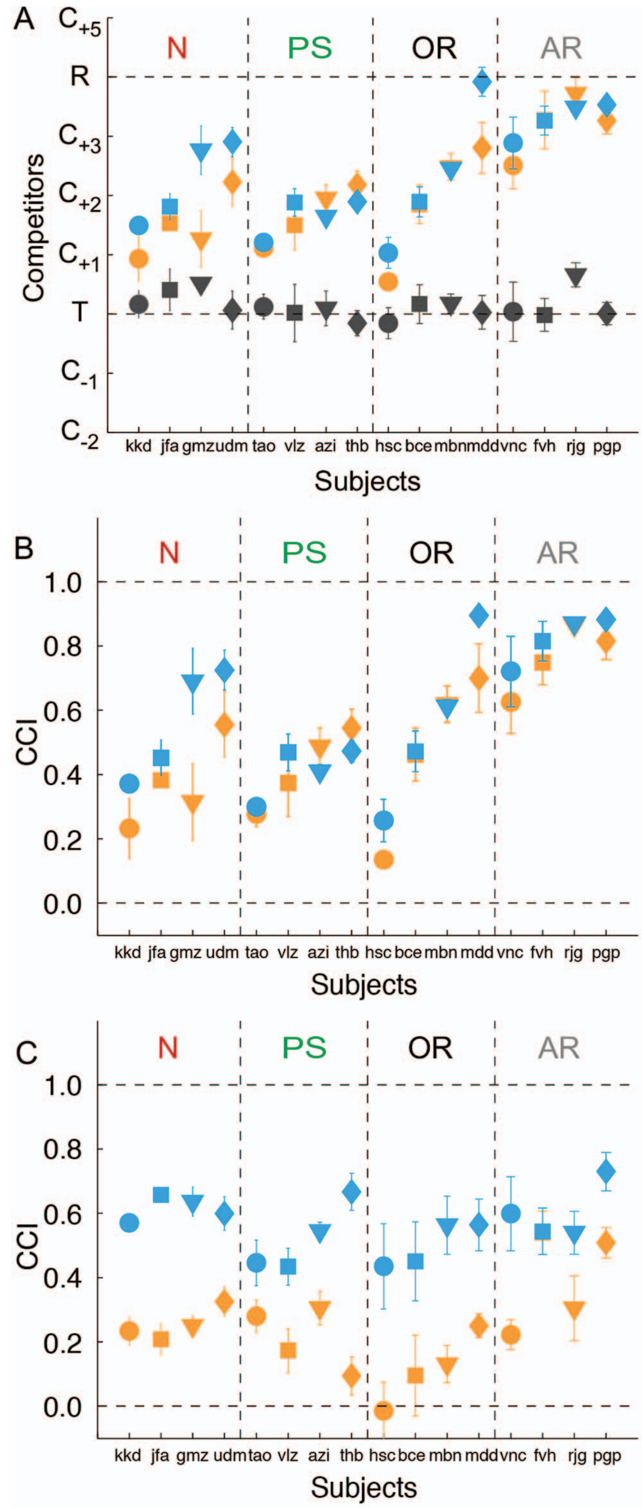

Figure 8. Experiment 2: Results. Panel A shows mean position of the subject's selection-based match (averaged across targets) relative to the space of competitors. For each subject selection-based color constancy indices are shown in Panel B and adjustment-based color constancy indices in Panel C. The figure follows the same representational conventions as Figure 3. See the online article for the color version of this figure.

