This paper appeared in Perception & Psychophysics, 70, 219-228 (2008).
To obtain a reprint, email reeves@neu.edu or d.h.foster@manchester.ac.uk

Color Constancy: Phenomenal or Projective?

ADAM J. REEVES

Northeastern University, Boston MA

KINJIRO AMANO and DAVID H. FOSTER

University of Manchester, Manchester, UK

Footnote to title page

This work was supported by the Engineering and Physical Sciences Research Council, UK, Grant GR/S43139/01, and by the Wellcome Trust, Grant 064669/Z/01/Z. We thank K. Żychaluk [Printer: dot over "Z"] for advice. Correspondence concerning this article may be addressed to A. J. Reeves at the Department of Psychology, Northeastern University, 360 Huntingdon Ave., Boston MA 02115, USA (email: reeves@neu.edu).

Editorial correspondence

Dr A. J. Reeves
Department of Psychology
Northeastern University
360 Huntingdon Ave.
Boston MA 02115, USA
email: reeves@neu.edu
Tel: 617-373-4708; Fax 617-373-8714

Suggested running head: Judging Color Constancy

Keywords: color vision, perception, color appearance, color constancy, hue and saturation, cone-excitation ratios.

Abstract

Naïve observers viewed a sequence of colored “Mondrian” patterns, simulated on a color monitor. Each pattern was presented twice in succession, first under one daylight illuminant of color temperature either 16000 K or 4000 K, and then under the other, to test for color constancy. Observers compared the central square of the pattern across illuminants, either rating it for sameness of material appearance, or for sameness of hue and saturation, or judging an objective property, that is, whether its change of color originated from a change in material or only in illumination. Average color-constancy indices were high for material-appearance ratings and binary judgments of origin, and low for hue-saturation ratings. Individuals’ performance varied, but judgments of material and of hue and saturation remained demarcated. Observers seem able to separate phenomenal percepts from their ontological projections of mental appearance onto physical phenomena: thus, even when a chromatic change alters perceived hue and saturation, observers can reliably infer the cause, the constancy of the underlying surface spectral reflectance.

INTRODUCTION

Color constancy refers to the constancy of perceived or apparent surface color under changes of the spectrum of an illuminant, or, in an extended sense, under changes in scene composition or configuration (Judd, 1940; Maloney, 1999). The phenomenon is a challenging one, for in some situations, it seems to involve a paradox: a separation of sensation from judgment. This separation was described by Lichtenberg in 1793 in a letter to Goethe thus: "In ordinary life we call white, not what looks white, but what would look white if it was set out in pure sunlight, or in a light whose quality did not differ much from sunlight. It is more the potential to be white and become white, in all its gradations, that we call white in some object, rather that the pure white colour itself" (Joost, Lee, & Zaidi, 2002, p. 302).

In the laboratory, color constancy is often measured by presenting simultaneously to an observer pairs of differently illuminated, geometric—usually checkerboard—patterns of colored surfaces, simulated on the screen of a computer-controlled color monitor. These checkerboard patterns are called "Mondrians" after their similarity to some of the paintings by Piet Mondriaan. Displays such as these have been used in many different laboratories and have the advantage that they contain no spatial cues to spectral reflectance based on familiar shapes or semantic content (e.g. Hansen, Olkkonen, Walter, & Gegenfurtner, 2006). The observer, while repeatedly looking from one pattern to the other (Arend & Reeves, 1986; Cornelissen & Brenner, 1995), attempts to match the surface color of a square in one pattern against the surface color of the corresponding square in the other pattern. On a continuous scale in which perfect constancy has the value 1 and perfect inconstancy the value 0 (Arend, Reeves, Schirillo, & Goldstein, 1991), reported levels of constancy in simultaneous asymmetric color matching of Mondrians have ranged from about 0.4 to about 0.8 (Amano &

Foster, 2004; Amano & Foster, 2005; Arend, Reeves, Schirillo, & Goldstein, 1991; Bäuml, 1999; Cornelissen & Brenner, 1995; Foster, Amano, & Nascimento, 2001; Lucassen & Walraven, 1996). For a given stimulus geometry, slightly higher scores may be obtained by presenting the two Mondrians to be matched sequentially in the same position, rather than simultaneously side by side (Foster, Amano, & Nascimento, 2001), possibly because of the generation of a transient color signal in the sequential presentation. Levels of constancy with simultaneously presented Mondrians fall within the range obtained in asymmetric matching across two- and three-dimensional physical tableaux (Brainard, Brunt, & Speigle, 1997; de Almeida, Fiadeiro, & Nascimento, 2004). (In general, perfect color constancy is impossible with real surfaces and illuminants because of the phenomenon of metamerism; see Box 1 of Foster (2003)).

The nature of the task given to observers is important. In the measurements by Arend & Reeves (1986) and Arend et al. (1991), later confirmed by others (Bäuml, 1999; Cornelissen & Brenner, 1995; Troost & de Weert, 1991), observers were given two “subjective” color-matching tasks concerned with stimulus appearance. In one, it was to adjust the color of a designated test square in the pattern so that it appeared as if it were “cut from the same piece of paper” as the corresponding standard square in the other pattern, that is, to match its surface color (a so-called “paper” match); in the other, it was to adjust the color of the test square so that its hue and saturation matched those of the standard square (a “hue-saturation” match). The first task produced the moderately high levels of color constancy just mentioned, whereas the second task produced much lower levels, from near zero to about 0.3. [FOOTNOTE 1] Thus, presented with the same stimuli, and presumably the same visual cues, observers could judge appearance in one way given one set of instructions about appearance and in a different way given another set of instructions about appearance.

Evidence of a more "objective" mode of color perception, concerned with what stimuli represent, has come from a different, operational approach to measuring color constancy introduced by Craven & Foster (1992). The task of the observer was to attribute changes in the appearance of a scene either to changes in the spectral composition of the illuminant or to changes in the spectral composition of the illuminant combined with changes in the reflecting properties of the scene, that is, the materials of which it was made. This aspect of color constancy was not concerned with the nature or extent of any changes in color appearance *per se*, but simply with the observer's interpretation of them. In the extreme (not proposed here), an observer could identify a surface as being unchanged under a change in illuminant without necessarily being able to identify the color of the surface itself (Craven & Foster, 1992). Observers were able to perform the task rapidly, reliably, and with little or no training (Foster et al., 2001). Levels of constancy by this operational method applied to Mondrians have been reported as 0.77–0.79 (e.g. Baraas, Foster, Amano, & Nascimento, 2004), within the range obtained with two-dimensional images of natural scenes (Foster, Amano, & Nascimento, 2006). High levels of performance have also been obtained in a related performance-based experimental task in which observers had to discriminate between colored filters placed over patterns of colored surfaces (Khang & Zaidi, 2002).

