PERGAMON

Vision Research 39 (1999) 1531–1550

Vision Research

Simultaneous color constancy: how surface color perception varies with the illuminant

Karl-Heinz Bäuml *

Institut für Psychologie, Universität Regensburg, 93040 Regensburg, Germany

Received 10 July 1997; received in revised form 9 February 1998

Abstract

In two experiments simultaneous color constancy was measured using simulations of illuminated surfaces presented on a CRT monitor. Subjects saw two identical Mondrians side-by-side: one Mondrian rendered under a standard illuminant, the other rendered under one of several test illuminants. The matching field was adjusted under the test illuminant so that it (a) had the same hue, saturation, and brightness (appearance match) or (b) looked as if it were cut from the same piece of paper (surface match) as a test surface under the standard illuminant. Matches were set for three different surface collections. The surface matches showed a much higher level of constancy than the appearance matches. The adjustment in the surface matches was nearly complete in the L and M cone data, and deviations from perfect constancy were mainly due to failures in the adjustment of the S cone signals. Besides this difference in amount of adjustment, the appearance and surface matches showed two major similarities. First, both types of matches were well described by simple parametric models. In particular, a model based on the notion of von Kries adjustment provided a good, although not perfect, description of the data. Second, for both types of matches the illuminant adjustment was largely independent of the surface collection in the image. The two types of matches thus differed only quantitatively, there was no qualitative difference between them. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Color; Color constancy; Appearance; Surface color; von Kries principle

1. Introduction

The light that is reflected from an illuminated object depends partly on the object's surface and partly on the illumination in the scene. This well-known fact raises a problem for our visual system. Since the illumination can vary drastically within days, illuminant changes can have a considerable impact on the light that is reflected from an object and, thus, on the receptor absorptions that result from it. Our visual system must compensate against this effect of illumination, in order to assign constant colors to objects.

The need for an adjustment to changes in illumination arises under quite different situations for our visual system. Two types of these situations were the primary focus of color constancy research in recent years. In the one type of situation, often referred to as successive color constancy, a scene is fairly uniformly illuminated by a light source, and illuminant changes are gradual and from one time to another. Under natural conditions this type of situation, for instance, can occur when a scene is lit by direct sunlight, say morning light, and then the spectral composition of the daylight changes gradually over the day. In the other type of situation, multiple light sources are present within the image, and the illuminant variation is relatively abrupt and within the image. Under natural conditions this effect may be caused, for instance, by shadowing, where one part of the scene is lit by direct sunlight while another part is cast in shadow. This situation is often referred to as simultaneous color constancy (Arend & Reeves, 1986; Brainard, 1998).

By using the method of asymmetric color matching (Wyszecki & Stiles, 1982), research in successive color constancy has repeatedly examined how the color appearance of objects varies with illumination. In this method, a subject sets a matching field under a test illuminant to match the appearance of a test surface under a standard illuminant. Changes in the subject's settings as a function of changes in illumination then

* E-mail: karl-heinz.baeuml@psychologie.uni-regensburg.de.

PII: S0042-6989(98)00192-8

reflect the visual system's adjustment to this change in the visual context. Typically, a considerable amount of illuminant adjustment was found in the subjects' matches, at least when there was enough time for the visual system to adapt to the illuminant change (Brainard & Wandell, 1992; Bäuml, 1995).

Only a low amount of illuminant adjustment was found in subjects' color appearance matches in the situation of simultaneous color constancy, which is in agreement with the general observation that adjustments in color appearance to changes in illumination occur quite slowly (Hunt, 1950; Fairchild & Reniff, 1995). At the same time, however, it has been shown that there is a considerable amount of adjustment in subjects' surface color matches in this type of situation. Arend & Reeves (1986), for instance, presented subjects with two identical Mondrians side-by-side, which were each illuminated by a different light source. The subjects were asked to do two types of asymmetric color matching. In the first step, they were asked to adjust a matching field in, say, the right Mondrian so that it had the same hue, saturation, and brightness as a test field in the left Mondrian (appearance match). In the second step, they were asked to adjust the matching field in the right Mondrian so that it looked as if it were cut from the same piece of paper as the test field in the left Mondrian, that is, as if it were the same surface, only rendered under a different illuminant (surface match). The main result from their study was that when setting appearance matches the subjects' matches were much closer to physical matches than to color constant matches. However, when setting surface matches they were closer to color constant matches than to physical matches. This pattern of results demonstrates that the same surface can have quite a different color appearance under two different illuminants but, at the same time, may be perceived as roughly the same surface rendered under two different illuminants. A number of consecutive studies confirmed this finding (Arend, Reeves, Schirillo & Goldstein, 1991; Troost & deWeert, 1991; Cornelissen & Brenner, 1995).

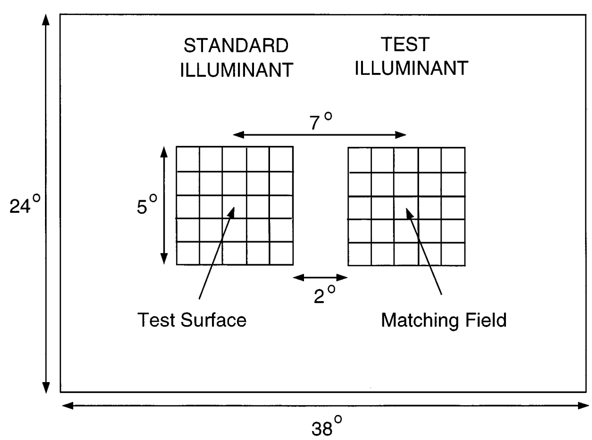

Fig. 1. The visual display. Subjects saw CRT simulations of two identical Mondrians, presented against a dark background field. The left-hand Mondrian was rendered under the standard illuminant, the right-hand Mondrian was rendered under one of four test illuminants. The Mondrians consisted of 25 rectangular fields of equal size. In the center field of the left-hand Mondrian the test surface was presented. The center field of the right-hand Mondrian (matching field) was adjusted by the subject to make it match in color (hue, saturation, and brightness or surface color) with the center field of the left-hand Mondrian, i.e. the test surface.

From experimental work investigating successive color constancy we know that, in situations with uniform and gradually changing illumination, the color appearance of objects is maintained across illuminant changes through an adjustment of the visual system that is largely in terms of the three cone signals, a principle often called von Kries adaptation (Kries von, 1905; Brainard & Wandell, 1992; Fairchild & Berns, 1993; Bäuml, 1995; Chichilnisky & Wandell, 1995; Brainard, Brunt & Speigle, 1997). Based on this von Kries principle, the cone signals that result from an illuminated object are scaled as a function of the illumination in the scene. For instance, if an illuminant mainly emits light in the long-wavelength part of the spectrum, then it is mainly the L cone signal which is scaled; if an illuminant mainly emits light in the short-wavelength part of the spectrum, it is mainly the S cone signal which is scaled. Due to this scaling, a large part of the illuminant effect that is originally present in the absorptions can be eliminated. In addition, there is evidence that, in this situation, the illuminant adjustment is to a considerable degree independent from the objects in the image. Both in CRT studies (Bäuml, 1994, 1997) and in more natural images (Brainard, 1998) it has been found that the same illuminant change induces largely the same illuminant adjustment under varying image surfaces.

While we know quite a lot about the illuminant adjustment of our visual system in successive color constancy, we do not know very much about its adjustment in simultaneous color constancy. Two open questions stand out. First, can we describe surface color matches by means of the same class of models as appearance matches? In particular, is there a principle similar to the von Kries principle which can account for the adjustment underlying surface color perception? Second, is there substantial variation in surface color matches as a function of the surfaces in the image, or is the illuminant adjustment largely independent of the image surfaces, as seems to be the case for appearance matches? In this paper, two series of experiments designed to address one of these questions each are described. Experiment 1 examines whether we can use the same simple models, which we use to describe appearance matches, to account also for surface matches; Experiment 2 then examines the influence of the surface collection in the image on this type of matches. The