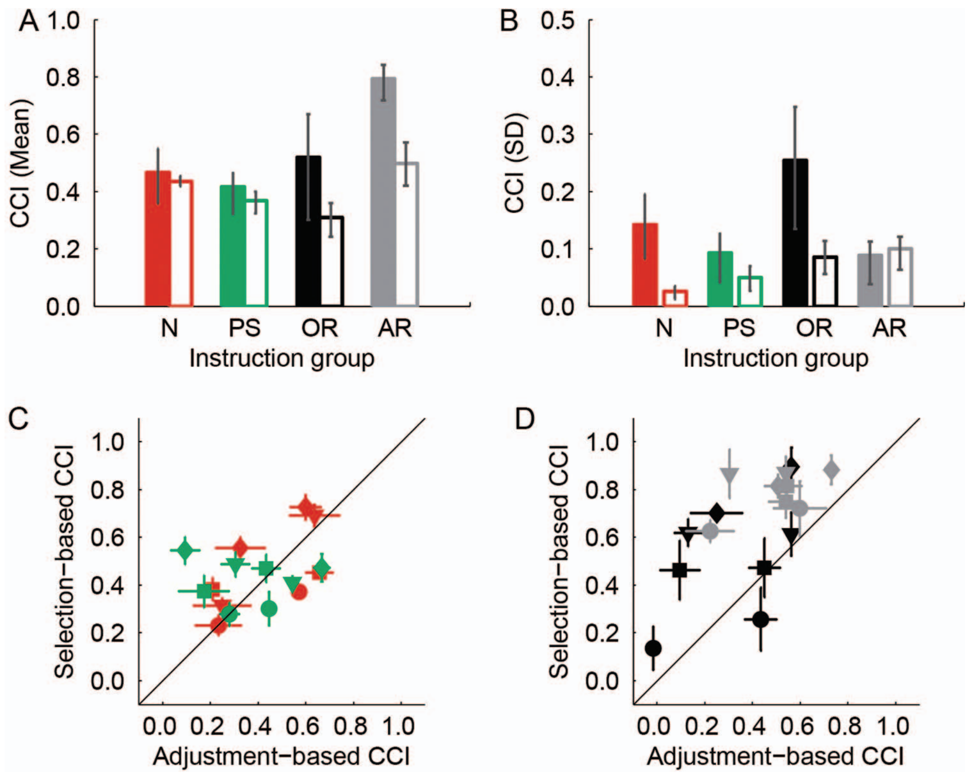

Figure 9. Experiment 2: Comparison of selection-based and adjustment-based matches across instructional groups. Panel A shows mean constancy indices and Panel B mean standard deviations for each group. Panel C plots mean selection-based indices against adjustment-based indices for the neutral and physical spectrum instruction groups. Panel D shows the corresponding plot for objective reflectance and apparent reflectance groups. The figure follows the same representational conventions as Figure 4. See the online article for the color version of this figure.

that they did not follow the pattern that we found for the color selection task or for both tasks in Experiment 1. Here, constancy was lower in the physical spectrum and objective reflectance groups ($CCI = 0.31$ and 0.37, respectively) than in the neutral and apparent reflectance groups ($CCI = 0.44$; $CCI = 0.50$). Furthermore, examining confidence intervals suggest that the difference in constancy between the neutral and the apparent reflectance groups on one hand and the objective reflectance and the physical spectrum groups on the other was significant. However, the difference between physical spectrum and objective reflectance groups was small relative to their confidence intervals, as was the difference between neutral and the apparent reflectance groups.

Comparing the standard deviations across groups (see Figure 9B) indicates that the differences in variability across groups are smaller than in Experiment 1, and do not show the systematic pattern we found in that experiment. In the color selection task, overall variability was the highest for the objective reflectance group and, judging from the bootstrapped confidence intervals, significantly higher than for both the physical spectrum and the apparent reflectance groups. In the asymmetric matching task, the overall variability was the lowest for the neutral instructions group and significantly lower than in the two reflectance groups.

One characteristic of the Experiment 2 data, which we did not observe in Experiment 1, is that for both tasks constancy was significantly higher for the bluish than for the yellowish illuminant change; this effect was particularly pronounced for the asymmetric matching task (Figure 8C; main effect of illuminant in color selection: $F(1, 36) = 11.78$, $p < .01$; in asymmetric matching: $F(3, 36) = 97.94$, $p < .001$). Currently, we do not have good explanation for this asymmetry. However, a number of other studies have reported a similar advantage in constancy for blue (relative to yellow) changes in illumination (Delahunt & Brainard, 2004; Pearce, Crichton, Mackiewicz, Finlayson, & Hurlbert, 2014; Winkler, Spillmann, Werner, & Webster, 2015). Better understanding of this result and its generality remains an open question for future research.

Across illuminants, constancy also varied across targets (Target × Illuminant interaction: $F(3, 36) = 9.30$, $p < .01$ in color selection; $F(3, 36) = 6.85$, $p < .01$ in asymmetric matching; main effect of target in asymmetric matching: $F(3, 36) = 6.77$, $p < .01$). For both tasks, however, these effects were independent of instructions: ANOVAs failed to reveal a significant Target × Instructions, Illuminants × Instruction, or Target × Illuminants × Instructions interaction (there was a trend toward Illuminant × Instruction interaction in the color selection task, $F(3, 36) = 3.08$, $p = .07$; values for all statistical tests are provided in the supporting material).

Figure 9C and D compare constancy across the two tasks. Although constancy variation with instructional group followed a different pattern than in Experiment 1, for consistency we plot the neutral and physical spectrum groups in one panel (9C) and the two reflectance groups in another (9D). The effects revealed by these plots are less pronounced than their counterparts for Experiment 1 (Figure 4C, D), but the broad patterns are similar. In particular, (a) the agreement across tasks was better for the neutral/physical spectrum group than for the two reflectance groups, and (b) in the reflectance-instructions category, constancy measured in the color selection task was systematically higher than that measured in the asymmetric matching task (15 out of 16 subject/illuminant-change comparison points lie above the diagonal).