One potential explanation of these different levels of constancy according to task was anticipated in Arend and Reeves' (1986) paper where they considered the role played by two kinds of constancy process, one which depends on the eye becoming accustomed to the new illuminant and involves both light adaptation (von Kries, 1905; Whittle, 1996) and contrast adaptation (Brown & MacLeod, 1997; Webster & Mollon, 1995), and another which involves little of this adaptation, as when the eye moves over a scene patterned by light and shade (Zaidi, Spehar, & DeBonet, 1997) or when a tungsten lamp is briefly turned on in a room partly illuminated by sky light. In the constancy process based on adaptation effects, hue and saturation are preserved under the change in illuminant. For example, a paper that looks "unique yellow" under direct sunlight would continue to look unique yellow under the greenish light reflected or transmitted from under a tree. In the other kind of constancy process, hue and saturation change when the illuminant changes, but they are interpreted as resulting from constant surface colors (constant spectral reflectances) under varying illumination. Thus, the paper that looks unique yellow under direct sunlight would in fact look greenish-yellow under a tree but be clearly identifiable as yellow paper. The differences in the completeness of these two processes define the level of color constancy achieved in the two kinds of task (Bäuml, 1999; Logvinenko & Maloney, 2006).

The aim of the present work was to compare directly measurements of color constancy from different subjective and objective tasks. Measurements of each have been made before, but in different laboratories and with different displays, not all together in the same laboratory. The two subjective tasks were adapted from Arend & Reeves (1986): observers had to judge, here using a rating method, whether the appearance of the center square of a Mondrian presented under one illuminant matched the appearance of the center square of the Mondrian presented under another illuminant. The goodness of this appearance match was defined against an ideal in which either the center squares in the two patterns appeared to be made from exactly the same piece of material or the hue and saturation of the center squares appeared exactly the same. Notice that the judgment of the goodness of the match is an entirely subjective one. Since chromatic rather than achromatic attributes were of the essence here, any perceived luminance differences were disregarded (results with and without luminance variations are reasonably similar: Arend, Reeves, Schirillo, & Goldstein, 1991; Foster, Amano, & Nascimento, 2001).

A rating method was used for the subjective measurements rather than giving the observer direct control of the stimulus chromaticity, as in traditional asymmetric color matching, for two reasons. First, the same range and randomization of stimulus chromaticities could then be used as in the binary-response task. Second, as Logvinenko & Maloney (2006) have shown, using ratings rather than matches may counter a general problem with asymmetric color matching in that observers may find it impossible to achieve a satisfactory unconstrained color match (Brainard, Brunt, & Speigle, 1997).

The objective task was taken from Craven & Foster (1992), as described earlier; that is, observers had to judge, using a binary response, whether a Mondrian presented under one illuminant and then under another illuminant differed solely by an illuminant change or an illuminant change with an additional material change (affecting the center square).

Measurements using all three tasks were made at both Northeastern University, Boston, USA, and the University of Manchester, UK, but with different experimental designs. In Experiment 1, at Northeastern, observers were divided into six different groups for the three experimental tasks and two directions of illuminant change, but each observer in each group was presented with the full range of test stimuli. In Experiment 2, at Manchester, observers were not grouped, and each observer was given all three experimental tasks, two directions of illuminant change, and the full range of test stimuli.

GENERAL METHODS

Observers

Forty-one normal trichromats from Northeastern University served as observers in Experiment 1, and eight from the University of Manchester, in Experiment 2. Their color vision was tested variously with Ishihara plates, the Farnsworth-Munsell 100-Hue test, and Rayleigh and Moreland anomaloscopy. Individual differences in the 100-Hue test and anomaloscopy were uncorrelated with the extent of color constancy in 14 pilot observers, so these measures were not further analyzed. All participants had normal or corrected-to-normal spatial visual acuity. They were unaware of the purpose of the experiment and none had specialist knowledge of color vision or color science.

Apparatus

In both laboratories, stimuli were generated by RGB color-graphics systems with nominal 15-bit intensity resolution on each gun (VSG 2/5, Cambridge Research Systems Ltd, Rochester, Kent, UK), controlled by a laboratory computer and displayed on a 20-inch RGB raster-scan monitor (GDM-F520, Sony Corp., Tokyo, Japan). Screen resolution was 800 × 600 pixels. The screen refresh rate was approx. 100 Hz. The display system at Northeastern University was calibrated with a colorimeter (ColorCAL, Cambridge Research Systems Ltd, Rochester, Kent, UK) and at the University of Manchester with a telespectroradiometer (SpectraColorimeter, PR-650, Photo Research Inc., Chatsworth, California) that had previously been calibrated by the National Physical Laboratory. In that system, errors in the displayed CIE 1931 (x, y, Y) coordinates of a white test square were < 0.005 in (x, y) and < 3% in Y (< 5% at lower light levels).

Stimuli

Mondrians, gray-scale depictions of which are shown in Fig. 1, were viewed binocularly at 90 cm in a darkened room. The luminance of each square ranged from approx. 2 to 32 cd m^{-2}. Each pattern consisted of an array of 49 (7 × 7) simulated colored surfaces, of side 1º visual angle, therefore subtending 7° × 7° as a whole. Surfaces were sampled from a dataset of spectral reflectances (rather than compositions of spectral basis functions) comprising 1059 out of 1269 possible surfaces (Parkkinen, Hallikainen, & Jaaskelainen, 1989) in the *Munsell Book of Color* (Munsell Color Corporation, 1976). The random sampling producing each pattern was repeated, if necessary, to eliminate any accidental similarities between the illuminated center test surface and the immediately surrounding surfaces (Foster, Amano, & Nascimento, 2001; Maloney, 1999); more precisely, the difference between the test and any surround was larger than 55% of the color difference between the two illuminants in the approximately uniform CIE 1976 (u', v') chromaticity diagram. Fresh random samples were drawn on each trial.

[Insert Fig. 1 about here. Printer: please print at width specified in figure.]

The area of the screen surrounding the Mondrians was dark so that observers could not use any others surfaces in the field of view as a reference. The test surface, whose reflectance was to be manipulated, was the center square of the pattern. The first pattern was presented under a fixed spatially uniform daylight of correlated color temperature 4000 K or 16000 K (the first global illuminant). Except for the center square, the second pattern was identical to the first, but presented under the other uniform daylight (the second global illuminant), i.e. 16000 K or 4000 K, respectively. The CIE 1931 (x, y) coordinates of the global illuminants were (0.259, 0.267) and (0.381, 0.382), respectively, and, in the CIE 1976 chromaticity diagram, the (u', v') coordinates were (0.182, 0.423) and (0.223, 0.504), respectively.

To simulate a material change, the chromaticity of the center square of the second Mondrian was changed by replacing the global illuminant over the square by an independent, spatially uniform local illuminant, also drawn from the daylight spectrum (Judd, MacAdam, & Wyszecki, 1964), as detailed elsewhere (Foster, Amano, & Nascimento, 2001). Its chromaticity coordinates, shown in Fig. 2 by crosses in the (u', v') chromaticity diagram, were drawn from nine possible values along the daylight locus. [FOOTNOTE 2].

[Insert Fig. 2 about here. Printer: please print at width specified in figure.]