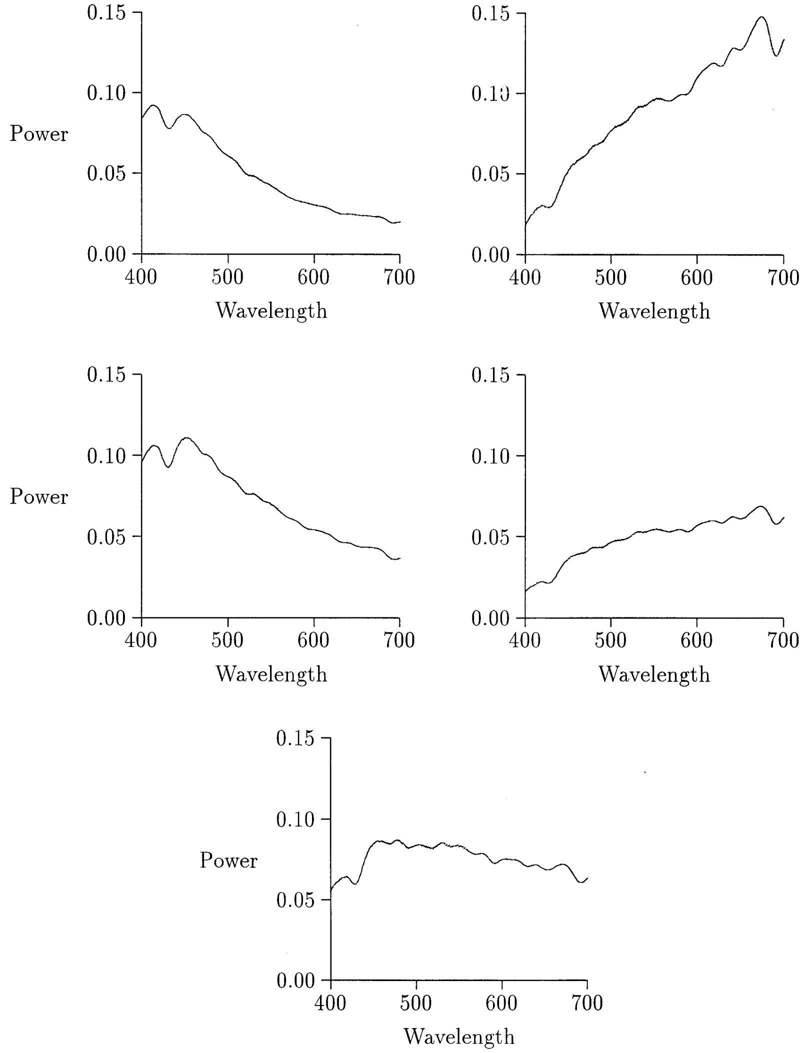

Fig. 2. Experimental illuminants. Each graph shows the spectral power distribution of one of the five illuminants used in the experiment. The bottom graph shows the standard illuminant which rendered the left-hand Mondrian. The other four graphs show the four test illuminants, which rendered the right-hand Mondrian. The two left-hand graphs show the two bluish test illuminants (T_1, T_2) and the two right-hand graphs the two yellowish test illuminants (T_3, T_4). All five illuminants are typical for natural daylight (see also Table A1 in the Appendix).

experimental results will provide us new information on how surface color perception varies with illumination. They will also cast light on the issue of whether appearance and surface matches differ quantitatively or qualitatively.

2. Experiment 1

2.1. Method

2.1.1. Visual display

Subjects saw CRT simulations of two identical Mondrians, in which one of the two Mondrians was rendered under the standard illuminant and the other was rendered under one of four test illuminants (Fig. 1). The Mondrians consisted of 25 rectangular fields of equal size. In the center field of the left-hand Mondrian, the test surface was presented. The center field of the right-hand Mondrian (matching field) was adjusted by the subject to make it match in color with the center field of the left-hand Mondrian, i.e. the test surface. The two Mondrians were presented against a dark uniform background field (< 0.01 cd/m^2), which subtended 24 vertical × 38 horizontal degrees of visual angle. The two Mondrians subtended 5° of visual angle each, both horizontally and vertically, and were separated from each other by 2° of visual angle. So, each of the single patches of the two Mondrians subtended 1° of visual angle. Subjects saw the screen without head restraints from a distance of about 0.5 m in an otherwise dark room.

The patterns were displayed on a computer-controlled color monitor (BARCO Calibrator CCID 7651)

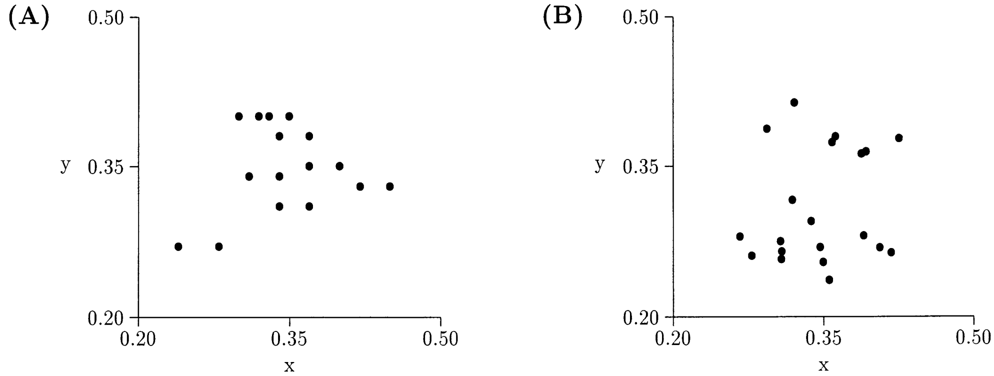

Fig. 3. Experimental surfaces. (A) The graph shows the CIE x, y chromaticity coordinates of the 16 surfaces that were used as test surfaces in this experiment; they were presented in the center field of the left-hand Mondrian. (B) The graph shows the CIE x, y chromaticity coordinates of the 24 surfaces which surrounded the test surface and matching field within the visual display. In both graphs, the coordinates of the surfaces that result when they are rendered under the standard illuminant are shown (see also Table A2 in the Appendix).

using a refresh rate of 71 Hz in non-interlaced video mode. The three channels of the monitor were controlled by an 8-bit digital-to-analog converter. The signals of the color channels could be varied in 256 steps from zero to maximal intensity for each pixel. The monitor's input signal was controlled by software, which corrected non-linearities in the tube's response function. The luminance of each color channel was measured with a high precision photometer (Fa. Lichtmesstechnik, Model L 1003).The CIE x, y coordinates of the phosphors were provided by the manufacturer. To compensate for local variations in the gamma curves, the above measurements were done separately for the screen locations where the two Mondrians were presented. Gamma curves measured at the center of each Mondrian were used to control the stimuli. The programming was done by using PXL subroutines (Irtel, 1997).

2.1.2. Experimental illuminants and surfaces

Fig. 2 shows the spectral power distributions of the illuminants that were used in this experiment. The bottom graph shows the standard illuminant which rendered the left-hand Mondrian. The other four illuminants represent the four test illuminants, which rendered the right-hand Mondrian. The two left-hand graphs show the two bluish test illuminants and the two right-hand graphs the two yellowish test illuminants. All five illuminants were drawn from the CIE daylight locus (Wyszecki & Stiles, 1982) and are typical for natural daylight. They span the whole range of chromaticities that is typically observed in natural daylight and vary also in their luminance. They were constructed from the three-dimensional linear model of Judd, MacAdam & Wyszecki (1964) (see Table A1 in the Appendix).

All experimental surfaces were drawn from the large set of simulated papers used by Brainard & Wandell (1992). These surfaces are approximations of Munsell papers. It is well known that the spectra of Munsell chips can be well approximated by appropriate linear combination of about six basis surface reflectance functions. Moreover, taking human photoreceptors into account the first three or four basis functions of the linear model are already sufficient to fit the chips' reflectance closely (Maloney, 1986). The chips' reflectance functions were approximated with a three-dimensional linear model where the three basis functions represent the first three principal components of the entire Kelly, Gibson & Nickerson (1943) data set. Fig. 3A shows the CIE x, y coordinates of the 16 surfaces which were used as test surfaces in this experiment. Their luminance varied between 7.0 and 21.0 cd/m^2. The coordinates are plotted that result when the surfaces are rendered under the standard illuminant. Fig. 3B shows the CIE x, y coordinates of the 24 surfaces which surrounded the test surface and matching field within the display. Their luminance varied between 2.0 and 25.6 cd/m^2. Again the coordinates are plotted that result when the surfaces are rendered under the standard illuminant (see also Table A2 in the Appendix).

Both the experimental illuminants and the experimental surfaces are described by three-dimensional linear models. Based on this type of modeling, for each illuminant ε a so-called light transformation matrix, Δ_ε, can be defined. This 3×3 matrix depends on the illuminant ε and provides a mapping from each 3×1 column vector ρ, which represents a surface, to the cone absorptions that result from this surface when rendered under illuminant ε (Wandell, 1995). The light transformation matrix was computed for the standard illuminant and each of the four test illuminants and these matrices were used to compute for each experimental surface the Smith & Pokorny (1975) cone coordinates that result when the surface is rendered under one of the illuminants. These coordinates were used to simulate the illuminated surfaces on the monitor. In addition, the coordinates were used to determine for each illuminant change the matches that a perfectly color constant