Although, as in Experiment 1, overall constancy was higher in the color selection than in the asymmetric matching task for all instructional groups (main effect of task: $F(1, 12) = 21.94$; $p = .001$) this difference across tasks was small for the neutral and physical spectrum instructions groups (0.03 and 0.05, respectively), but fairly large for two reflectance groups (objective: 0.21; apparent: 0.29), leading to a significant Task × Instruction interaction, $F(3, 12) = 4.21$, $p < .05$. The main effect of instructions was also significant, $F(3, 12) = 4.73$, $p < .05$.

We also examined the stability of selection-based constancy across sessions, but did not observe the marked systematic change across sessions that occurred for some subjects in Experiment 1. Figure 10A shows constancy indices across sessions for two subjects from different instructional groups (udm: neutral; mdd: objective reflectance). Figure 10B shows slopes of the line fitted to each subject's constancy indices across sessions in the two illuminant conditions.

As in Experiment 1, the precision of subject matches in the illuminant-constant condition did not differ significantly across instructional groups in either task (main effect of instructions and Instruction × Target interaction were not significant; $F < 2$, *ns*), although, as in Experiment 1, there was some variation in precision of asymmetric matches across targets (main effect of target: $F(3, 36) = 5.05$, $p < .01$). In terms of absolute precision the matches were reasonably good for both tasks: mean match-to-target distance across subjects was 2.3 ΔE in the color selection task (ranging from 1.1 to 3.2 ΔE) and 3.6 ΔE in the asymmetric matching task (ranging from 2.0 to 5.8 ΔE). Interestingly, however, the overall match precision was worse than in Experiment 1. A repeated measure ANOVA with target as within-subject and experiment and instructional group as between-subjects factors revealed a main effect of experiment in both color selection, $F(1, 24) = 35.56$, $p < .001$ and asymmetric matching, $F(1, 24) = 41.13$, $p < .001$.

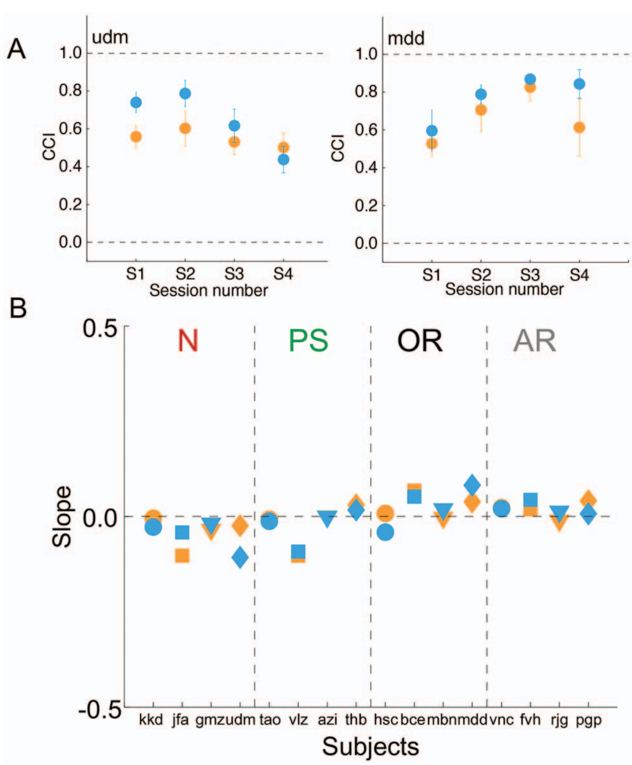

Figure 10. Experiment 2: Selection-based constancy across sessions. Panel A shows constancy indices across sessions for two sample subjects from different instructional groups (left: udm from neutral instructions group; right: mdd from objective reflectance group). Panel B shows slope for linear regression fits to subject's data across sessions. The figure follows the same representational conventions as Figure 5. See the online article for the color version of this figure.

Discussion

We measured instructional effects on color constancy using four different types of instructions, two different tasks and two classes of stimuli. Consistent with previous reports, we find that instructions can modulate experimentally measured constancy. This result holds for both simplified and naturalistic stimuli and for both the color selection and the asymmetric matching tasks. The key novel result we report is that the size of the instructional effects varies both with task and class of stimuli.

Simplified stimuli. When the stimuli were simplified (Experiment 1), varying instructions had a large effect on constancy: subjects who were asked to make task judgments based on the reflectance properties of the test patches made matches that indicated excellent constancy. In contrast, the matches of subjects who were asked to make judgments based on the reflected light, or who received nonspecific neutral instructions, indicated poor constancy. This difference was particularly large for the color selection task, where constancy indices across different instructional conditions spanned practically the full range of 0 to 1.