The advantage of this technique is that it ensures that the colorimetric change that occurs with a material change is of the same kind as the colorimetric change that occurs with a global illuminant change, so that the observer cannot respond merely on the basis of an “aberrant” color. It also it has certain technical advantages (Foster, Amano, & Nascimento, 2001) in that it quantifies changes in material chromaticity independent of the reflectance of the material. It does, however, effectively average chromaticity changes over different regions of color space, but the (u', v') chromaticity diagram used here is sufficiently uniform for the present purposes (Arend & Reeves, 1986), so a measure of constancy based on an average is likely to be reasonably representative.

The local illuminants were carefully chosen to range from the clearly too bluish to the clearly too reddish, so that, at the extremes, observers readily detected a material change. The nine local illuminants were chosen to include the two global illuminants (as indicated in Fig. 2). Thus on a random one-ninth of trials, the test surface changed in chromaticity but the local illuminant of the test signified no material change: an observer with perfect color-constancy would identify these trials as showing the same material (i.e. a pure illuminant change), and reject the rest. On a different one-ninth of the trials, the test surface did not change in chromaticity (i.e., it was given the local illuminant appropriate for no change in global illumination), and an observer with no color constancy would select these trials as showing a perfect hue-saturation match, and reject the rest. Although the chromaticity changes to the test surface were here constrained to the daylight locus, in unconstrained measurements (Foster, Amano, & Nascimento, 2006) the modes of observers’ responses lie on or close to this curve.

Tasks

In the rating tasks, observers were asked to "rate the quality of the simulations" of material changes or changes in hue and saturation by moving a mouse-controlled pointer on a vertical scale displayed on the monitor after the Mondrians were presented. The scale ranged from 100% at the top to 0% at the bottom. For observers rating material changes, 100% meant that the center square of the second pattern "looks as if it is made from exactly the same piece of paper (or material) as in the first pattern", and 0% meant that it "looks as if it is made from a quite different piece of paper (or material)". For observers rating hue-saturation changes, 100% meant that the "hue and saturation of the center square in the two patterns looks exactly the same", and 0% meant that they "look quite different". Observers were told to ignore any brightness changes. In the binary "same-material"-judgment task, observers pressed one of two keys to indicate whether the material of the center square of the pattern was the same in the first pattern as in the second, as far as possible ignoring any color change due to the change in illumination.

Procedure

Two Mondrians were presented sequentially at the same position, for 1 s each, with no temporal interval between them. Observers were allowed to move their eyes freely (Cornelissen & Brenner, 1995), and were given unlimited time to make each response. Data from the first block of trials were discarded as practice. Each local illuminant was presented 4 times in each block. Each participant in Experiment 1 gave 48 responses to each of 9 local illuminants, for a total of 432 judgments. Participants in Experiment 2 made 72 material-appearance ratings, 72 hue-saturation ratings, and 360 binary "same-material" judgments. The ordering of tasks was randomized across experimental sessions, but within each session, lasting no more than 1 hr, the task remained constant.

Performance measure (color-constancy index CCI)

An observer's performance in each task may be represented by the pattern of his or her mean ratings or proportion of binary responses r_i distributed over the nine local illuminants, $i = 1, 2, \ldots, 9$, identified with the corresponding (u_i', v_i') -coordinate on the daylight locus in the CIE 1976 chromaticity diagram shown in Fig. 2. Points on the daylight locus may be continuously parameterized by their distance s along the locus, so that if the nine local illuminants fall at $s_1, s_2, \ldots, s_9$, the effect of each task on performance may be summarized by the central tendency $\bar{s}$ of the distribution of the r_i. The position of the maximum of the raw distribution was itself too unstable to allow reliable inferences about $\bar{s}$. Instead, three robust measures were used: the "mode" of the distribution estimated by the position of the maximum over the continuous interval from s_1 to s_9 of a global quadratic regression of r_i on s_i; the mode of the distribution estimated by the position of the maximum over this interval of a locally weighted quadratic regression (i.e. quadratic loess: e.g. Cleveland, 1979; Fan & Gijbels, 1996) with bandwidth determined by cross-validation (Fan & Gijbels, 1996); and a weighted mean of the distribution defined by $\bar{s} = \sum_{i=1}^{9} r_i' s_i \Big/ \sum_{i=1}^{9} r_i'$, with $r_i' = r_i - \min_j r_j$. In fact, as made clear later, all three measures produced similar values of $\bar{s}$.

For ease of interpretation and for comparison with previous studies, these modes or means $\bar{s}$ were converted into standard color-constancy indices of the kind introduced by Arend *et al.* (1991). Thus, in the (u', v') chromaticity diagram, let a be the distance between $\bar{s}$ and the global illuminant on the second Mondrian, and let b be the distance between the two global illuminants [FOOTNOTE 3]; then the color-constancy index CCI is given by $1 - a/b$. Perfect constancy therefore corresponds to an index of unity, and the greater the error the lower the index.

There are other ways of analyzing the data that might be considered, especially methods relating to optimal observer behavior (e.g. Duda, Hart, & Stork, 2001). But these methods involve additional assumptions, and the degree to which observers are suboptimal is secondary to the more direct question of how their matches are affected by the task. The CCI has the advantage that it represents matching performance without reference to the magnitude of the particular response measure, whether based on material or hue-saturation ratings or on binary "same-material" judgments.

EXPERIMENT 1

The aim of the first experiment was to compare observers' performance in the three tasks using the same stimuli in the same laboratory. Observers were assigned a single task each, in order to prevent any possible confounding across tasks. Six observers made a rating of material appearance with a global illuminant change of 16000 K to 4000 K, and six with the opposite illuminant change; six observers made a rating of hue and saturation with a global illuminant change of 16000 K to 4000 K, and six with the opposite illuminant change; and nine observers made a binary judgment of origin with a global illuminant change of 16000 K to 4000 K, and eight with the opposite illuminant change;

Training

Observers were trained to distinguish changes in illumination, material, and hue and saturation on a computer display. First, they viewed successive images of a countryside scene, photographed at mid-day, in the late afternoon, and in the late afternoon with one object (a pale flowering bush) edited to a vivid orange. It was easy to discern that the illumination was the same in the second and third displays but differed in the first one; that the hue and saturation of the bush was different in each display; and that the "material" (flowers) of which the bush was made was different only in the third display. The third display was now re-edited by pasting in the flowering bush from the mid-day scene. The hue and saturation of the bush were the same in displays one and three, but different in display two, while the material was the same in displays one and two, but different in display three. This case was more difficult and observers typically needed several iterations of the displays to convince themselves that the same hue and saturation could imply a different material. Once these displays were understood, the observer was introduced to a similar sequence of Mondrians, with the center square playing the role of the bush. Training typically took 15–20 min. As all observers received training to the same level, differences in CCIs across tasks could not be attributed to differences in expertise.