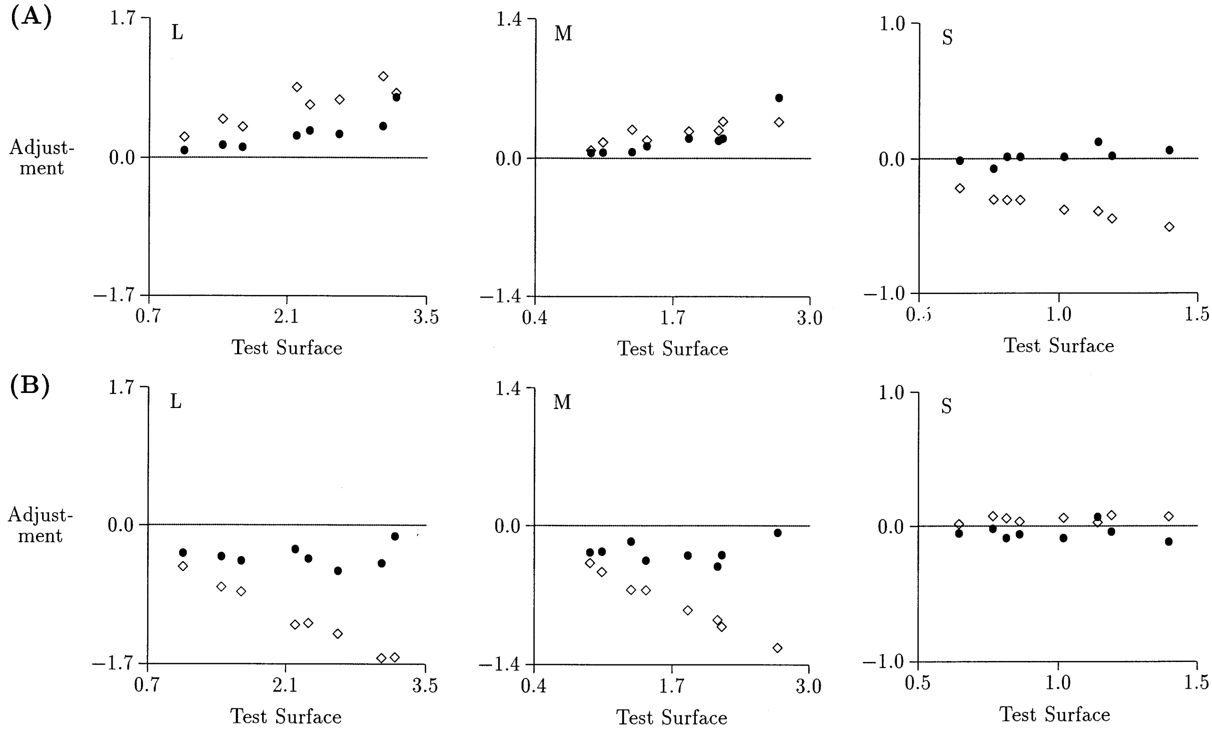

Fig. 4. Appearance matches. (A) Matches of subject CL set for the bright yellow illuminant. (B) Matches of subject NC set for the dark blue illuminant. The LMS coordinates of the subjects' adjustment to the illuminant changes are shown as a function of the coordinates of eight test surfaces. The subjects' matches are represented by the solid symbols (●). The open symbols (◇) represent the matches that a perfectly color constant observer would have set. To the extent that, for each of the three cone types, the subjects' matches fall on the horizontal line, they indicate that there is no illuminant adjustment at all; to the extent that the subjects' matches and the theoretical matches overlap, complete illuminant adjustment is indicated.

observer would have set to each of the 16 test surfaces. Comparing these theoretical matches with the matches a subject really sets provides information on the degree of the subject's illuminant adjustment.

2.1.3. Subjects

Seven subjects took part in the experiment. They had all normal color vision. With the exception of one subject (TE), they were naive about the purposes of the experiment.

2.1.4. Procedure

Each of the subjects set matches to eight different test surfaces. Three of the subjects set appearance matches: subjects CL and TK set appearance matches between the standard and one bluish and one yellowish test illuminant, and subject NC set appearance matches between the standard and all four test illuminants. Four of the subjects set surface matches: two of them (EM, TE) set surface matches between the standard and all four test illuminants, the other two (AR, BG) between the standard and one bluish and one yellowish test illuminant[1].

[1] Four of the subjects also set matches when the two Mondrians were equally illuminated (symmetric matches). These matches served as a control for possible left-right asymmetries in the matches. Data here were omitted as no noteworthy asymmetry was found for any of the subjects.

In each experimental session only one test illuminant was presented to each subject. Subjects made two matches to each of the eight test surfaces during a session. Each test illuminant was presented in two to four different sessions to each subject, resulting in four to eight settings for each test surface under each test illuminant. Each subject began an experimental session with 1 min of dark adaptation. Then the two Mondrians were presented to the subject. The subjects did not adapt to the Mondrians and immediately started setting either appearance or surface matches.

To set an appearance match, the subjects were instructed to adjust the matching field so that it had the same hue, saturation, and brightness as the test field. They were instructed to disregard, as much as possible, other areas of the screen. To set a surface match, the subjects were instructed to adjust the matching field so that it looked as if it were cut from the same piece of paper as the test field. It was pointed out that the two displays consisted of identical surfaces, which were only differently illuminated. Subjects were informed that taking the colors of other patches within a Mondrian into account might be helpful for this task (Arend & Reeves, 1986). While setting the matches, the subjects were instructed to spend about the same time looking at each Mondrian and to alternate between the two

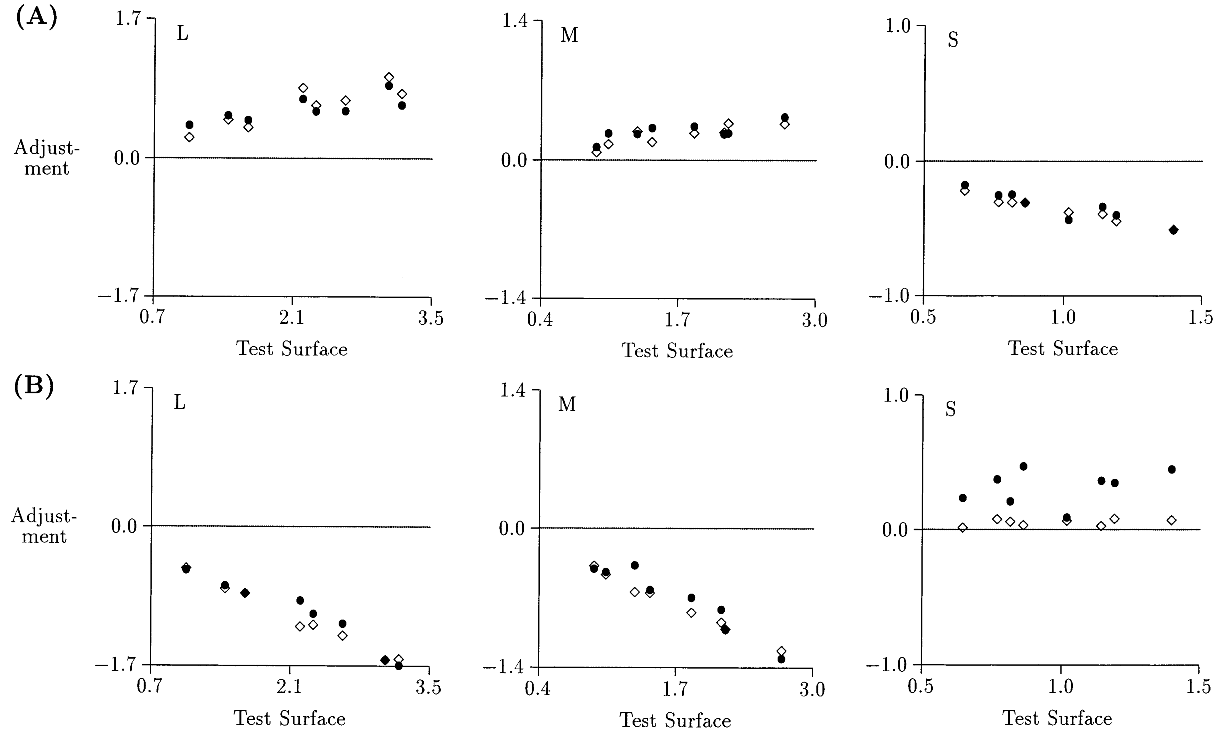

Fig. 5. Surface matches. (A) Matches of subject EM set for the bright yellow illuminant. (B) Matches of subject TE set for the dark blue illuminant. The LMS coordinates of the subjects' adjustment to the illuminant changes are shown as a function of the coordinates of eight test surfaces. The subjects' matches are represented by the solid symbols (●). The open symbols (◇) represent the matches that a perfectly color constant observer would have set. To the extent that, for each of the three cone types, the subjects' matches fall on the horizontal line, they indicate that there is no illuminant adjustment at all; to the extent that the subjects' matches and the theoretical matches overlap, complete illuminant adjustment is indicated.

displays in a roughly 2-s period. This was done to minimize adaptation to each of the two Mondrians. On average, each of the seven subjects made about 200 matches.

2.1.5. Data analysis

Tests of several models of the asymmetric matching data are reported. To choose the best parameters for each model, the difference between theoretically predicted and empirically observed matches was minimized using an error term that is normalized by the estimated covariance matrix, Λ_T, of a subject's match settings under a certain test illuminant T (Mahalanobis metric). Suppose the column vector, e_i, denotes the difference between the predicted and observed match coordinates. Then, the model parameters were estimated subject to minimization of the quantity

$$\frac{1}{n}\sum_{i=1}^{n}(e_i^t\Lambda_T^{-1}e_i)^{1/2}$$

where n is the number of matches that a subject set under test illuminant T. Brent's algorithm was used to perform the error minimizations (Gegenfurtner, 1992). Intuitively, this error measure is equivalent to (a) transforming the model deviations into a new coordinate frame where the distribution of errors are independent and have unit variance; and (b) using the Euclidean distance in that coordinate frame as the error measure. This approach to model fitting had already been used by Poirson and Wandell (1993) and Bäuml and Wandell (1996).

2.2. Results

2.2.1. Appearance matches

Fig. 4 shows the mean appearance matches of two subjects for one test illuminant each. The Smith & Pokorny (1975) LMS coordinates of the subjects' adjustments to the illuminant change are plotted as a function of the coordinates of eight test surfaces. The subjects' matches are represented by the solid symbols. The open symbols in the graphs represent theoretical matches, namely the matches that a perfectly color constant observer would have set. To the extent that, for each of the three cone types, the subjects' matches fall on the horizontal line, they indicate that there is no illuminant adjustment at all. To the extent that, for each of the three cone types, the subjects' matches and the theoretical matches overlap, complete illuminant adjustment is indicated.