To evaluate the effect different types of instructions had on constancy we compared the pattern of results for each instructional group to that we obtained for neutral instructions. In the neutral instructions group (a) overall constancy was low for both the color selection and the asymmetric matching task, (b) variability between subjects was also low, (c) selection-based and adjustment-based constancy measures were in good agreement, and (d) selection-based constancy remained stable across sessions. In the physical spectrum instructions group subject performance was essentially identical to that in the neutral group, but in the two reflectance instructions groups it differed considerably. For these groups, (a) overall constancy was high, (b) intersubject variability was also high, (c) selection-based constancy deviated systematically from adjustment-based constancy, and (d) for some subjects, the selection-based constancy increased systematically across sessions.

We believe this pattern of responses across instructional groups speaks against the hypothesis that different instructions probe different aspects of a common perceptual representation or evoke different perceptual modes. Instead, our results suggest that when asked to make reflectance-based judgments about simplified stimuli subjects employ a different task strategy—one based on explicit reasoning. We base this conclusion primarily on the findings that different types of instructions had different effects for the two tasks we used. If different instructions simply probed different aspects of common underlying perceptual representation (or a different perceptual mode), then constancy for the two tasks should be similar within each instructional group. Indeed, this prediction might be taken as an operational definition of what it means for performance to be based directly on a perceptual representation. We do find a good agreement between constancy across tasks for the neutral (as well as physical spectrum) instructions group, but not for the two reflectance groups.

Further, if the reflectance-based instructions simply probed a different perceptual representation than the neutral or the physical spectrum instructions, we would expect to find a different degree of constancy for the reflectance groups, but one which is stable across sessions. This prediction is clearly violated by the subset of subjects who show large shifts in constancy across sessions. For these subjects performance is more consistent with a gradual shift in strategy over the course of the experiment, elicited by repeated delivery of instructions at the beginning of each session. Our interpretation is that in the early sessions subject's performance is based on the perceptual representation which, consistent with the results from the neutral instructions group, indicates poor constancy for our simplified stimuli. In contrast, the later sessions reveal the subject's evolving instruction-driven tendency to reason about reflectance from that same perceptual representation.

Some previous studies speculated that under reflectance-based instructions subjects tend to match "what they should see" instead of "what they do see," which leads to "higher uncertainty of matches" (Troost & de Weert, 1991). This is consistent with our observations and can account for the finding that constancy was systematically higher in the color selection task. For a subject who is trying to match how a surface would look under different illumination, it is fairly easy to develop a strategy that would guide matches in the color selection task: They would simply need to select the competitor that shows a larger perceptual shift in the direction of the illumination change. Relying on this strategy could lead to overconstancy; and, in fact, examining across-sessions data revealed overconstancy for some selection-based matches in at least one session for seven (out of eight) subjects in the reflectance groups. In contrast, in the asymmetric matching task, it may be harder for subjects to decide how far they need to compensate in the direction of the illumination change to achieve the prompted-for reflectance match. We found no overconstancy asymmetric matches in any sessions for any subjects.

Previous studies which found large instructional effects using naive subjects reported that some training is necessary for subjects to follow and understand the task instructions (Bauml, 1999; Reeves et al., 2008; Troost & de Weert, 1991; see also Doerschner, Boyaci, & Maloney, 2007). This is certainly consistent with our experience. In our study we employed extensive training of naive observers combined with a strict instructional regime to achieve instructional effects. This included: (a) an induction procedure, in which the difference between surface reflectance and reflected light was both demonstrated and verbally explained; (b) lengthy task instructions in which subject's task was reiterated several times and explicitly related to what was learned in the induction procedure; (c) repeating the induction procedure before the first illuminant-changed session; and (d) repeating task instructions at the beginning of each experimental session.

Even with our instructional regime, not all subjects in the reflectance instructions group showed high constancy. Matches for one subject (jds; objective reflectance instructions) were low in both tasks and conformed to the pattern we found for the neutral and physical spectrum groups. For another two subjects (zdc and nfh) constancy indices in the asymmetric matching task also fell within range of those measured in the neutral/physical spectrum groups, although their selection-based indices were considerably higher. In sum, introducing reflectance-based instructions led to large intersubject variability in both tasks. We attribute this to different degrees to which subjects in the reflectance instructions groups use strategies based on reasoning and different degrees of success that their idiosyncratic strategies have in achieving constancy. Although large intersubject variability in the reflectance groups could occur due to individual differences, it would be unlikely if all subjects who tend to exhibit high constancy were randomly assigned to the reflectance groups and it would be hard to explain why they would exhibit high constancy in one task but low in another.

Our hypothesis that subjects in the reflectance groups employed strategies based on explicit reasoning is supported by some of the reports provided in the postexperiment questionnaires. This information has a fairly informal character but we believe that, in combination with the rich behavioral data we report, it can inform our understanding of the nature of instructional effects (Wagner, 2012). When asked to describe the strategy they used to complete the two experimental tasks, descriptions provided by subjects in the neutral and physical spectrum group are generally procedural and simply describe the actions they took to complete the task:

> I would narrow the choice down to two [tests] and then compare them against one another by looking back and forth between each one and the target square in the middle"(subject hfe, neutral instructions; color selection task).