Results and Comment

In the material-appearance rating task, a perfect observer would judge the appearance of the center square in relation to the rest of the Mondrian, and would give the highest quality rating when the local illuminant on the center square of the second Mondrian coincided with the second global illuminant. In the hue-saturation rating task, a perfect observer would be able to judge the appearance of the center square independently of the rest of the Mondrian, and would give the highest quality rating when the local illuminant on the center square of the second Mondrian coincided with the first global illuminant. In the binary judgment-of-origin task, a perfect observer would always respond “same material” when the local illuminant on the center square of the second Mondrian coincided with the second global illuminant, and respond “different material” when it did not.

Figure 3(a) shows the proportion of binary “same-material” responses (solid symbols) and the mean material-appearance rating (open symbols) plotted as a function of the u'-coordinate of the local illuminant labeling the change in reflectance of the center square of the

Mondrian (the increments in u' values increase because of the curvature of the daylight locus; see Fig. 2). The global illuminant change was from a correlated color temperature of 16000 K to one of 4000 K (indicated by the gray vertical lines). The rating scale on the right ordinate has been aligned with the proportion scale on the left ordinate for maximum overlap between the two sets of data. Since different observers were free to interpret the rating of stimulus quality along their own subjective ranges, the amplitude of the function has limited meaning, but the shape is important. The correlation between binary "same-material" judgments and material-appearance ratings was high, with Pearson's $\rho = 0.87$ ($p = .002$). Figure 3(b) shows the corresponding results for the opposite direction of illuminant change, from a correlated color temperature of 4000 K to one of 16000 K, for which Pearson's $\rho = 0.91$ ($p < .0001$). The peaks of the binary-judgment and material-rating response curves with the two directions of global illuminant change fall reasonably close to the u' values of the global illuminants of the second Mondrian: 4000 K in (a) and 16000 K in (b).

[Insert Fig. 3 about here. Printer: please print at width specified in figure.]

Figure 3(c) shows mean hue-saturation ratings (solid symbols) and, for comparison, the mean material-appearance rating (open symbols) from (a), plotted as a function of the u'-coordinate of the local illuminant. The rating scale on the right ordinate is identical to the one on the left. Figure 3(d) shows the corresponding results for the opposite direction, from a correlated color temperature of 4000 K to one of 16000 K. The peaks of the hue-saturation- and material-appearance-rating curves fall towards the opposite ends of the u'-range.

As might be anticipated, the interaction between reflectance change (u'-coordinate of local illuminant) and task was significant in ANOVAs with the nine local illuminants as a within-subjects factor and the task as a between-subjects factor, both for the data shown in

Fig. 3(c) ($F(8,80) = 23.7$, $p<.001$) and for the data shown in Fig.3(d) ($F(8,80) = 7.69$, $p<.001$). There was a close correspondence between observers' pooled responses for the two directions of global illuminant change, with Pearson's $\rho = 0.93$ ($p < .0001$).

The color-constancy indices for all three tasks, two directions of illuminant change, and three methods of estimation are listed in Table 1. The CCIs for material-appearance ratings and binary "same-material" judgments ranged from 0.63 to 0.85, and for hue-saturation ratings from 0.29 to 0.36. Indices averaged over the two directions of illuminant change were slightly higher for binary "same-material" judgments than for material-appearance ratings.

[Insert Table 1 about here.]

These values are compatible with values from previous studies cited in the Introduction, confirming them for the same displays and same observer training. Recall that the CCIs for binary "same-material" judgments or material-appearance ratings range from about 0.4 to about 0.8, depending on the experimental conditions and observers. Hue-saturation ratings with Mondrians also depend on the surround field, not least through adjacent and remote color-contrast effects (Barnes, Wei, & Shevell, 1999; Brenner, Ruiz, Herráiz, Cornelissen, & Smeets, 2003; Hurlbert & Wolf, 2004; Jameson & Hurvich, 1961; Krauskopf, Zaidi, & Mandler, 1986; Monnier & Shevell, 2003; Shevell & Wei, 2000; Tiplitz Blackwell & Buchsbaum, 1988a; Wachtler, Albright, & Sejnowski, 2001), and CCIs generally range from about 0.1 to about 0.4, depending on the experimental conditions and observers (Arend, Reeves, Schirillo, & Goldstein, 1991; Bäuml, 1999; Cornelissen & Brenner, 1995; Troost & de Weert, 1991). Higher indices for hue-saturation-brightness matches have been reported with rapidly alternating Mondrians (Barbur, de Cunha, Williams,

& Plant, 2004) and reflective stimuli in what was considered to be an equivalent color-appearance task (Brainard, Brunt, & Speigle, 1997). High indices have also been reported in a “red-green”–“blue-yellow” classification task with single presentations of patterns of colored surfaces (Smithson & Zaidi, 2004).

EXPERIMENT 2

Experiment 2 was conducted in the same way as Experiment 1, except that each observer participated in all three tasks. Notwithstanding the risk of confounding across tasks, the aim was to test if observers were task-consistent, that is, whether an individual with a high CCI in a material task would have a low CCI in the hue-saturation task (and vice versa), or were self-consistent, in that individuals with relatively higher CCIs in either task would have relatively higher CCIs in the other task.

Training

Observers first read the instructions for the experiment, and were then given a demonstration of a Mondrian undergoing the two kinds of change, as in the actual experiment, with the “material” of the center square remaining the same and with the hue and saturation of the center square remaining the same. Observers were required to explain the tasks back to experimenter to confirm that they had fully understood. Technical practice, making ratings and binary responses with the mouse, was performed over 10–20 trials. Training typically took 15–20 min.

Results

As in Experiment 1, there was a close correspondence between observers' pooled responses for the two directions of global illuminant change, with Pearson's $\rho = 0.92$ ($p < .0001$). Individual observers' mean material-appearance and hue-saturation ratings and proportion of binary "same-material" responses were therefore each averaged over the two directions. Color-constancy indices were calculated for individuals with quadratic loess, as described in General Methods.

Table 2 summarizes observers' CCIs in the three tasks. As expected, the difference between material-appearance and hue-saturation ratings was highly significant ($t = 4.1$, d.f. = 7, $p = 0.005$, 2-tailed test), but the difference between binary "same-material" responses and material-appearance ratings was not ($t = 0.19$, d.f. = 7, $p > 0.5$, 2-tailed test). The correlation between observers' CCIs in material-appearance and hue-saturation ratings was not significant (Pearson's $\rho = -0.15$, $p > 0.5$), nor between hue-saturation ratings and binary "same-material" responses ($\rho = 0.01$, $p > 0.5$), nor between binary "same-material" responses and material-appearance ratings ($\rho = -0.16$, $p > 0.5$).

[Insert Table 2 about here.]

The separation between observers' hue-saturation ratings and their binary "same-material" judgments was complete. Figure 4 shows CCIs from hue-saturation ratings plotted against CCIs from material-appearance ratings (open symbols) and CCIs from binary "same-material" judgments against CCIs from material-appearance ratings (solid symbols). Each point is numbered by observer, from 1 to 8. For some observers, the difference between their CCIs from binary "same-material" judgments and hue-saturation ratings was large (observer #1: 0.64) and for others, very small (observer #2: 0.18), but the difference was always positive. In fact, it was just possible to partition the data points with a straight line, from a logistic discriminant analysis.