The subjects' appearance matches show a tendency towards color constancy. Indeed, in many cases the

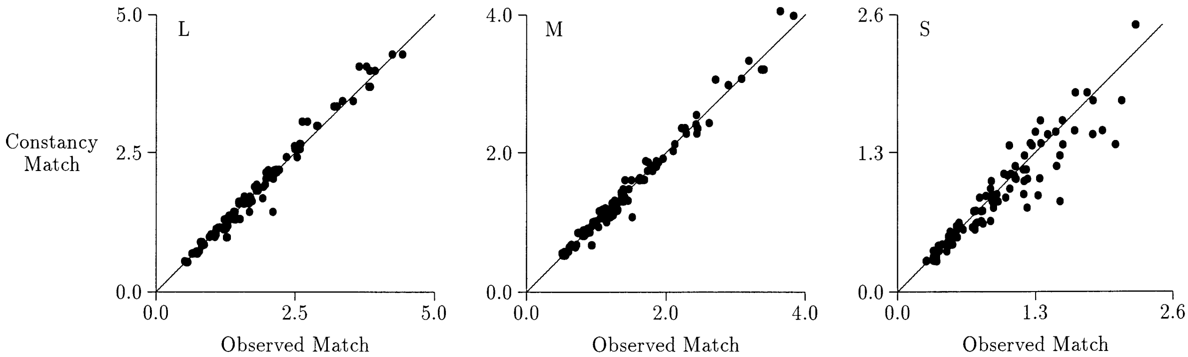

Fig. 6. Degree of constancy of the subjects' surface matches. The LMS coordinates of the surface matches of all four subjects are compared with the LMS coordinates of the matches that a perfectly color constant observer would have set. To the extent that the points deviate from the diagonal they indicate deviations from complete illuminant adjustment.

matches shift in the right direction. These adjustments, however, are generally small and are far from perfectly color constant matches. This pattern of results also holds for the other test illuminants under which the subjects set appearance matches and for subject TK, whose results are similar to those of subjects CL and NC.

The method proposed by Arend, Reeves, Schirillo and Goldstein (1991) was used to quantify the degree of constancy that the subjects showed. In this method two Euclidean distances are measured: (1) the distance between the test surfaces and perfectly color constant matches (u); and (2) the distance between perfectly color constant matches and the matches the subject set under the test illuminant (v). Based on these distances the term $1 - v/u$ is interpreted as a constancy index. Typically this method results in a constancy index between 0 and 1, whereby 0.0 reflects no adjustment at all to the illuminant and 1.0 reflects complete adjustment. The distances were computed using CIELUV metric with the color coordinates of the respective test illuminant as the nominally white light (Wyszecki & Stiles, 1982). A mean constancy index of 0.21 was found for subject NC, 0.28 for subject CL, and 0.15 for subject TK. These numbers reveal a low degree of illuminant adjustment.

2.2.2. Surface matches

Fig. 5 shows the mean surface matches of two subjects for one test illuminant each. The symbols within each graph have the same meaning as in Fig. 4, however, the solid symbols now represent surface matches rather than appearance matches. The matches show a clear shift towards color constancy. Moreover, in many cases there is a considerable overlap between the two types of symbols, which indicates a high level of constancy. This pattern of results also generalizes to the other test illuminants under which the subjects set surface matches, and similarly holds true for the other two subjects, AR and BG. Again, constancy indices for the subjects' matches were computed. A mean constancy index of 0.75 for subject EM, 0.81 for subject TE, 0.83 for subject AR, and 0.73 for subject BG was found. These numbers reveal a high degree of illuminant adjustment[2].

The surface matches show extensive color constancy. However, as can be seen from Fig. 5B, there may also be some clear deviations from constancy. These deviations are to some degree systematic. Indeed, all four subjects show a very high level of constancy in their L and M cone data, and it is only for some of the S cone data that major deviations from constancy arise. Fig. 6 illustrates this point. In this figure the LMS coordinates of all the surface matches from all four subjects are compared with the corresponding theoretical perfectly color constant matches. Perfect constancy, with respect to each cone class, is indicated by the extent to which all the points in each graph fall on the diagonal. As can be seen, the deviations from constancy are relatively small for the L and M cone data, while there are some clear deviations for the S cone data.

2.3. Models

2.3.1. Appearance matches

As mentioned above, there is strong evidence that, at least when adaptation is more or less complete, the effect of illuminant changes on the color appearance of objects is roughly consistent with the von Kries principle. Whether the appearance matches in the present paradigm, in which adaptation is quite reduced, are

[2] In informal observations, the observed difference between appearance and surface matches were established as, is not caused by the fact that different subjects set the appearance and surface matches in this experiment. Two of the subjects who set appearance matches also made a couple of surface matches (CL, TK), and two of the subjects who set surface matches also made a couple of appearance matches (AR, TE). As expected, all four subjects' appearance matches showed a small amount of illuminant adjustment and all four subjects' surface matches showed a large amount of illuminant adjustment.

Table 1
Residual errors for model fits

Subject	Task	Precision	Affine	Linear	von Kries	Perfect	Effect
NC	CA	1.52	1.87	2.13	2.54	13.72	5.15
CL	CA	1.47	1.83	2.05	2.20	12.67	4.63
TK	CA	1.46	1.55	1.68	1.70	7.61	3.57
EM	SC	1.48	1.66	1.74	1.98	2.68	11.17
TE	SC	1.52	1.88	1.97	2.32	3.02	16.32
AR	SC	1.34	1.94	2.08	2.56	2.97	15.64
BG	SC	1.44	1.59	1.62	1.85	3.47	12.34

The mean residual errors for the affine linear model, the linear model, the von Kries model, and the perfect-constancy model are shown, separately for each subject and matching task.
CA, color appearance; SC, surface color.
In addition, the precision of the matches and the size of the illuminant effect on the subjects' matches are shown.

also in agreement with this principle, was examined. A series of nested models was proposed—the affine linear model, the linear model, and the von Kries principle—and the extent to which these models account for the illuminant effect in this experiment was tested.

The affine linear model assumes that, after discounting some background level from each of the two Mondrians (Walraven, 1976; Shevell, 1978), the cone coordinates of the test surfaces, t, are linearly transformed into the cone coordinates of the matching field, m:

$$m = N_\varepsilon t + \gamma_\varepsilon,$$

where N_ε is a 3×3 matrix, γ_ε is a 3×1 column vector, and both N_ε and γ_ε vary with the test illuminant, ε. This model has 12 parameters for each illuminant change. Two more restrictive models were constructed. In the first step, the affine linear model was restricted by assuming that no background level is discounted from the Mondrians, i.e. $\gamma_\varepsilon = 0$. This results in a model with nine parameters for each illuminant change, in which the cone coordinates of the test surfaces are now linearly mapped into the cone coordinates of the matches. In the second step, the additional restriction was imposed that the matrix N_ε be diagonal. That is, consistent with the von Kries model, we assume that the adjustment to an illuminant change be in terms of the three cone coordinates. This model has only three parameters to account for a change in illumination.

Table 1 shows the residual errors that were found for each of the three models, separately for the three subjects. They were computed using the Mahalanobis metric (see Section 2.1). Averaged over subjects, an error of 1.75 units for the affine linear model, 1.95 units for the linear model, and 2.15 units for the von Kries model was found. The precision of the matches was 1.48 units. These results closely parallel those that are typically found for appearance matches when adaptation is fairly complete (Brainard & Wandell, 1992; Bäuml, 1995). They suggest that all three nested models provide a good description of the data and do about equally well. In fact, at least in a first approximation, the von Kries principle is reasonably close to the precision of the subjects' matches. Fig. 7A shows scatterplots comparing the whole set of appearance matches of one subject (NC) with the predictions of these matches based on this principle. There is good agreement between the matches and predictions. Table 2 provides the von Kries coefficients that were estimated for each of the three subjects under the single test illuminants.

Fig. 4 and the constancy indices reported above demonstrate that the subjects' appearance matches are far from perfectly constant. To quantify the deviation from constancy in terms of the models proposed above, the error was computed when using the perfectly color constant (theoretical) matches to predict the subjects' matches. This model is a special case of a linear model. Averaged over subjects, an error of 11.33 units was found (Table 1), which indicates a poor fit of the perfect constancy model. The fact that the error is even larger than the size of the effect in this experiment tells us that assuming complete illuminant adjustment leads to a much worse prediction of the matches than assuming that there is no illuminant adjustment at all.

2.3.2. Surface matches

The same set of nested models as used for the appearance matches were tested, in order to account for the subjects' surface matches. Table 1 shows, separately for the four subjects, the residual errors that were found for the three models. The pattern of results is quite similar to the one found for the appearance matches. Averaged over subjects, an error of 1.77 units for the affine linear model, 1.85 units for the linear model, and 2.18 units for the von Kries model was found. The precision of the matches was 1.44 units. Again, the results suggest that all three models do about equally well, and that all three models provide a reasonable description of the data. Fig. 7B shows scatterplots comparing subject TE's surface matches under