> . . . I would first try to match intensity. Then I would increase the chroma to get the right color range. Then I would decrease adjustment size and adjust until I had a good match. I also tried to focus on the space between the target and adjusted each square such that both were out of focus, than I could tell better if they were a match" (subject iul, neutral instructions, asymmetric matching task).

In contrast, reports of subjects in the reflectance groups often revealed the use of mental imagery and explicit reasoning.

> . . . I would imagine the color was projecting a light onto the colored square. With this imaginary veil I would quickly scan the options and then pick the color I felt looked closest (subject nfh, apparent reflectance instructions, color selection task).

> In trails [sic] where the illumination changed, I tried to imagine as if the outer ring of illumination was placed over the center square. I also looked at the ringer where the circle changed, to see the effect that it

> had on the background colors. That was especially useful in the second experiment, where there was more precision involved. With the first experiment, generally there was only one square that was in the general region of what I thought was correct. However, with the second experiment, it was very difficult to determine what the center square would look like under the changed illumination, so I had to think more about it (eoh; apparent reflectance instructions; completed color selection task first; comment provided in response to the experimenter's request for clarification of initial response on questionnaire).

Naturalistic stimuli. When the stimuli were naturalistic, the size of the instructional effects was smaller, but the data still revealed the behavioral signature of instructional effects that we found in Experiment 1. Most of the patterns in the data that, we argue, are best understood as resulting from two distinct task strategies—one purely perceptual and the other based on explicit reasoning—are still present in Experiment 2. In particular:

1. In the color selection task constancy was higher for the apparent reflectance groups than for the neutral and physical spectrum groups. Although still substantial, this difference was half as large as that we found with simplified stimuli (mean *CCI*s: 0.67 vs. 0.35).

2. The agreement between the selection-based and adjustment-based constancy measures was better for the neutral/physical spectrum groups than for the two reflectance groups.

3. In the two reflectance groups, selection-based constancy was systematically higher than adjustment-based constancy.

4. There were large differences in constancy between subjects in the reflectance groups in both tasks. These differences spanned practically the whole constancy range, suggesting that these subjects differ in extent to which they rely on instruction-driven reasoning to make matches as well as the degree of constancy they can achieve using their idiosyncratic strategy.

As in Experiment 1, comments of some subjects in the reflectance groups provided in the postexperiment questionnaires reveal conscious attempts to compensate for the effect of illumination change when making matches. For example:

> First I decided whether the illumination was the same on the two sides of the cube. If it was the same, I looked between the target and the test button several times to decide which matched. If the illumination was different, I tried to identify how: I identified types of illumination by lighter or darker, more yellow or more blue. Then I tried to keep the image of the target button in my mind's eye while imagining how it would look under that type of illumination to find the right test button (subject rjg, apparent reflectance instructions, color selection task, last sentence in response to experimenter's request for clarification).

In Experiment 2 comments on the illumination also appear for subjects in the neutral and physical spectrum group, but it remains unclear whether that is due to the stimulus, which conveyed a fairly realistic depiction of a multilit scene, due to the fact that the postexperiment questionnaire was modified for Experiment 2 to include explicit questions about the illumination (see supporting material), or due to the fact that the experimenter more routinely asked for clarification of some subject's responses after the questionnaire had been completed, or some combination of these factors. When explicitly asked to discuss whether and how they consciously considered the change in illumination when making matches, most subjects in the neutral/physical spectrum groups reported that they noticed occasional changes in illumination in the experiment but that they either did not consider these when making matches or that they tried to consciously ignore them. For example:

> I often pondered if I should select a color or adjust a color to compensate for different illumination, but I thought it would be best if I made decision based on my observations instead of what I believed would compensate for different illumination situations (subject jfa, neutral instructions).

Interestingly, comments of one subject from the neutral instructions group revealed considerations of illumination that suggests the use of strategies similar to those described by the subjects in the reflectance groups (but note that the compensation described by this subject was in the direction opposite of constancy).

> Usually if the background was brightly illuminated I felt that the brighter circle matched the target more closely (in the second task). In the first task, I took it into considerations while making adjustments; since different faces of the cube had various "shadows:" If the target was shadowed, I ended up making the test button brighter to account for the difference (subject kkd, neutral instructions, completed asymmetric matching first).

The complete set of subjects' responses to the questionnaire is provided in the supporting material online.

Returning to the data, the smaller size of the instructional effects we found in Experiment 2 may result from the fact that the naturalistic stimuli produce a perceptual representation that is more color constant than the representation produced by simple stimuli. This would leave less room for subjects to employ strategies based explicit reasoning, and, thus, result in smaller instructional effects.

