Color constancy: Phenomenal or projective?

ADAM J. REEVES
Northeastern University, Boston, Massachusetts

AND

KINJIRO AMANO AND DAVID H. FOSTER
University of Manchester, Manchester, England

Naive 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 with a correlated color temperature of either 16,000 or 4,000 K and then under the other, to test for color constancy. The observers compared the central square of the pattern across illuminants, either rating it for sameness of material appearance or 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 from a change 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.

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 of 1 and perfect inconstancy the value of 0 (Arend, Reeves, Schirillo, & Goldstein, 1991), reported levels of constancy in simultaneous asymmetric color matching of Mondrians have ranged from about .4 to about .8 (Amano & Foster, 2004; Amano, Foster, & Nascimento, 2005; Arend et al., 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 2-D and 3-D 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 in Foster, 2003.)

The nature of the task given to observers is important. In the measurements by Arend and Reeves (1986) and Arend et al. (1991), later confirmed by others (Bäuml, 1999;

A. J. Reeves, reeves@neu.edu

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 0 to about .3.[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 and 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 are able to perform the task rapidly, reliably, and with little or no training (Foster, Nascimento, et al., 2001). Levels of constancy attained with this operational method applied to Mondrians have been reported as about .77 (e.g., Baraas, Foster, Amano, & Nascimento, 2004), within the range obtained with 2-D 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 for the finding that levels of constancy differ according to task was anticipated in Arend and Reeves's (1986) article, in which they considered the role played by two kinds of constancy process: one that depends on the eye's 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 that 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 the changes 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 would 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 tasks (Bäuml, 1999; Logvinenko & Maloney, 2006).

The aim of the present work was to compare measurements of color constancy from different subjective and objective tasks directly. 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 and 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 to be exactly the same. Note that the judgment of the goodness of the match was 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 et al., 1991; Foster, Amano, & Nascimento, 2001).

Rather than giving the observers direct control of the stimulus chromaticity, as in traditional asymmetric color matching, a rating method was used for the subjective measurements for two reasons. First, the range and randomization of stimulus chromaticities could then be the same as those used as in the binary response task. Second, as Logvinenko and 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 et al., 1997).

The objective task was taken from Craven and Foster (1992), as described above; that is, the 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 by an illuminant change with an additional material change (affecting the center square).

Measurements using all three tasks were made at both Northeastern University and the University of Manchester, but with different experimental designs. In Experiment 1, at Northeastern, the 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, the 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 METHOD

Observers

Forty-one normal trichromats from Northeastern University served as observers in Experiment 1, and 8 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 the 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 graphic systems with nominal 15-bit intensity resolution on each gun (VSG 2/5; Cambridge Research Systems, Rochester, U.K.), controlled by a laboratory computer and displayed on a 20-in. RGB raster-scan monitor (GDM-F520; Sony, Tokyo). Screen resolution was 800×600 pixels. The screen refresh rate was approximately 100 Hz. The display system at Northeastern University was calibrated with a colorimeter (ColorCAL; Cambridge Research Systems, Rochester, U.K.) and at the University of Manchester with a telespectroradiometer (SpectraColorimeter, PR-650; Photo Research, Chatsworth, CA) 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, grayscale depictions of which are shown in Figure 1, were viewed binocularly at 90 cm in a darkened room. The luminance of each square ranged from approximately 2 to 32 cd/m². Each pattern consisted of an array of 49 (7×7) simulated colored surfaces, 1º of visual angle per side; therefore the array subtended 7º × 7º as a whole. The surfaces were sampled from a data set of spectral reflectances (rather than compositions of spectral basis functions) comprising 1,059 out of 1,269 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-surface color and any surround color 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.

The area of the screen surrounding the Mondrians was dark, so that the observers could not use any other surface 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 with a correlated color temperature of 4,000 or 16,000 K (the first global illuminant). Except for the center square, the second pattern was identical to the first but was presented under the other uniform daylight (the second global illuminant)—that is, 16,000 or 4,000 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
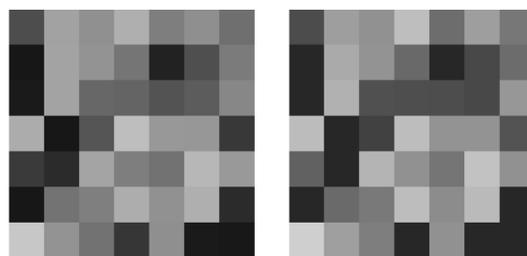
Figure 1. An example of a pair of simulated Mondrians, depicted here in grayscale. They consisted of the same colored Munsell papers, but the pattern on the left was illuminated by a bluish light with a correlated color temperature of 16,000 K, and that on the right by a more yellow light of 4,000 K. Although the grayscale representation makes the patterns appear closely similar, it can be seen that some squares that are not discriminable on the left are discriminable on the right and vice versa. Examples in color are given in Foster (2003).