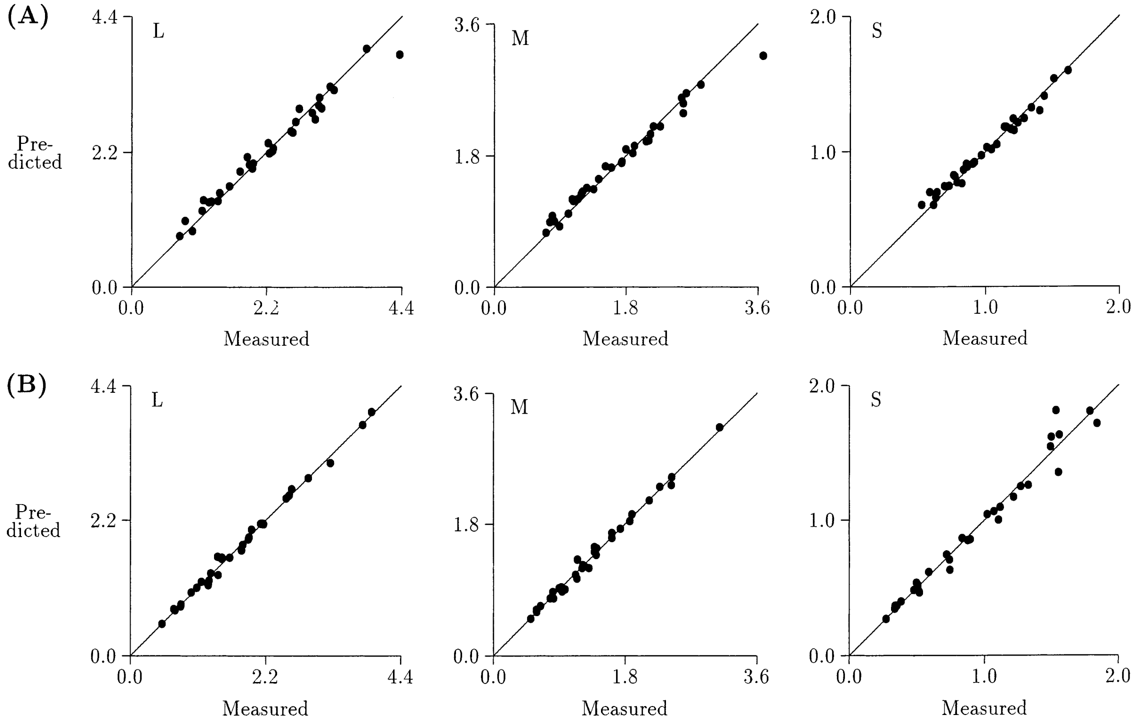

Fig. 7. Scatterplot of von Kries model fit. Scatterplots are shown which compare the LMS coordinates of subject NC's mean appearance matches (A) and of subject TE's mean surface matches (B) with the predictions of these matches based on the von Kries model. The data for all four illuminant changes are shown. If the model held perfectly for a type of matches, all data points would fall on the diagonal line.

all four test illuminants with the predictions of these matches based on the von Kries model. Again, there is good agreement between the matches and predictions. Table 2 provides the von Kries coefficients that were estimated for each of the four subjects under the single test illuminants.

Fig. 6 above reveals that in some cases the subjects' matches deviate from perfectly color constant matches, particularly with respect to the S cone data. To quantify these deviations, the error was computed when using the perfectly color constant theoretical matches, in order to predict the subjects' matches. Averaged over subjects, the residual error of this model was 3.03 units (Table 1).This error is about one unit larger than the error from the linear model or the von Kries model, which indicates that these weaker models provide a substantially better description of the adjustment than the assumption of complete illuminant adjustment. Still, the results confirm the above finding that the subjects' matches show a large amount of adjustment.

2.3.3. Comparing appearance and surface matches

Fig. 8 serves as a summary of the results of the present experiment. It shows, separately for the two types of tasks, the mean error of the von Kries principle in predicting the subjects' matches compared to the subjects' precision in the task and the size of the subjects' adjustment to the illuminant changes (effect). In addition, it shows the quality of fit of the less restrictive hypotheses of an affine linear adjustment and a general linear adjustment. Three points may be stressed. First, consistent with what has been shown in Fig. 7, the subjects' matches, both appearance and surface, are roughly consistent with the von Kries principle. Second, the von Kries principle accounts about as well for the data as do the more general hypotheses of an affine linear and a linear adjustment. Third, and most interesting, the pattern of results is the same for the appearance and the surface matches, with only one big difference: The surface matches show a much higher amount of adjustment to the illuminant changes than the appearance matches.

2.3.4. Von Kries scaling versus scaling in another color space

The von Kries principle assumes that the adjustment to an illuminant change is in terms of the three cone types. We might also think about an opponent version of this principle, proposing that the adjustment is in terms of three opponent colors signals. The surface match data were examined to see if they are better described by the original cone-based version of the principle or an opponent version. The Jameson and Hurvich (JH) opponent space (Jameson & Hurvich, 1955) and the Derrington, Krauskopf and Lennie (DKL) space (Derrington, Krauskopf & Lennie, 1984) was used for this analysis. In addition, the extent of Finlayson, Drew and Funt's (FDF) space of spectrally

Table 2
Estimated von Kries coefficients

Subjects	Task	T_1			T_2			T_3			T_4		
		L	M	S	L	M	S	L	M	S	L	M	S
NC	CA	0.80	0.81	0.92	0.90	0.91	0.96	0.88	0.86	0.80	1.14	1.12	1.03
CL	CA	0.83	0.85	0.95	—	—	—	—	—	—	1.19	1.15	0.97
TK	CA	—	—	—	0.93	0.95	1.04	0.84	0.83	0.89	—	—	—
EM	SC	0.46	0.51	0.98	0.76	0.81	1.21	0.70	0.67	0.43	1.22	1.15	0.66
TE	SC	0.49	0.55	1.22	0.81	0.86	1.31	0.69	0.65	0.44	1.23	1.14	0.63
AR	SC	0.52	0.57	0.98	—	—	—	—	—	—	1.27	1.17	0.70
BG	SC	0.54	0.59	1.09	—	—	—	—	—	—	1.24	1.17	0.83

The three parameters that resulted when fitting the von Kries model to the subjects' matches are shown, separately for each subject and test illuminant.
CA, color appearance; SC, surface color; T_i, test illuminant i; L, gain of L cone class; M, gain of M cone class; S, gain of S cone class. A scale factor of 1.0 means that there is no adjustment with respect to this cone class.

sharpened sensors was tested (Finlayson, Drew & Funt, 1994) might still improve the fit of the original von Kries principle. This space was deduced from a computational approach to color constancy, to provide the most complete illuminant adjustment that may result from an independent adjustment of three color codes. It differs from cone space mainly through the fact that the long-wavelength sensor is pushed further toward the long-wavelength end of the spectrum, that it shows some degree of red-green opponency, and is spectrally sharpened.

Averaged over subjects, an error of 5.89 units for the JH space, 3.04 units for the DKL space, and 2.22 units for the FDF space was found. The mean error for the original von Kries principle was 2.18 units (see above).Thus, the two opponent versions of the principle do worse than the original cone-based version of the principle. This is particularly obvious for the JH space, while the DKL space leads to an error which is similar to the one that results from the perfect color constancy model (3.03 units)[3]. The FDF space leads to about the same error as the cone space, which is not too surprising because the two spaces do not differ very much.

3. Experiment 2

3.1. Method

The methods used in this Experiment 2 are largely identical to those used in Experiment 1. The main difference is that fewer experimental illuminants and more experimental surface collections were used than in Experiment 1.

3.2. Experimental illuminants and surfaces

Three illuminants from Experiment 1 were used: the standard illuminant, which rendered the left-hand Mondrian, and the dark bluish (T_1) and the bright yellowish (T_4) test illuminants, which both rendered the right-hand Mondrian (see Fig. 2 and Table A1 in the Appendix).

Three surface collections were used: a standard collection, which is identical to the surface collection used in Experiment 1, and two test collections. One of the two test collections, referred to as the bright collection, had similar average chromaticity as the standard collection ($x = 0.343$, $y = 0.302$ vs. $x = 0.321$, $y = 0.333$), but had a higher average luminance than the standard collection ($L = 19.65$ vs. $L = 10.47$). The other test collection had a chromaticity that was on average reddish ($x = 0.413$, $y = 0.314$) and a somewhat smaller average luminance ($L = 7.42$) relative to the standard collection; this collection is called the red collection.

Fig. 9 shows the CIE x, y chromaticity coordinates of all the surfaces from the three experimental surface collections when rendered under the standard illuminant (see also Table A2 in the Appendix). All the surfaces were drawn from the large set of surfaces used by Brainard and Wandell (1992) and are approximations of Munsell papers. Finally, eight of the 16 test surfaces that were used in Experiment 1 were also used in this experiment.

3.2.1. Subjects

Seven subjects took part in the experiment. They had normal color vision and were all naive about the purpose of the experiment[4]. Each of the subjects set

[3] A few other opponent spaces, like those discussed in the studies of Poirson & Wandell (1993) and Bäuml & Wandell (1996) were tested. In all cases the average residual error was larger than 3.0 units.

[4] Another subject started with the experiment but decided to quit after a few practice sessions. She felt she could not find any criterion to set reliable surface matches. Previous studies reported on subjects, who could not set surface matches across differently illuminated Mondrians, as well (Arend & Reeves, 1986; Cornelissen & Brenner, 1995).