There were some characteristics of the data from Experiment 1 that we did not find in Experiment 2. First, we did not find a systematic change in constancy across sessions for any of the subjects in Experiment 2. This might be because the reduced room for employing strategies based on reasoning made any such effects too small to measure. Second, intersubject variability in constancy was not systematically higher for the two reflectance groups than for the neutral/physical spectrum groups, although in the asymmetric matching tasks we recorded some trends in that general direction. Finally, there were some notable differences in the pattern of constancy across instructional groups in Experiment 2, when compared with Experiment 1, for which we do not have an explanation (e.g., in the asymmetric matching constancy was similarly high in the neutral and apparent reflectance group and higher than in the physical spectrum and objective reflectance groups).

Conclusion

Although we used four different types of instructions, we found only two distinctive behavioral patterns. First, subjects' performance for physical spectrum instructions group was essentially identical to that for the neutral group. We believe that this is because naive subjects have great difficulty dissociating the concept of the appearance of the proximal stimulus from their percept of the stimulus (Gilchrist, 2012; MacLeod, 2012) and therefore simply do the same thing as the subjects who received neutral instructions. These, we believe, tend naturally to elicit judgments based on the percept. It is possible that if we had provided even more extensive training, perhaps of the sort that is delivered in courses on painting and photography, we might have been able to induce a different kind of strategy based on reasoning in the physical spectrum instructions group—one where subjects would adjust their matches in the direction opposite of constancy, resulting in another distinctive behavioral pattern. Effects of this sort would be expected to be largest for stimulus conditions where constancy under neutral instructions is high.

Second, subjects in the objective and apparent reflectance groups also behaved very similarly to each other. In Experiment 1, constancy for these two groups was indistinguishable. In Experiment 2, constancy was significantly higher in the apparent reflectance group, but the large intersubject variability in the objective reflectance group makes us cautious about attributing this result to a real difference between the two types of instructions. Rather, we believe that the subtle difference in wording between the two reflectance instructions is not meaningful to naive subjects.

We recognize that not all theorists will agree that the pattern of the data for the reflectance instructions, particularly in Experiment 1, is most parsimoniously understood as revealing explicit, conscious reasoning on the part of subjects. Thus, a challenge for future research is to extend the results we report here to more sharply define the nature of the processes that generate the patterns we observe. For example, it might be possible to explicitly test the role of reasoning in the reflectance instructions group by asking subjects to perform a concurrent cognitive task while making color matches and measure how such interference interacts with the effect of instructions.

Our study did not examine the potential role that early adaptation processes play in the instructional effects we have measured. Previous research that modeled adaptation processes in an asymmetric matching task in which instructions were varied, however, showed that differences in adaptation that could arise due to different stimulus scanning patterns under different instructions cannot account for the size of instructional effects (Cornelissen & Brenner, 1995).

We want to make a clear distinction between the task strategies based on explicit reasoning we discuss here and perception based on unconscious inferences, a theoretical account proposed by von Helmholtz (1866) that is often invoked to explain mechanisms that support color constancy as well as visual perception more generally. The task strategies we consider are based on conscious and explicit problem solving (*What is the best match I can make given this percept and what I know about the relationship between reflectance and illumination*), while unconscious inference refers to cognitively impenetrable processes which produce the percept. Thus, although we believe that the question of how subjects can reason from their percepts is interesting in its own right, we think it is a different question from how objects appear. Our view is that instructional effects are telling us about the former, and that to get at the latter neutral instructions are the most likely to succeed.

Although our results speak against a dual perceptual representation for the stimuli we used, introspection suggests that some stimuli do produce perceptual scission, where a region of the image can be naturally associated with two distinct, essentially equally vivid percepts. We have observed such scission ourselves, for example, in real object-and-illuminant setups in which manipulating perceived depth can have a large effect on perceived lightness (Gilchrist, 1980; Radonjić & Gilchrist, 2013) or in visual illusions that invoke an impression of haze or transparency (Adelson, 2000; Anderson & Winawer, 2005; Singh & Anderson, 2002). An interesting open question is whether the effects of instructions for such stimuli would show a different pattern from those we observe, perhaps one where large instructional effects are not accompanied by the signature of strategies based on reasoning.

Considerations of such stimuli inspired Logvinenko and his collaborators (Logvinenko & Maloney, 2006; Logvinenko, Petrini, & Maloney, 2008) to explore the dimensionality of perceptual experience via multidimensional scaling method. While this approach is novel and potentially fruitful it is not clear that it eliminates the potential for subjects to respond on the basis of explicit reasoning. In addition, the generality of the results of Logvinenko and colleagues (2006, 2008) remains unclear (Madigan & Brainard, 2014).