illuminant over the square with 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 Figure 2 by crosses in the (u', v') chromaticity diagram, were drawn from nine possible values along the daylight locus.[2]

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 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.
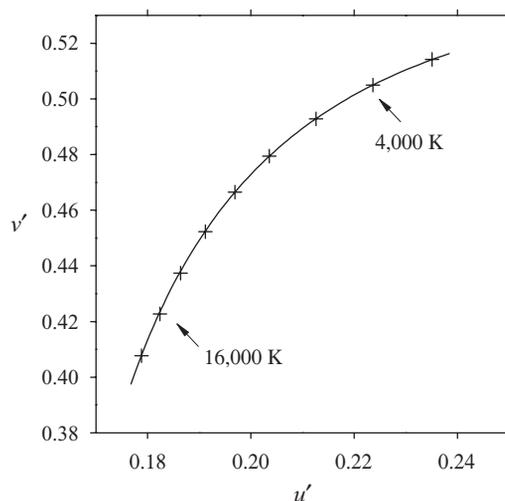
Figure 2. Chromaticity coordinates of the local daylight illuminant used to simulate a change in spectral reflectance of the center square of the second of two successively presented Mondrians under global illuminants with correlated color temperatures of 16,000 and 4,000 K (arrows). Points are plotted in the CIE 1976 (u', v') chromaticity diagram. The smooth curve is the daylight locus (Judd, MacAdam, & Wyszecki, 1964).

The local illuminants were carefully chosen to range from the clearly too bluish to the clearly too reddish, so that, at the extremes, the observers readily detected a material change. The nine local illuminants were chosen so as to include the two global illuminants (as indicated in Figure 2). Thus, on a random one ninth of the 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 would 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 would reject the rest. Although the chromaticity changes to the test surface were here constrained to the daylight locus, in unconstrained measurements (Foster et al., 2006), the modes of observers' responses lay on or close to this curve.

Tasks

In the rating tasks, the 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 had been presented. The scale ranged from 100% at the top to 0% at the bottom. For the 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 the 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." The observers were told to ignore any brightness changes. In the binary same-material judgment task, the 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, ignoring, as far as possible, any color change due to the change in illumination.

Procedure

Two Mondrians were presented sequentially at the same position for 1 sec each, with no temporal interval between them. The 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 four times in each block. Each participant in Experiment 1 gave 48 responses to each of nine local illuminants, for a total of 432 judgments. The 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 h, the task remained constant.

Performance Measure (Color Constancy Index)

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') coordinates on the daylight locus in the CIE 1976 chromaticity diagram shown in Figure 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; see, 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

$$\bar{s} = \frac{\sum_{i=1}^{9} r_i' s_i}{\sum_{i=1}^{9} r_i'},$$

with $r_i' = r_i - \min(r_1, r_2, \ldots, r_9)$. In fact, as will be made clear below, 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 (CCIs) 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;[3] then the CCI is given by $1 - a/b$. Perfect constancy therefore corresponds to an index of unity, and the less the constancy 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 the observers' performance on the three tasks, using the same stimuli in the same laboratory. Each of the observers was assigned a single task, in order to prevent any possible confounding across tasks. Six observers made a rating of material appearance with a global illuminant change of 16,000 to 4,000 K, and 6 with the opposite illuminant change; 6 observers made a rating of hue and saturation with a global illuminant change of 16,000 to 4,000 K, and 6 with the opposite illuminant change; and 9 observers made a binary judgment of origin with a global illuminant change of 16,000 to 4,000 K, and 8 with the opposite illuminant change.

Training

The 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 midday, 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 reedited by pasting in the flowering bush from the midday scene. The hue and saturation of the bush were the same in the first and third displays, but different in the second display, whereas the material was the same in the first and second displays, but different in the third display. 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 observers were introduced to a similar sequence of Mondrians, with the center square playing the role of the bush. Training typically took 15–20 min. Since all the 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 would respond "different material" when it did not.