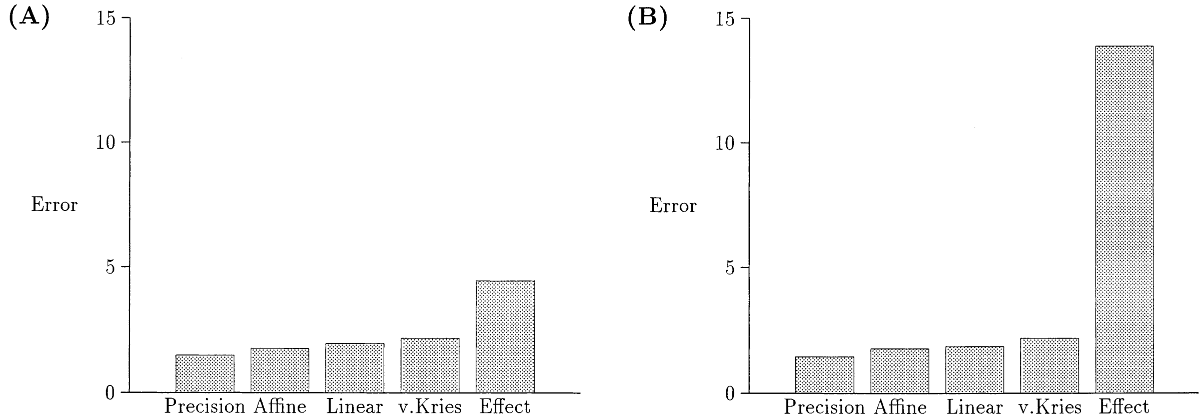

Fig. 8. Quality of model fits. The mean residual errors for the affine linear model, the linear model, and the von Kries model are shown together with the precision of the subjects' matches and the size of the illuminant effect on the subjects' matches. (A) Appearance matches. (B) Surface matches.

matches to the eight test surfaces. Five of them set matches between the standard and the yellowish test illuminant: Subjects MC and MK set appearance matches, and subjects BK, CM, and MR set surface matches. Two of the subjects set matches between the standard and the bluish test illuminant: Subject BP set appearance matches, and subject ST set surface matches. All seven subjects set their matches in all three experimental surface collections, that is, when the display's two identical Mondrians consisted of the standard collection, the bright collection, and the red collection. Each of the seven subjects made about 160 matches, on average.

3.3. Results

3.3.1. Degrees of constancy

In Experiment 1 subjects' surface matches showed much more illuminant adjustment than subjects' appearance matches (78 vs. 20%). This feature can also be met in this Experiment 2. The method proposed by Arend, Reeves, Schirillo and Goldstein (1991) to compute constancy indices was used. In order to maximize comparability of results with those from Experiment 1, the constancy indices were computed only for the matches set in the standard collection. In the appearance matching task a constancy index of 0.16 for subject MC, 0.27 for subject MK, and 0.32 for subject BP was found. In the surface matching task a constancy index of 0.80 for subject BK, 0.81 for subject CM, 0.82 for subject MR, and 0.74 for subject ST was found. So, on average, there is again a much higher degree of adjustment in the surface matching task (79%) than in the appearance matching task (25%).

3.3.2. Appearance matches

Fig. 10A, B show the mean appearance matches of subject MC (yellowish illuminant) for the three surface collections. Panel (A) compares the matches set in the standard collection with those set in the bright collection, panel (B) compares the matches set in the standard collection with those set in the red collection. Similarly, Fig. 10C, D show the mean appearance matches of subject BP (bluish illuminant), where panel (C) compares the matches set in the standard collection with those set in the bright collection, and panel (D) compares the matches set in the standard collection with those set in the red collection. In each panel, the Smith & Pokorny (1975) LMS coordinates of the subjects' adjustments to the illuminant change are plotted as a function of the coordinates of the eight test surfaces. The solid symbols represent the matches set in the standard collection, the triangles those set in the bright collection, and the plus signs those set in the red collection. To the extent that, for each of the three cone classes, the types of symbols overlap, they indicate that there is no effect on the subjects' adjustment to an illuminant change as a result of a change in surface collection.

The surface collection systematically influenced subject MC's adjustment to the yellowish illuminant. This is particularly obvious when the matches set in the standard collection are compared with those set in the red collection (Fig. 10B). In general, MC needed higher L, M, and S coordinates in the red collection than in the standard collection in order to set matches to the single test surfaces. There is also a slight indication that MC needed higher L, M, and S coordinates in the standard collection than in the bright collection (Fig. 10A). A similar pattern of results arose for subject MK. His S cone data, however, were practically the same in the standard and bright collection.

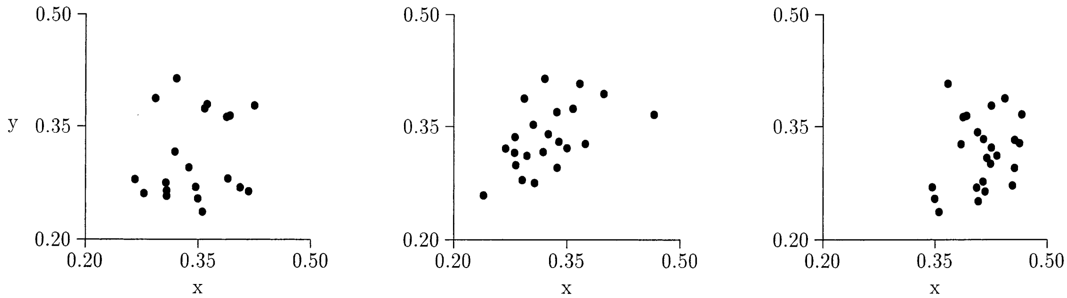

Fig. 9. Experimental surface collections. Each graph shows the CIE x, y chromaticity coordinates of the 24 surfaces of one experimental surface collection. The left graph shows the standard collection, the middle graph the bright collection, and the right graph the red collection. The coordinates of the surfaces that result when they were rendered under the standard illuminant are shown (see also Table A2 in the Appendix).

The surface collection had only a minor influence on subject BP's adjustment to the bluish illuminant. In tendency, subject BP needed slightly higher L and M coordinates in the standard than in the bright collection in order to set the matches (Fig. 10C). The difference between the matches set in the standard and those set in the red collection (Fig. 10D) was even smaller. These results for the bluish illuminant and the previous ones for the yellowish illuminant suggest that the surface collection can have a systematic effect on subjects' adjustments to an illuminant change.The size of this effect seems to vary somewhat with the illumination in the image. Over all, however, it is relatively small.

3.3.3. Surface matches

Fig. 11A, B show the mean surface matches of subject BK (yellowish illuminant) and Fig. 11C,D show the mean surface matches of subject ST (bluish illuminant). Panels (A) and (C) compare the matches set in the standard collection (solid symbols) with those set in the bright collection (triangles), panels (B) and (D) compare the matches set in the standard collection with those set in the red collection (plus signs).

As was true for the appearance matches, the surface matches show some variation with surface collection under the yellowish illumination. In general, subject BK needed higher L, M, and S coordinates to set the matches in the standard collection than in the bright collection (Fig. 11A), and also needed higher L and M coordinates in the red collection than in the standard collection (Fig. 11B).The same was true for subject MR with respect to the L and M cone data, but a different pattern arose with respect to the S cone data. She needed higher S coordinates in the red collection than in the standard collection and about the same S coordinates in the standard and the bright collection. Subject CM showed essentially the same picture as subject BK for the standard and the bright collection, but showed hardly any systematic differences between the matches set in the standard and those set in the red collection.

Again the surface collection had only slight effects on the adjustments to the bluish illuminant change. Indeed, the matches that subject ST set under the bluish illuminant were more or less the same in the standard and the bright collection (Fig. 11C). It is only for the red collection that, in general, ST needed slightly smaller L and M coordinates to set the matches than she needed in the standard collection (Fig. 11D). Again these results and the previous ones for the yellowish illuminant indicate that the surface collection can systematically influence the subjects' adjustments to an illuminant change. Again, however, this effect seems to vary with the illumination, and it is relatively small.

In summary, these results demonstrate that both the appearance and the surface matches can be affected by the image surfaces, and that the influence of image surfaces is quite similar in the two types of tasks. For instance, for both types of matches the influence of image surfaces is larger for the yellowish than for the bluish illuminant change. Or, changing the image surfaces from the standard to the red collection, on average, induces a larger effect in the illuminant adjustment than when changing the surfaces from the standard to the bright collection. Besides this qualitative similarity the matches are also similar in quantitative respects. This point is addressed in more detail in the next paragraphs.

3.4. Models

3.4.1. Illuminant adjustment

Both the linear model and the von Kries model were fitted to the subjects' matches, separately for each surface collection. For the appearance matches a mean residual error of 1.95 units for the linear model and of 2.31 units for the von Kries model was found. For the surface matches, a mean residual error of 1.91 units for the linear model and of 2.18 units for the von Kries model was found. The precision of both types of matches was about 1.5 units (see Table 3 for details). These results mirror those from Experiment 1 and indicate again that not only the linear model, but also the von Kries model provide a reasonable description of the two types of matches. Table 4 shows the von