In summary, our results suggest that certain types of instructions cause subjects to employ strategies based on explicit reasoning—which are grounded in their perceptions and developed using the information provided in the instructions and training—to achieve the response they believe is requested by the experimenter. The shifts in constancy caused by such instructions are particularly large with simplified stimuli, which have been predominantly used to study instructional effects and which, historically, have been widely used in color constancy studies in general. Our study shows that with such stimuli, constancy is poor when probed with nonspecific neutral instructions (Radonjić et al., 2015b). This may be because the simple stimuli do not provide the cues necessary for the visual system to parse them as illuminated objects seen across a change of illumination (see Figure 1). The more naturalistic stimuli used in Experiment 2 presumably provide more such cues which leads to higher constancy with neutral instructions (see Figure 6).

Although our experiments differed in degree of stimulus naturalness, both sets of stimuli we used were simulated scenes. Thus, characterizing whether and to what degree different instructions modulate constancy in scenes that consist of real illuminated objects remains an open question. In such natural scenes, the degree of constancy is often shown to be high (e.g., Kraft & Brainard, 1999) and, judging by our results, this would leave even less room for the effect of reflectance-based instructions. Our general view, however, is that future studies that use either real stimuli or their naturalistic simulations together with neutral instructions are those that will lead to a characterization of perceptual constancy that is representative of natural viewing.

References

Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 339–351). Cambridge, MA: MIT Press.

Anderson, B. L., & Winawer, J. (2005). Image segmentation and lightness perception. *Nature, 434,* 79–83. http://dx.doi.org/10.1038/nature03271

Arend, L. E., & Goldstein, R. (1987). Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America, 4,* 2281–2285. http://dx.doi.org/10.1364/JOSAA.4.002281

Arend, L. E., Jr., Reeves, A., Schirillo, J., & Goldstein, R. (1991). Simultaneous color constancy: Paper with diverse Munsell values. *Journal of the Optical Society of America, 8,* 661–672. http://dx.doi.org/10.1364/JOSAA.8.000661

Arend, L., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America, 3,* 1743–1751. http://dx.doi.org/10.1364/JOSAA.3.001743

Arend, L. E., & Spehar, B. (1993a). Lightness, brightness, and brightness contrast: 1. Illuminance variation. *Perception & Psychophysics, 54,* 446–456. http://dx.doi.org/10.3758/BF03211767

Arend, L. E., & Spehar, B. (1993b). Lightness, brightness, and brightness contrast: 2. Reflectance variation. *Perception & Psychophysics, 54,* 457–468. http://dx.doi.org/10.3758/BF03211768

Bäuml, K. H. (1999). Simultaneous color constancy: How surface color perception varies with the illuminant. *Vision Research, 39,* 1531–1550. http://dx.doi.org/10.1016/S0042-6989(98)00192-8

Blakeslee, B., & McCourt, M. E. (2015). What visual illusions tell us about underlying neural mechanisms and observer strategies for tackling the inverse problem of achromatic perception. *Frontiers in Human Neuroscience, 9,* 205. http://dx.doi.org/10.3389/fnhum.2015.00205

Blakeslee, B., Reetz, D., & McCourt, M. E. (2008). Coming to terms with lightness and brightness: Effects of stimulus configuration and instructions on brightness and lightness judgments. *Journal of Vision, 8,* 1–14. http://dx.doi.org/10.1167/8.11.3

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

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, 14,* 2091–2110. http://dx.doi.org/10.1364/JOSAA.14.002091

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.

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. http://dx.doi.org/10.2190/EM.28.2.f

Cornelissen, F. W., & Brenner, E. (1995). Simultaneous colour constancy revisited: An analysis of viewing strategies. *Vision Research, 35,* 2431–2448. http://dx.doi.org/10.1016/0042-6989(94)00318-1

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

Doerschner, K., Boyaci, H., & Maloney, L. T. (2007). Testing limits on matte surface color perception in three-dimensional scenes with complex light fields. *Vision Research, 47,* 3409–3423. http://dx.doi.org/10.1016/j.visres.2007.09.020

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman and Hall/CRC. http://dx.doi.org/10.1007/978-1-4899-4541-9

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

Gibson, J. J. (1950). *The perception of the visual world*. Westport, CT: Greenwood Press.

Gilchrist, A. L. (1980). When does perceived lightness depend on perceived spatial arrangement? *Perception & Psychophysics, 28,* 527–538. http://dx.doi.org/10.3758/BF03198821

Gilchrist, A. L. (2012). Objective and subjective sides of perception. In G. Hatfield & S. R. Allred (Eds.), *Visual experience: Sensation, cognition, and constancy* (pp. 105–121). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199597277.003.0006

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,* 6. http://dx.doi.org/10.1167/14.2.6

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

Kingdom, F. A. A. (2011). Lightness, brightness and transparency: A quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research, 51,* 652–673. http://dx.doi.org/10.1016/j.visres.2010.09.012

Kirby, J. R., Moore, P. J., & Schofield, N. J. (1988). Verbal and visual learning styles. *Contemporary Educational Psychology, 13,* 169–184. http://dx.doi.org/10.1016/0361-476X(88)90017-3

Koffka, K. (1935). *Principles of Gestalt psychology*. New York, NY: Harcourt, Brace and Company.