Figure 3A 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 Figure 2). The global illuminant change was from a correlated color temperature of 16,000 K to one of 4,000 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 $\rho = .87$ ($p = .002$). Figure 3B shows the corresponding results for the opposite direction of illuminant change, from a correlated color temperature of 4,000 K to one of 16,000 K, for which Pearson's $\rho = .91$ ($p < .0001$). The peaks of the binary judgment and material rating response curves with the two directions of global illuminant
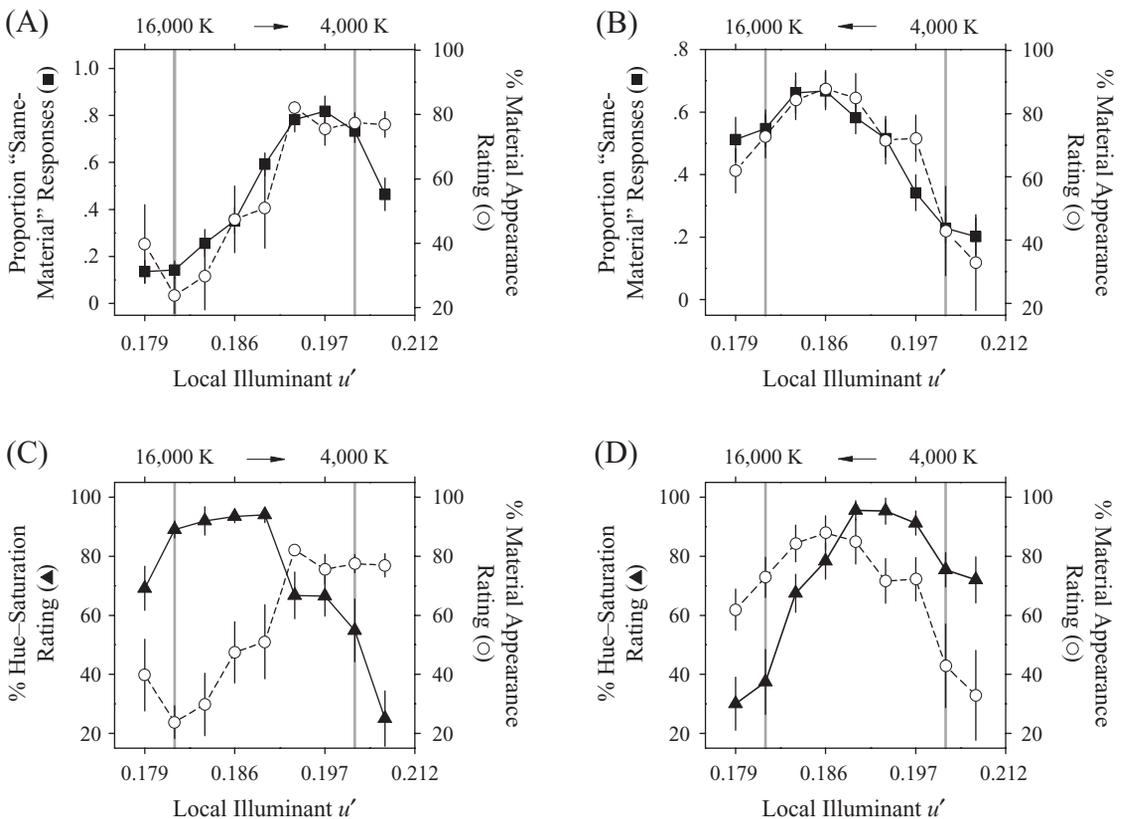
Figure 3. Observers' responses as a function of change in spectral reflectance of the center square of a Mondrian for different tasks and directions of global illuminant change. (A) Proportion of binary same-material responses (solid symbols) and the mean material appearance rating (open symbols) as 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 16,000 K to one of 4,000 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 (see Figure 2). (B) Corresponding results for the opposite global illuminant change, from a correlated color temperature of 4,000 K to one of 16,000 K. (C) Mean hue–saturation rating (solid symbols) and, for comparison, the mean material appearance rating (open symbols) from panel A as a function of the u'-coordinate of the local illuminant. (D) Corresponding results for the opposite global illuminant change. Data are based on 41 observers, 6–9 for each task and direction of global illuminant change (Experiment 1).

change fall reasonably close to the u' values of the global illuminants of the second Mondrian: 4,000 K in panel A and 16,000 K in panel B.