[Insert Fig. 4 about here. Printer: please print at width specified in figure.]

DISCUSSION

Observers seem able to separate their judgments about color appearance from their judgments of the objective properties of reflecting surfaces under different illuminants. Experiment 1 showed that one group of observers, given the task of rating the extent to which the hue and saturation of the center square of two Mondrians was the same, made responses that were close to ideal, behaving almost as colorimeters. Another group of observers, given the task of rating the extent to which the material of the center square of the two Mondrians appeared the same, made a quite different pattern of responses which was also reasonably close to ideal, corresponding to the reflecting properties of the center square rather than to the spectrum of the reflected light. Finally, another group of observers, given the task of deciding whether the material of the Mondrians was the same in the first display as in the second, ignoring any color change due to the change in illumination, gave a pattern of responses that was closely parallel to those based on judgments of material appearance. Each pattern of observer responses was much the same whether the global illuminant change was from a daylight of correlated color temperature 16000 K to one of 4000 K or the opposite.

The methods used in this study to obtain CCIs do not distinguish between changes in sensitivity and changes in response criterion between tasks. In related experiments, however, Van Es, Vladusich, and Cornelissen (2007) asked observers to report whether a target color patch centered in an array of colored patches stayed the same across a simulated change in the illuminant (a "local" color judgment), or changed in a manner consistent with the illuminant change (a "global" color judgment). Judgments were binary ("Yes"/"No") rather than ratings. Target patches (a) stayed physically the same or (b) changed in a way entirely consistent with the illuminant change or (c) changed half-way. A color-constancy index was defined as the number of "Yes" trials in (b) plus half the "Yes" trials in (c) divided by the total number of "Yes" trials. This index, which unlike the one used here does not rely on the central tendency of the stimuli but rather on the response proportions, averaged between 70% and 80% in the global (constancy) task and 20% to 25% in the local (inconstancy) task, consistent with the estimates obtained here.

This ability to separate judgments of color appearance from judgments of physical origin is not simply an artifact of a group behavior, but, as Experiment 2 showed, it is demonstrated by individual observers. The same observer viewing the same stimuli could — to a greater or lesser extent — judge independently the hue and saturation of the stimulus and whether it represented a physically realizable reflecting surface under a change in illuminant. The latter judgment may depend on a mechanism by which we unconsciously "project" a subjective experience, such as color, back onto the physical world as an object property. It is not necessary that such a projection be completely identified with the subjective appearance, and for color, the data here suggest that it is not.

The ontological difference between the subjective experience and the protection of it back into the world is not especially intuitive for color, where careful measurements are needed to reveal it, but it is vivid for one's feelings of warmth. As one moves towards a fire, the sensation of warmth on the skin increases, but the "heat" that one attributes to the fire does not; the projected quality (heat in the fire) is clearly distinguishable from the sensation. Interestingly, constancy is also illustrated here, as distance is discounted when the inferred heat is projected back into the fire and experienced as a property of the fire. [FOOTNOTE 4]

How, then, were observers able to achieve these two seemingly contradictory modes of performance: the one concerning physical origin, the other color appearance? One possibility is that their judgments of material properties were mediated by a cue based on the relations between perceived colors under changes in illuminant rather than on the perceived colors themselves (Foster & Nascimento, 1994). These relations could, in turn, be represented by the spatial ratios of cone-receptor excitations generated in response to light reflected from pairs of surfaces or groups of surfaces. Such ratios, which may also be calculated across post-receptoral combinations and spatial averages of cone signals, have the property of being almost exactly invariant under changes in illuminant, both with natural scenes (Nascimento, Ferreira, & Foster, 2002) and with Mondrians of colored papers (Foster & Nascimento, 1994), as used here. [FOOTNOTE 5] The approximate constancy of cone-excitation ratios under illuminant changes may explain performance in several color-vision tasks, including the spatially parallel detection of reflectance changes (Foster et al., 2001), judgments of transparency (Ripamonti & Westland, 2003; Westland & Ripamonti, 2000), and asymmetric color matching (Tiplitz Blackwell & Buchsbaum, 1988b). As shown experimentally, cone-excitation ratios, or ratios of post-receptoral signals, provide a compelling cue to observers trying to distinguish between illuminant and reflectance changes in scenes, even when they sometimes correspond to highly unlikely natural events (Nascimento & Foster, 1997). Consistent with their proposed role in these judgments, cone-excitation ratios are not informative about individual surface reflectances (Tiplitz Blackwell & Buchsbaum, 1988b), and need not be involved in generating surface-color percepts, *per se*. An interesting variation of this approach has been proposed for matching hue, chroma, and value of Munsell surfaces whereby cone-excitation ratios are calculated not with respect to the physical, but the perceived background color (Daugirdiene, Murray, Vaitkevicius, & Kulikowski, 2006).

Models of color constancy have traditionally concentrated on the problem of estimating the illuminant color, for knowing the color, it is, in principle, possible for an observer to eliminate its effects on the perception of the scene (von Helmholtz, 1867). But judgments of illuminant color depend on many sources of information (Maloney, 2002; Smithson, 2005). For example, the scene could be assumed to be colorimetrically unbiased, so that no particular color predominates (the "gray-world" assumption: Buchsbaum, 1980; Land, 1986), when the space-average color then coincides with the illuminant color. With natural scenes, both these and higher-order statistics might also be exploited (Foster, Amano, & Nascimento, 2006; Golz & MacLeod, 2002). Another common assumption is that the surface with the highest luminance is white (Land & McCann, 1971; McCann, McKee, & Taylor, 1976). When pitted against each other, information about the space-average seems to take priority over the highest luminance (Linnell & Foster, 2002). Other cues such as mutual illumination (Bloj, Kersten, & Hurlbert, 1999) and specularities with non-uniform surfaces (Yang & Maloney, 2001) may also be used to infer the illuminant color.

In the present experiments, however, it would have been possible to perform all three tasks without estimating the illuminant color (Foster, 2003). Hue-saturation judgments, of course, merely require the observer to judge the local colorimetric properties of the test square in the Mondrian, defined by the excitations within the affected cones. For material judgments, either as ratings of appearance or as judgments of the objective properties of reflecting surfaces, it suffices to make relational judgments between the test square and one or more squares of the rest of the Mondrian, as just described: if the color relations provided by spatial cone-excitation ratios between the squares are not preserved, then the observer infers a material change.

Even so, performance in all three tasks was not perfect. That judgments of color appearance phenomena depend partly on objective properties and that judgments of those properties depend partly on color appearance parallels results obtained in asymmetric lightness and brightness perception (Arend & Goldstein, 1987; Arend & Spehar, 1993; Gilchrist et al., 1999; Logvinenko & Maloney, 2006). But within the limits of this interdependence, it seems that observers can separate their phenomenal percepts from their mental projections of those percepts onto the physical world (Joost, Lee, & Zaidi, 2002; Judd, 1940). That is, even when the quality of a particular chromatic change alters perceived hue and saturation, observers can reliably infer the cause, namely the constancy of the underlying reflecting surface.