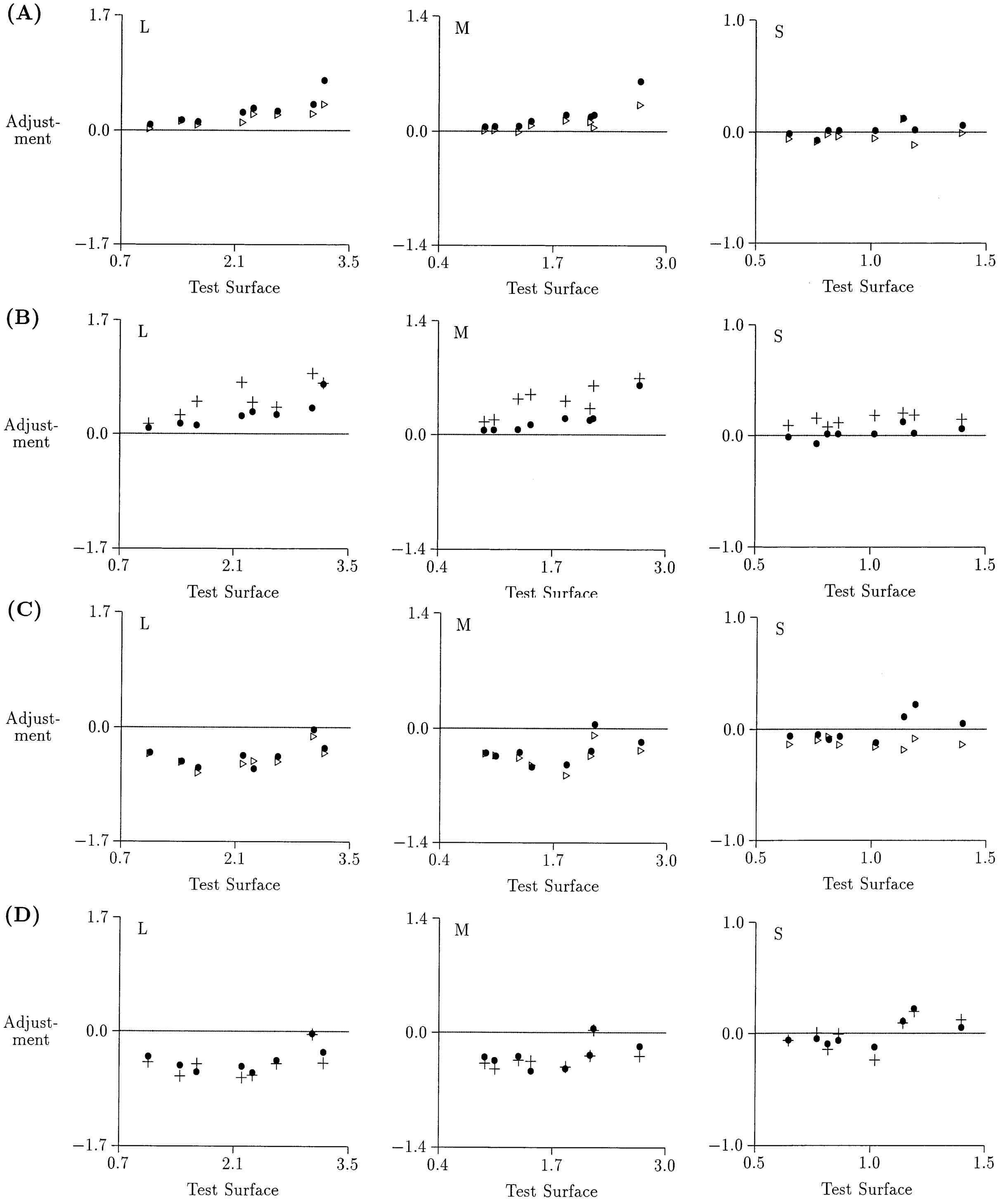

Fig. 10. Appearance matches in the three surface collections. (A), (B) Matches of subject MC set for the yellowish illuminant. (C), (D) Matches of subject BP set for the bluish illuminant. The LMS coordinates of the subjects' adjustment to the illuminant changes are shown as a function of the coordinates of the eight test surfaces (●, standard collection; ▷, bright collection; +, red collection).

Kries coefficients that were estimated for each of the subjects for the single surface collections.

3.4.2. *Influence of surface collection*

In the analyses of the preceding paragraphs, linear transformations were fitted to the subjects' appearance and surface matches, which were free to vary both with the illuminant and with the surface collection. By fitting linear transformations to the same data, which are only free to vary with the illuminant but not to vary with the surface collection, we gain some insight into the size of the influence of surface collection on the subjects' adjustments to illuminant changes.

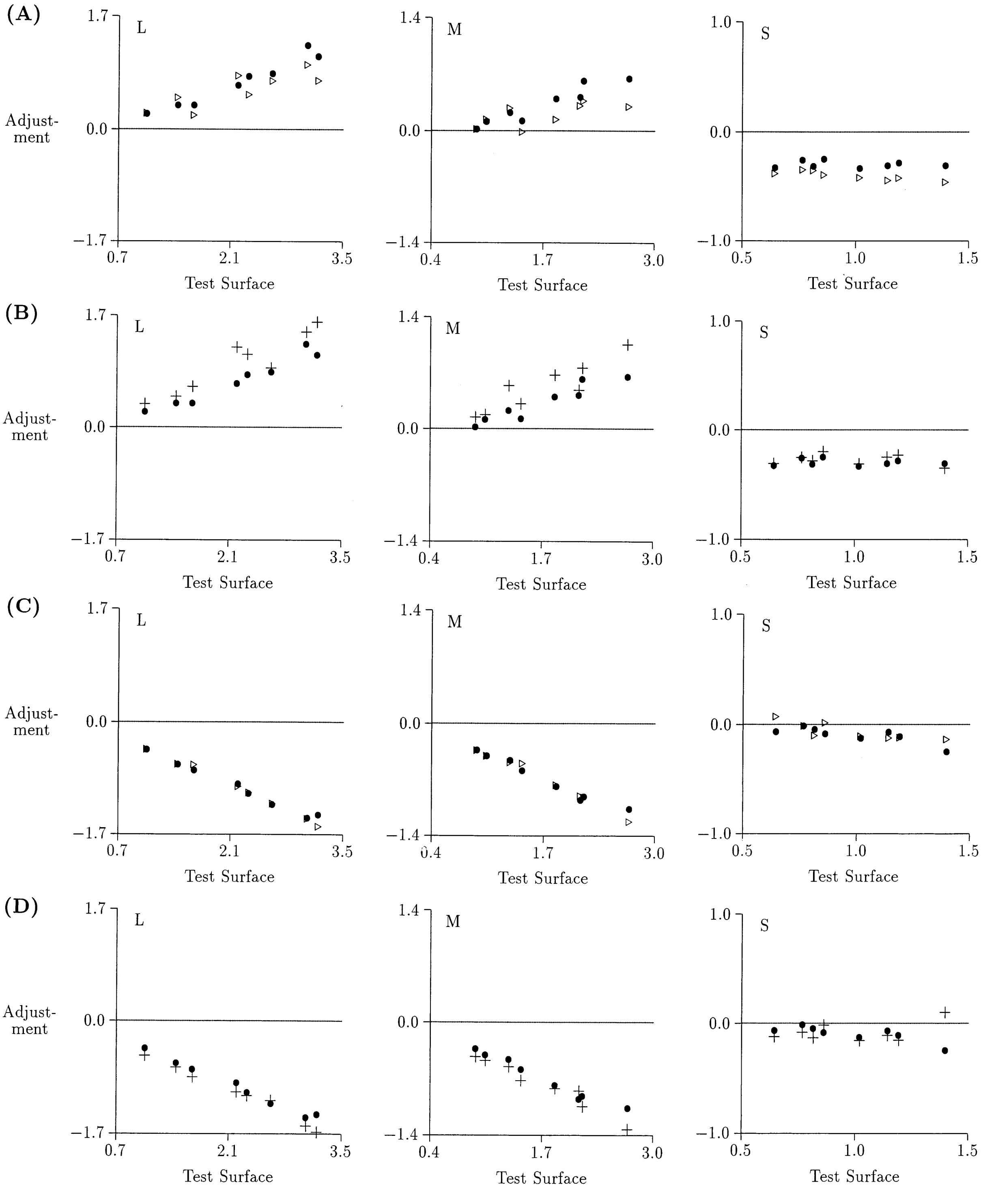

Fig. 11. Surface matches in the three surface collections. (A), (B) Matches of subject BK set for the yellowish illuminant. (C), (D) Matches of subject ST set for the bluish illuminant. The LMS coordinates of the subjects' adjustment to the illuminant changes are shown as a function of the coordinates of the eight test surfaces (●, standard collection; ▷, bright collection; +, red collection).

Fitting the restricted linear transformations to the appearance match data increased the average error of the linear model by 0.42 units and of the von Kries model by 0.27 units. Fitting the restricted transformations to the surface match data increased the average error of the linear model by 0.28 units and of the von Kries model by 0.19 units (Table 3)[5]. So, consistent with the visual impression from Figs. 10 and 11 and

[5] The matches set under the (dark) bluish illuminant were more reliable than the matches set under the (bright) yellowish illuminant,

Table 3
Residual errors for model fits

Subjects	Task	Illuminant	Precision	Linear (G)	von Kries (G)	Linear (R)	von Kries (R)	Effect
BP	CA	T_1	1.43	2.17	2.36	2.49	2.58	6.00
MC	CA	T_4	1.52	1.89	2.47	2.43	2.77	5.93
MK	CA	T_4	1.50	1.78	2.10	2.20	2.41	4.56
ST	SC	T_1	1.41	1.80	2.07	2.03	2.24	17.01
BK	SC	T_4	1.50	2.10	2.41	2.58	2.77	10.82
CM	SC	T_4	1.54	1.89	2.16	2.02	2.24	8.02
MR	SC	T_4	1.51	1.84	2.06	2.14	2.24	9.02

The mean residual errors for the linear model and the von Kries model are shown, separately for each subject and matching task.
CA, color appearance; SC, surface color; T_i, test illuminant *i*.
The errors are shown when the transformations were allowed to vary with surface collection (G, general) and when they were restricted to be the same across collections (R, restricted).
In addition, the precision of the matches and the size of the illuminant effect on the subjects' matches are shown.

Table 4, the effect of surface collection on the subjects' adjustments to an illuminant change is not very large.

3.4.3. Comparing appearance and surface matches

Fig. 12 may serve as a summary of the results of the present experiment. It compares, separately for the appearance and surface matches, the mean error of the von Kries model, when its parameters are free to vary with the surface collection, with that of the von Kries model, when its parameters are restricted to being the same across surface collections. Similar comparisons are shown when using the general linear model instead of the von Kries model. The errors from these models are compared with the precision of the matches and with the size of the subjects' adjustments to the illuminant changes (effect).