Figure 3C shows mean hue–saturation ratings (solid symbols) and, for comparison, the mean material appearance rating (open symbols) from panel 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 3D shows the corresponding results for the opposite direction, from a correlated color temperature of 4,000 K to one of 16,000 K. The peaks of the hue–saturation and material appearance rating curves fall toward the opposite ends of the u'-range.

As might be anticipated, the interaction between reflectance change (u'-coordinate of the 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 Figure 3C [$F(8,80) = 23.7, p < .001$] and for the data shown in Figure 3D [$F(8,80) = 7.69, p < .001$]. There was a close correspondence between observers' pooled responses for the two directions of global illuminant change, with Pearson's $\rho = .93$ ($p < .0001$).

The CCIs 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 .63 to .85, and for hue–saturation ratings from .29 to .36. Indices averaged over the two directions of illuminant change were slightly higher for binary same-material judgments than for material appearance ratings.

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 .4 to about .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 .1 to about .4, depending on the experimental conditions and observers (Arend et al., 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 et al., 1997). High indices have also been reported in a red–green and 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 whether the observers would be 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

The 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. The observers were required to explain the tasks back to the experimenter to confirm

Table 1
Color Constancy Indices for Three Observer Tasks, Two Directions of Global Illuminant Change, and Three Methods of Estimation

Method of Estimation	Observer Task	Global Illuminant		
		16,000 to 4,000 K	4,000 to 16,000 K	M
Quadratic mode	Binary same-material judgment	.85	.77	.81
	Material appearance rating	.83	.64	.74
	Hue–saturation rating	.29	.32	.31
Loess mode	Binary same-material judgment	.82	.73	.78
	Material appearance rating	.86	.65	.76
	Hue–saturation rating	.29	.32	.31
Weighted mean	Binary same-material judgment	.75	.70	.73
	Material appearance rating	.75	.63	.69
	Hue–saturation rating	.36	.35	.35

Note—Values of the color constancy index (CCI) were estimated from the mode or mean of the distribution of responses in Figure 3 (see the General Method section). 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 are based on 41 observers, 6–9 for each task and each direction of global illuminant change (Experiment 1).

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 $\rho = .92$ ($p < .0001$). Individual observers' mean material appearance and hue–saturation ratings and proportions of binary same-material responses were, therefore, each averaged over the two directions. CCIs were calculated for individuals with quadratic loess, as described in the General Method section.

Table 2 summarizes the observers' CCIs in the three tasks. As was expected, the difference between material appearance and hue–saturation ratings was highly significant [$t(7) = 4.1$, $p = .005$, two-tailed test], but the difference between binary same-material responses and material appearance ratings was not [$t(7) = 0.19$, $p > .5$, two-tailed test]. The correlation between the observers' CCIs in material appearance and hue–saturation ratings was not significant (Pearson's $\rho = -.15$, $p > .5$), nor were those between hue–saturation ratings and binary same-material responses (Pearson's $\rho = .01$, $p > .5$) and between binary same-material responses and material appearance ratings (Pearson's $\rho = -.16$, $p > .5$) significant. (These correlations over observers should not be confused with the correlations over local illuminants calculated in Experiment 1.)

The separation between the 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 the 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: .64) and for others, very small (Observer 2: .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.

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 to be the same, made a quite different pattern of responses that 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

Table 2
Means and Standard Errors of Color Constancy Indices for Three Observer Tasks

Observer Task	M	SE
Binary same-material judgment	.77	.04
Material appearance rating	.75	.07
Hue–saturation rating	.35	.05

Note—Data are based on 8 observers, who performed all three tasks with both directions of global illuminant change, from a correlated color temperature of 16,000 K to one of 4,000 K and the opposite (Experiment 2).