FOOTNOTE 1

With maximum adaptation to each illuminant, Arend (1993) was able to obtain constancy indices for unique-hue settings of 0.66 on average. But with brief adaptation, the indices fell to 0.34 on average.

FOOTNOTE 2

In principle, the stimulus provided by any individual test square can be decomposed into any pair of illumination-reflection functions whose products yield metamers. Thus, at the extreme, an observer might perceive the display as a large sheet of gray paper with 49 tightly focused square-shaped illuminants falling on it (Adelson & Pentland, 1996). But an observer who is color-constant will perceive the display as 49 different squares under a single illuminant. Such an observer, if he or she detects the anomalous local illuminant, will conclude that the material of which the center test is made has been changed, which is the most parsimonious interpretation (Craven & Foster, 1992).

FOOTNOTE 3

Calculating constancy indices from Euclidean distances between points in CIE 1976 (u', v') chromaticity diagram and from distances between points measured along the daylight locus produced values within <1% of each other.

FOOTNOTE 4

As pointed out by a reviewer, one can imagine that the retinal signal generates a variety of cues that can be used as aids in the brain's interpretation of the signal, some of which relate to surface color, and some to illuminant color. The observer carries out a task by selectively combining a subset of these cues, with different subsets selected based on the observers' understanding of the instructions given. A weighted linear combination of such cues (Maloney, 2002) would be a "projection", but not the ontological projection discussed here.

FOOTNOTE 5

The rank orderings of pairs of cone excitations are also preserved under an illuminant change (Dannemiller, 1993), which is necessary since the rank ordering of two numbers is a weaker property than their ratio. But, critically, the converse does not hold: constant cone rankings do not imply an illuminant change, whereas constant cone ratios do, almost exactly, and this is how observers interpret ratio information (Nascimento & Foster, 1997).

REFERENCES

Adelson, E. H., & Pentland, A. P. (1996). The perception of shading and reflectance. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian Inference* (pp. 409-423). Cambridge: Cambridge University Press.

Amano, K., & Foster, D. H. (2004). Colour constancy under simultaneous changes in surface position and illuminant. *Proceedings of the Royal Society of London Series B-Biological Sciences,* **271**, 2319-2326.

Amano, K., & Foster, D. H. (2005). Minimalist surface-colour matching. *Perception,* **34**, 1009-1013.

Arend, L., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A-Optics Image Science and Vision,* **3**, 1743-1751.

Arend, L. E. (1993). How much does illuminant color affect unattributed colors? *Journal of the Optical Society of America A-Optics Image Science and Vision,* **10**, 2134-2147.

Arend, L. E., & Goldstein, R. (1987). Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America a-Optics Image Science and Vision,* **4**, 2281-2285.

Arend, L. E., Jr, Reeves, A., Schirillo, J., & Goldstein, R. (1991). Simultaneous color constancy: papers with diverse Munsell values. *Journal of the Optical Society of America A-Optics Image Science and Vision,* **8**, 661-672.

Arend, L. E., & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 1. Illuminance variation. *Perception & Psychophysics,* **54**, 446-456.

Baraas, R. C., Foster, D. H., Amano, K., & Nascimento, S. M. C. (2004). Protanopic observers show nearly normal color constancy with natural reflectance spectra. *Visual Neuroscience,* **21**, 347-351.

Barbur, J. L., de Cunha, D., Williams, C. B., & Plant, G. (2004). Study of instantaneous color constancy mechanisms in human vision. *Journal of Electronic Imaging,* **13**, 15-28.

Barnes, C. S., Wei, J., & Shevell, S. K. (1999). Chromatic induction with remote chromatic contrast varied in magnitude, spatial frequency, and chromaticity. *Vision Research,* **39**, 3561-3574.

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

Bloj, M. G., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature,* **402**, 877-879.

Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image. 1. Asymmetric matches. *Journal of the Optical Society of America A-Optics Image Science and Vision,* **14**, 2091-2110.

Brenner, E., Ruiz, J. S., Herráiz, E. M., Cornelissen, F. W., & Smeets, J. B. J. (2003). Chromatic induction and the layout of colours within a complex scene. *Vision Research,* **43**, 1413-1421.

Brown, R. O., & MacLeod, D. I. A. (1997). Color appearance depends on the variance of surround colors. *Current Biology,* **7**, 844-849.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute,* **310**, 1-26.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association,* **74**, 829-836.

Cornelissen, F. W., & Brenner, E. (1995). Simultaneous colour constancy revisited: an analysis of viewing strategies. *Vision Research,* **35**, 2431-2448.

Craven, B. J., & Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research,* **32**, 1359-1366.

Dannemiller, J. L. (1993). Rank orderings of photoreceptor photon catches from natural objects are nearly illuminant-invariant. *Vision Research,* **33**, 131-140.

Daugirdiene, A., Murray, I. J., Vaitkevicius, H., & Kulikowski, J. (2006). Cone contrast computations: physical versus perceived background and colour constancy. *Spatial Vision,* **19**, 173-192.

de Almeida, V. M. N., Fiadeiro, P. T., & Nascimento, S. M. C. (2004). Color constancy by asymmetric color matching with real objects in three-dimensional scenes. *Visual Neuroscience,* **21**, 341-345.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). New York: Wiley.

Fan, J., & Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall.

Foster, D. H. (2003). Does colour constancy exist? *Trends in Cognitive Sciences,* **7**, 439-443.

Foster, D. H., Amano, K., & Nascimento, S. M. C. (2001). Colour constancy from temporal cues: better matches with less variability under fast illuminant changes. *Vision Research,* **41**, 285-293.

Foster, D. H., Amano, K., & Nascimento, S. M. C. (2006). Color constancy in natural scenes explained by global image statistics. *Visual Neuroscience,* **23**, 341-349.

Foster, D. H., & Nascimento, S. M. C. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings of the Royal Society of London Series B-Biological Sciences,* **257**, 115-121.

Foster, D. H., Nascimento, S. M. C., Amano, K., Arend, L., Linnell, K. J., Nieves, J. L., et al. (2001). Parallel detection of violations of color constancy. *Proceedings of the National Academy of Sciences of the United States of America,* **98**, 8151-8156.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X. J., et al. (1999). An anchoring theory of lightness perception. *Psychological Review,* **106**, 795-834.

Golz, J., & MacLeod, D. I. A. (2002). Influence of scene statistics on colour constancy. *Nature,* **415**, 637-640.

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience,* **9**, 1367-1368.

Hurlbert, A., & Wolf, K. (2004). Color contrast: a contributory mechanism to color constancy. *Progress in Brain Research,* **144**, 147-160.

Jameson, D., & Hurvich, L. M. (1961). Opponent chromatic induction: experimental evaluation and theoretical account. *Journal of the Optical Society of America,* **51**, 46-53.