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences of the United States of America, 96,* 307–312. http://dx.doi.org/10.1073/pnas.96.1.307

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. http://dx.doi.org/10.1167/14.1.24

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

Logvinenko, A. D., Petrini, K., & Maloney, L. T. (2008). A scaling analysis of the snake lightness illusion. *Perception & Psychophysics, 70,* 828–840. http://dx.doi.org/10.3758/PP.70.5.828

Logvinenko, A. D., & Tokunaga, R. (2011). Lightness constancy and illumination discounting. *Attention, Perception, & Psychophysics, 73,* 1886–1902. http://dx.doi.org/10.3758/s13414-011-0154-2

Lotto, R. B., & Purves, D. (1999). The effects of color on brightness. *Nature Neuroscience, 2,* 1010–1014. http://dx.doi.org/10.1038/14808

MacLeod, D. I. A. (2012). A mechanistic perspective on the given. In G. Hatfield & S. R. Allred (Eds.), *Visual experience: Sensation, cognition, and constancy* (pp. 122–134). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199597277.003.0007

Madigan, S. C., & Brainard, D. H. (2014). Scaling measurements of the effect of surface slant on perceived lightness. *i-Perception, 5,* 53–72. http://dx.doi.org/10.1068/i0608

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *Journal of Vision, 3,* 573–585. http://dx.doi.org/10.1167/3.8.5

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64,* 17–24. http://dx.doi.org/10.1111/j.2044-8295.1973.tb01322.x

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

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,* e87989. http://dx.doi.org/10.1371/journal.pone.0087989

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. http://dx.doi.org/10.1163/156856897X00366

Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2015a). Color constancy in a naturalistic, goal-directed task. *Journal of Vision, 15,* 3. http://dx.doi.org/10.1167/15.13.3

Radonjić, A., Cottaris, N. P., & Brainard, D. H. (2015b). Color constancy supports cross-illumination color selection. *Journal of Vision, 15,* 13. http://dx.doi.org/10.1167/15.6.13

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. http://dx.doi.org/10.1068/i0575

Reeves, A. J., Amano, K., & Foster, D. H. (2008). Color constancy: Phenomenal or projective? *Perception & Psychophysics, 70,* 219–228. http://dx.doi.org/10.3758/PP.70.2.219

Ripamonti, C., Bloj, M., Hauck, R., Kiran, M., Greenwald, S., Maloney, S. I., & Brainard, D. H. (2004). Measurements of the effect of surface slant on perceived lightness. *Journal of Vision, 4,* 747–763. http://dx.doi.org/10.1167/4.9.7

Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.

Rudd, M. (2010). How attention and contrast gain control interact to regulate lightness contrast and assimilation: A computational neural model. *Journal of Vision, 10*. http://dx.doi.org/10.1167/10.14.40

Singh, M., & Anderson, B. L. (2002). Toward a perceptual theory of transparency. *Psychological Review, 109,* 492–519. http://dx.doi.org/10.1037/0033-295X.109.3.492

Troost, J. M., & de Weert, C. M. M. (1991). Naming versus matching in color constancy. *Perception & Psychophysics, 50,* 591–602. http://dx.doi.org/10.3758/BF03207545

Van Es, J. J., Vladusich, T., & Cornelissen, F. W. (2007). Local and relational judgements of surface colour: Constancy indices and discrimination performance. *Spatial Vision, 20,* 139–154. http://dx.doi.org/10.1163/156856807779369733

von Helmholtz, H. (1866). *Treatise on physiological optics* (2nd ed., Vol. 1962). New York, NY: Dover.

Wagner, M. (2012). Sensory and cognitive explanations for a century of size constancy research. In G. Hatfield & S. R. Allred (Eds.), *Visual experience: Sensation, cognition, and constancy* (pp. 63–86). Oxford, UK: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780199597277.003.0004

Winkler, A. D., Spillmann, L., Werner, J. S., & Webster, M. A. (2015). Asymmetries in blue-yellow color perception and in the color of "the dress." *Current Biology, 25,* R547–R548. http://dx.doi.org/10.1016/j.cub.2015.05.004

Wright, W. (2013). Color constancy reconsidered. *Acta Analytica, 28,* 435–455. http://dx.doi.org/10.1007/s12136-013-0187-3

Received September 2, 2015
Revision received November 6, 2015
Accepted November 10, 2015 ■