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 with a correlated color temperature of 16,000 K to one of 4,000 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 (1) stayed physically the same, (2) changed in a way entirely consistent with the illuminant change, or (3) changed halfway. A CCI was defined as the number of *yes* trials in (2) plus half the *yes* trials in (3) divided by the total number of *yes* trials. This
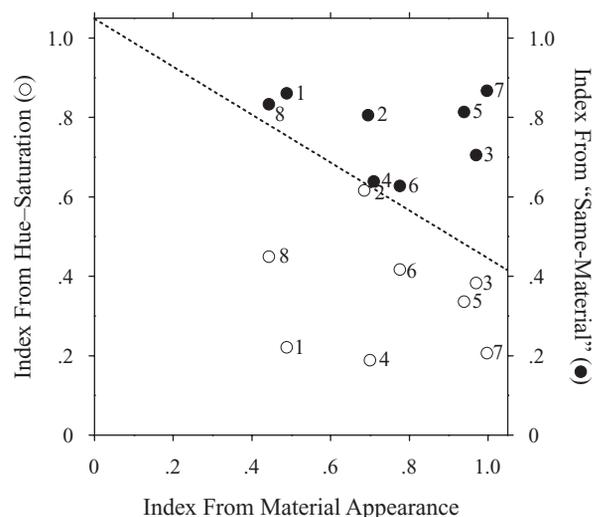
Figure 4. Color constancy indices (CCIs) from hue–saturation ratings plotted against CCIs from material appearance ratings (open symbols) and CCIs from binary same-material judgments plotted against CCIs from material appearance ratings (solid symbols). Overlapping points have been shifted away from each other by .01. Each point is numbered by observer, from 1 to 8 (Experiment 2). The dotted line is from a logistic discriminant analysis.

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 projection 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 toward 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, since distance is discounted when the inferred heat is projected back into the fire and experienced as a property of the fire.[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 postreceptoral 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), such as those used here.[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, Nascimento, et al., 2001), judgments of transparency (Ripamonti & Westland, 2003; Westland & Ripamonti, 2000), and asymmetric color matching (Tiplitz Blackwell & Buchsbaum, 1988b). As has been shown experimentally, cone excitation ratios, or ratios of postreceptoral 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 adjusted for 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 (Helmholtz, 1867/1924). 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 et al., 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 nonuniform 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, 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 parallel 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 et al., 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.

AUTHOR NOTE

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 B*, **271**, 2319-2326.

AMANO, K., FOSTER, D. H., & NASCIMENTO, S. M. C. (2005). Minimalist surface–colour matching. *Perception*, **34**, 1007-1011.

AREND, L. E., [JR.] (1993). How much does illuminant color affect unattributed colors? *Journal of the Optical Society of America A*, **10**, 2134-2147.

AREND, L. E., [JR.], & GOLDSTEIN, R. (1987). Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America A*, **4**, 2281-2285.

AREND, L. [E., JR.], & REEVES, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A*, **3**, 1743-1751.

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*, **8**, 661-672.

AREND, L. E., [JR.], & SPEHAR, B. (1993). Lightness, brightness, and brightness contrast: 1. Illuminance variation. *Perception & Psychophysics*, **54**, 446-456.

BAARAS, 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*, **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 B*, **257**, 115-121.

FOSTER, D. H., NASCIMENTO, S. M. C., AMANO, K., AREND, L., [E., JR.], LINNELL, K. J., NIEVES, J. L., ET AL. (2001). Parallel detection of violations of color constancy. *Proceedings of the National Academy of Sciences*, **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.

HELMHOLTZ, H. VON (1924). *Helmholtz's treatise on physiological optics* (J. P. C. Southall, Ed.). Washington, DC: Optical Society of America. (Original work published 1867)

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." *Color Research & 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*, **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: Author.

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*, **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 B*, **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*, **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*, **20**, 1673-1680.

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

SMITHSON, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society B*, **360**, 1329-1346.

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

TIPLITZ BLACKWELL, K., & BUCHSBAUM, G. (1988a). The effect of spatial and chromatic parameters on chromatic induction. *Color Research & Application*, **13**, 166-173.

TIPLITZ BLACKWELL, K., & BUCHSBAUM, G. (1988b). Quantitative studies of color constancy. *Journal of the Optical Society of America A*, **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 KRIES, J. (1905). Die Gesichtsempfindungen. In W. Nagel (Ed.), *Physiologie der Sinne: Vol. 3. Handbuch der Physiologie des Menschen* (pp. 109-282). Braunschweig: Vieweg & 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*, **17**, 255-264.

WHITTLE, P. (1996). Perfect von Kries contrast colours. *Perception*, **25**(Suppl.), 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*, **14**, 2608-2621.

NOTES

1. With maximum adaptation to each illuminant, Arend (1993) was able to obtain constancy indices for unique-hue settings of .66 on average. But with brief adaptation, the indices fell to .34 on average.

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).

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

4. As was 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 on the basis of 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.

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).