Joost, U., Lee, B. B., & Zaidi, Q. (2002). Lichtenberg's letter to Goethe on "Färbige Schatten" - Commentary. *Color Research and Application,* **27**, 300-303.

Judd, D. B. (1940). Hue saturation and lightness of surface colors with chromatic illumination. *Journal of the Optical Society of America,* **30**, 2-32.

Judd, D. B., MacAdam, D. L., & Wyszecki, G. (1964). Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America,* **54**, 1031-1040.

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

Krauskopf, J., Zaidi, Q., & Mandler, M. B. (1986). Mechanisms of simultaneous color induction. *Journal of the Optical Society of America A-Optics Image Science and Vision,* **3**, 1752-1757.

Land, E. H. (1986). Recent advances in Retinex Theory. *Vision Research,* **26**, 7-21.

Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America,* **61**, 1-11.

Linnell, K. J., & Foster, D. H. (2002). Scene articulation: dependence of illuminant estimates on number of surfaces. *Perception,* **31**, 151-159.

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.

Lucassen, M. P., & Walraven, J. (1996). Color constancy under natural and artificial illumination. *Vision Research,* **36**, 2699-2711.

Maloney, L. T. (1999). Physics-based approaches to modeling surface color perception. In K. R. Gegenfurtner & L. T. Sharpe (Eds.), *Color Vision: From Genes to Perception* (pp. 387-416). Cambridge: Cambridge University Press.

Maloney, L. T. (2002). Illuminant estimation as cue combination. *Journal of Vision,* **2**, 493-504.

McCann, J. J., McKee, S. P., & Taylor, T. H. (1976). Quantitative studies in Retinex theory. A comparison between theoretical predictions and observer responses to the "Color Mondrian" experiments. *Vision Research,* **16**, 445-458.

Monnier, P., & Shevell, S. K. (2003). Large shifts in color appearance from patterned chromatic backgrounds. *Nature Neuroscience,* **6**, 801-802.

Munsell Color Corporation. (1976). *Munsell Book of Color - Matte Finish Collection.* Baltimore, MD: Munsell Color Corporation.

Nascimento, S. M. C., Ferreira, F. P., & Foster, D. H. (2002). Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America A-Optics Image Science and Vision,* **19**, 1484-1490.

Nascimento, S. M. C., & Foster, D. H. (1997). Detecting natural changes of cone-excitation ratios in simple and complex coloured images. *Proceedings of the Royal Society of London Series B-Biological Sciences,* **264**, 1395-1402.

Parkkinen, J. P. S., Hallikainen, J., & Jaaskelainen, T. (1989). Characteristic spectra of Munsell colors. *Journal of the Optical Society of America A-Optics Image Science and Vision,* **6**, 318-322.

Ripamonti, C., & Westland, S. (2003). Prediction of transparency perception based on cone-excitation ratios. *Journal of the Optical Society of America A-Optics Image Science and Vision,* **20**, 1673-1680.

Shevell, S. K., & Wei, J. (2000). A central mechanism of chromatic contrast. *Vision Research,* **40**, 3173-3180.

Smithson, H., & Zaidi, Q. (2004). Colour constancy in context: Roles for local adaptation and levels of reference. *Journal of Vision,* **4**, 693-710.

Smithson, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences,* **360**, 1329-1346.

Tiplitz Blackwell, K., & Buchsbaum, G. (1988a). The effect of spatial and chromatic parameters on chromatic induction. *Color research and application,* **13**, 166-173.

Tiplitz Blackwell, K., & Buchsbaum, G. (1988b). Quantitative studies of color constancy. *Journal of the Optical Society of America A-Optics Image Science and Vision,* **5**, 1772-1780.

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

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.

von Helmholtz, H. (1867). *Handbuch der Physiologischen Optik, Vol. II, 1st Edition, published by Leopold Voss, Leipzig. Translation 3rd Edition (Helmholtz's Treatise on Physiological Optics, 1909) edited by J.P.C. Southall, Optical Society of America, Washington, D. C., 1924, pp. 286-287. Republished by Dover Publications, New York, 1962.*

von Kries, J. (1905). Die Gesichtsempfindungen. In W. Nagel (Ed.), *Handbuch der Physiologie des Menschen* (Vol. 3 of Physiologie der Sinne, pp. 109-282). Braunschweig: Vieweg und Sohn.

Wachtler, T., Albright, T. D., & Sejnowski, T. J. (2001). Nonlocal interactions in color perception: nonlinear processing of chromatic signals from remote inducers. *Vision Research,* **41**, 1535-1546.

Webster, M. A., & Mollon, J. D. (1995). Color constancy influenced by contrast adaptation. *Nature,* **373**, 694-698.

Westland, S., & Ripamonti, C. (2000). Invariant cone-excitation ratios may predict transparency. *Journal of the Optical Society of America A-Optics Image Science and Vision,* **17**, 255-264.

Whittle, P. (1996). Perfect von Kries contrast colours. *Perception,* **25 (supplement)**, 16.

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

Zaidi, Q., Spehar, B., & DeBonet, J. (1997). Color constancy in variegated scenes: role of low-level mechanisms in discounting illumination changes. *Journal of the Optical Society of America A-Optics Image Science and Vision,* **14**, 2608-2621.

TABLES

Table 1

Color-constancy indices for three observer tasks, two directions of global illuminant change, and three methods of estimation

		Global illuminant		
Method of estimate	**Observer task**	16000 K to 4000 K	4000 K to 16000 K	**Mean**
quadratic mode	binary “same-material” judgment	0.85	0.77	0.81
	material-appearance rating	0.83	0.64	0.74
	hue-saturation rating	0.29	0.32	0.31
loess mode	binary “same-material” judgment	0.82	0.73	0.78
	material-appearance rating	0.86	0.65	0.76
	hue-saturation rating	0.29	0.32	0.31
weighted mean	binary “same-material” judgment	0.75	0.70	0.73
	material-appearance rating	0.75	0.63	0.69
	hue-saturation rating	0.36	0.35	0.35

Note—Values of the color-constancy index (CCI) were estimated from the mode or mean of the distribution of responses in Fig. 3 (see General Methods). Global illuminants were characterized by their correlated color temperatures. An ideal observer would have a CCI of 1 for a binary “same-material” judgment and a material-appearance rating and a CCI of 0 for a hue-saturation rating. The last column shows the CCI averaged over the two directions of global illuminant change. Data based on 41 observers, 6–9 for each task and each direction of global illuminant change (Experiment 1).

Table 2

Means and standard errors of color-constancy indices for three observer tasks.

Observer task	**Mean**	**SE**
binary “same-material” judgment	0.77	0.04
material-appearance rating	0.75	0.07
hue-saturation rating	0.35	0.05

Note—Data based on 8 observers, who performed all three tasks with both directions of global illuminant change, from a correlated color temperature of 16000 K to one of 4000 K and the opposite (Experiment 2).

FIGURE CAPTIONS

Figure 1. An example of a pair of simulated Mondrians, depicted here in gray-scale. They consist of the same colored Munsell papers, but the pattern on the left was illuminated by a bluish light of correlated color temperature 16000 K and that on the right by a more yellow light of 4000 K. Although the gray-scale representation makes the patterns appear closely similar, it can be seen that some squares which are not discriminable on the left are discriminable on the right, and vice versa. Examples in color are given in Foster (2003).

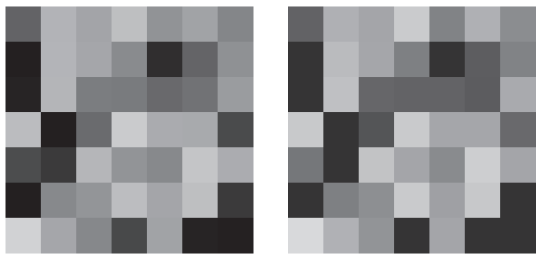
Figure 2. Chromaticity coordinates of the local daylight illuminant used to simulate a change in spectral reflectance of the center square of the second of two successively presented Mondrians under global illuminants of correlated color temperature 16000 K and 4000 K (arrowed). Points are plotted in the CIE 1976 (u', v') chromaticity diagram. The smooth curve is the daylight locus (Judd, MacAdam, & Wyszecki, 1964).

Figure 3. Observers' responses as a function of change in spectral reflectance of the center square of a Mondrian for different tasks and directions of global illuminant change. The plots in (a) show the proportion of binary "same-material" responses (solid symbols) and the mean material-appearance rating (open symbols) as a function of the CIE 1976 u'-coordinate of the local illuminant on the test surface. The global illuminant change was from a correlated color temperature of 16000 K to one of 4000 K (indicated by the gray vertical lines). The rating scale on the right ordinate has been aligned with the proportion scale on the left for maximum overlap between the two sets of data. The increments in u' values increase because of the curvature of the daylight locus (Fig. 2). The plots in (b) show the corresponding results for the

opposite global illuminant change, from a correlated color temperature of 4000 K to one of 16000 K. The plots in (c) show the mean hue-saturation rating (solid symbols) and, for comparison, the mean material-appearance rating (open symbols) from (a) as a function of the *u'*-coordinate of the local illuminant. The plots in (d) show the corresponding results for the opposite global-illuminant change. Data based on 41 observers, 6–9 for each task and direction of global illuminant change (Experiment 1).

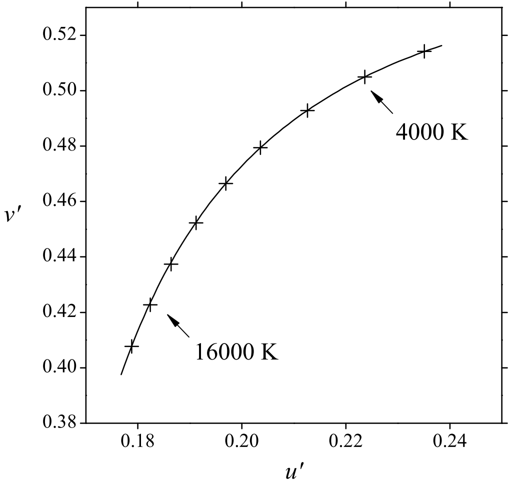
Figure 4. Color-constancy indices (CCIs) from hue-saturation ratings plotted against CCIs from material-appearance ratings (open symbols) and CCIs from binary “same-material” judgments against CCIs from material-appearance ratings (solid symbols). Overlapping points have been shifted away from each other by 0.01.Each point is numbered by observer, from 1 to 8 (Experiment 2). The dotted line is from a logistic discriminant analysis.

Reeves et al. Fig. 1
Printer: Do not edit or change scale; otherwise contact authors. Total width should be 70 mm



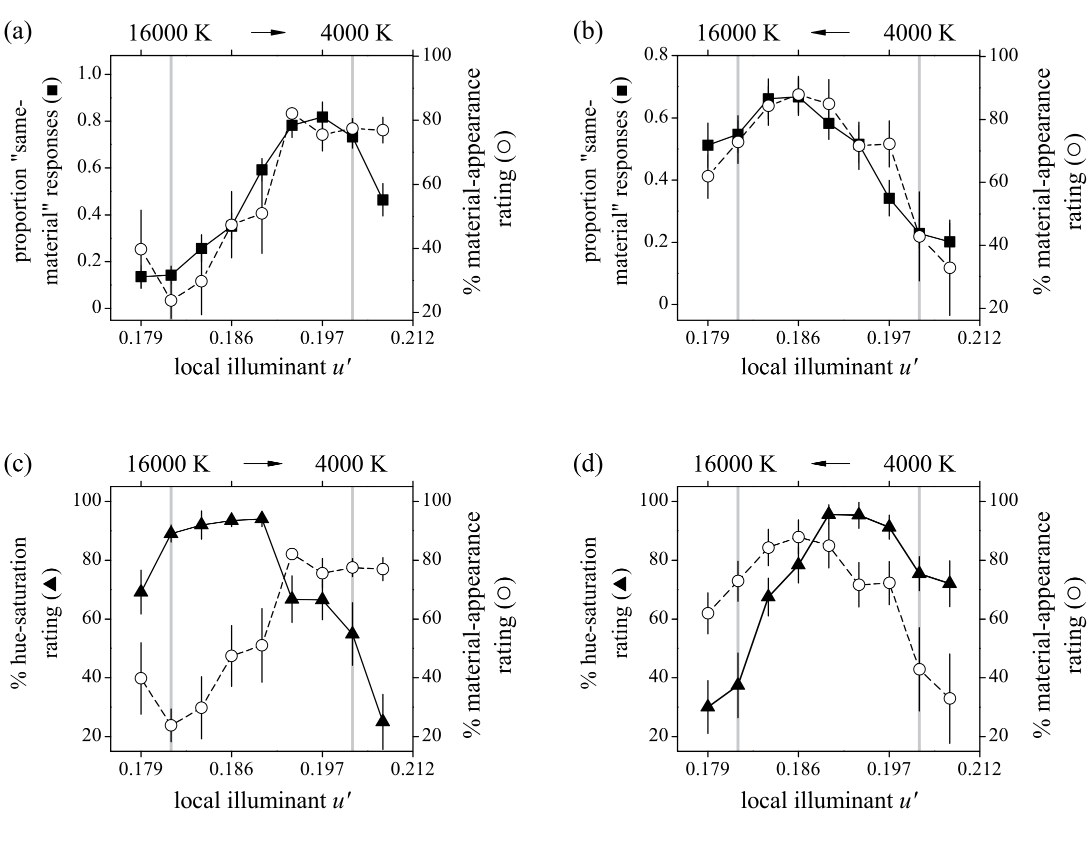
Reeves et al. Fig. 2

Printer: Do not edit or change scale; otherwise contact authors. Total width should be 67 mm



Reeves et al. Fig. 3

Printer: Do not edit or change scale; otherwise contact authors. Total width should be 142 mm



Reeves et al. Fig. 4

Printer: Do not edit or change scale; otherwise contact authors. Total width should be 71 mm