The figure shows three things. First, the von Kries model and the general linear model describe the matches about equally well. This holds both when the transformations are estimated individually for the single surface collections and when the restriction, that the surface collection does not affect the adjustment, is imposed. Second, the effect of surface collection on the description of the adjustment is relatively small. The increase in error that is introduced by ignoring the role of surface collection is comparable to the increase in error that is introduced when we accept the von Kries model as, as good a description of the adjustment as the general linear model. Third, the same pattern of results shows up for the appearance and the surface matches. Thus, the main difference between the two types of matches is again that the surface matches show a much higher amount of adjustment to the illuminant changes than the appearance matches.

thus leading to different covariance matrices. This explains why the visually larger effects under the yellowish illuminant did not lead to much larger errors in the analyses than the visually smaller effects under the bluish illuminant. this point holds for both the appearance and the surface matches.

4. Discussion

4.1. Appearance versus surface matches

The present study replicates findings from previous studies by showing that, in simultaneous color constancy, surface matches show a much higher level of illuminant adjustment than appearance matches. It goes beyond prior work by demonstrating that the adjustment in the surface matches is nearly complete in the L and M cone data—thus generalizing results from Arend and Goldstein (1987) who found nearly perfect lightness constancy in achromatic Mondrian patterns—and that deviations from perfect surface color constancy are mainly due to failures in the adjustment of the S cone signals. These failures may be the result of an underestimation or an overestimation of the illuminant effect (Fig. 5), and they may vary both within subjects, as a function of the test illuminant, and between subjects (Table 2). This variability in the S cone data is in agreement with a recent result by Nascimento & Foster (1998), who also found that the S cone signals are of minor relevance for the subjects' surface color constancy performance. Notice, however, that also appearance matches show more variability in the S cone than in the L and M cone data (Tables 2 and 4).

Just as in the previous studies by Arend and colleagues (Arend & Reeves, 1986; Arend, Reeves, Schirillo & Goldstein, 1991), in the present experiments quite low amounts of illuminant adjustment were found in the appearance matching task-on the order of 25%. Recently, Brainard, Brunt and Speigle (1997) reported an asymmetric color matching experiment in which subjects set appearance matches in more natural images. On average, these researchers found degrees of constancy which were substantially higher than the ones Arend and colleagues and this study found—on the order of 60%. The higher level of constancy that Brainard and colleagues reported might have to do with the more natural stimulus conditions that they used

Table 4
Estimated von Kries coefficients

Subjects	Task	Illuminant	Standard			Bright			Red		
			L	M	S	L	M	S	L	M	S
BP	CA	T_1	0.79	0.82	0.99	0.72	0.75	1.02	0.74	0.79	0.98
MC	CA	T_4	1.13	1.11	1.04	1.10	1.07	0.99	1.27	1.28	1.17
MK	CA	T_4	1.14	1.10	0.95	1.11	1.06	0.95	1.21	1.21	1.05
ST	SC	T_1	0.54	0.58	0.88	0.53	0.57	0.91	0.49	0.53	0.88
BK	SC	T_4	1.33	1.20	0.71	1.27	1.13	0.61	1.43	1.30	0.73
CM	SC	T_4	1.30	1.17	0.73	1.22	1.11	0.65	1.26	1.13	0.70
MR	SC	T_4	1.27	1.23	0.74	1.22	1.18	0.76	1.39	1.39	0.86

The three parameters that resulted when fitting the von Kries model to the subjects' matches are shown, separately for each subject and surface collection.
CA, color appearance; SC, surface color; T_i, test illuminant i; standard, standard surface collection; bright, bright surface collection; red, red surface collection; L, gain of L cone class; M, gain of M cone class; S, gain of S cone class.
A scale factor of 1.0 means that there is no adjustment with respect to this cone class.

(Kaiser & Boynton, 1996). However, other possibilities remain as well, since their experimental setup differed in several ways from the one used in the present study. For instance, Brainard and colleagues presented relatively rich scenes with quite large test fields and spatially gradual illuminant changes; here relatively reduced scenes with rather small fields and abrupt illuminant changes were presented. In fact, the role of degree of naturalness of images on color constancy performance is still an open issue (Bäuml, 1997; Brainard, 1998).

4.2. von Kries adjustment

The appearance matches of both experiments are well described by simple parametric models. In particular, a model based on the notion of von Kries adjustment provides a good description of the data. One of the major questions of this study was whether we need different models to describe appearance and surface matches in simultaneous color constancy, or whether we can use the same class of models to account for the two types of matches. Of particular interest was the question of whether there is a principle similar to the von Kries principle which can account for the surface matches. The present data provide clear-cut answers on these questions. They demonstrate that the surface matches can be described by the same set of models as the appearance matches, and that the von Kries principle provides an equally good explanation of the appearance match data as of the surface match data.

Finding appearance matches to be roughly consistent with the von Kries notion has often been interpreted based on the idea that the adjustment takes place at the level of the photoreceptors in our retina. This interpretation of the von Kries principle may be adequate for the constancy mechanisms involved in setting appearance matches (Werner & Walraven, 1982; Chaparro, Stromeyer, Chen & Kronauer, 1995; Chichilnisky & Wandell, 1995). It seems unlikely, however, that it is appropriate in the case of surface matches. Indeed, we expect that the adjustment which underlies the surface matches occurs at much higher levels of the visual system (Cornelissen & Brenner, 1995). One possibility would be that the cone signals, after being recoded into opponent signals still within the retina, are separated again in a more central representation, or at least close versions of the cone signals (Zeki, 1993; Finlayson, Drew & Funt 1994), and that the adjustment which underlies the surface matches is largely in terms of these separated signals. But, of course, other possibilities remain as well.

Whatever the physiological explanation of the von Kries adjustment in the case of surface matches, the cone signals that result from an image's surfaces carry useful information for the visual system's inference about whether a contextual change reflects a change in illumination or a change in surfaces, and our visual system seems to be able to use this information. Corresponding evidence has also been provided in two recent studies by Foster and Nascimento (1994) and Nascimento and Foster (1998). In a simulation study, Foster & Nascimento (1994) examined how the cone ratios of two adjacent surfaces, x and y, vary as a function of changes in illumination. While changes in illumination have a large effect on the absolute cone values that result from the illuminated surfaces, they found that illuminant changes leave the cone ratios L_x/L_y, M_x/M_y, and S_x/S_y largely invariant. Furthermore, Nascimento and Foster (1998) presented subjects with pairs of successive images of Mondrian patterns undergoing illuminant changes, in which one of the changes was adjusted so that cone ratios for any two surfaces in the Mondrian were held constant. Subjects were asked which of the two changes was more like an illuminant change. It was found that subjects systematically

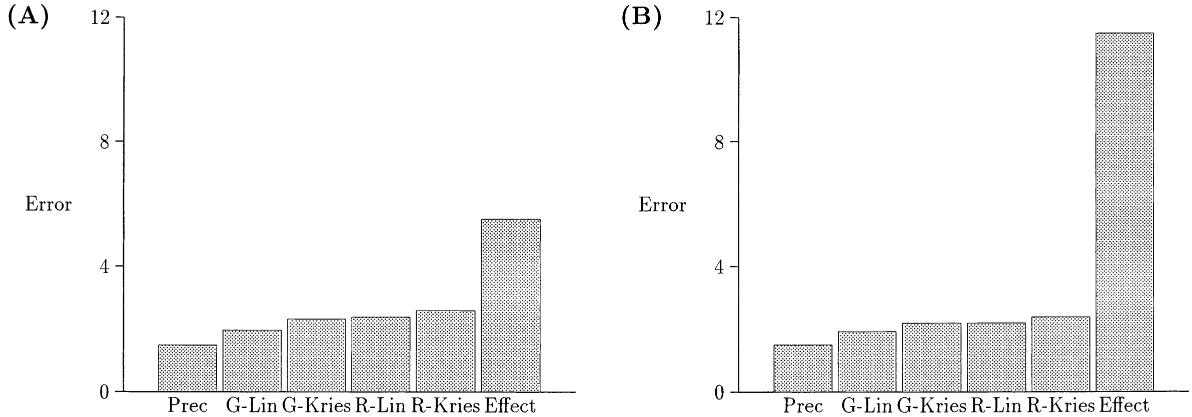

Fig. 12. Quality of model fits. The mean residual errors for the linear model (Lin) and the von Kries model (Kries) are shown, when the model transformations are free to vary with surface collections (G, general) and when they are restricted to be constant across surface collections (R, restricted). In addition, the precision of the matches and the size of the illuminant effect on the subjects' matches are shown. (A) Appearance matches. (B) Surface matches.

misidentified the changes with the corrected cone ratios as the illuminant changes, even when those corrected images corresponded to highly improbable natural events.

These results together with those from the present experiments suggest that von Kries adjustment is an appropriate model to account for the illuminant adjustment which underlies surface matches. At least in two respects, however, such a view is a simplification. First, the von Kries principle provides a reasonable description of the surface matches but can not account for all the variation in the data. This feature is not unique to the surface matches and does equally hold for appearance matches, in both successive color constancy (Brainard & Wandell, 1992; Bäuml, 1995) and simultaneous color constancy (Brainard, Brunt & Speigle, 1997) (Figs. 8 and 12). Second, the experimental data can not distinguish between an adjustment in terms of the cone classes and an adjustment in terms of three closely related color codes, like, for instance, the spectrally sharpened sensors proposed by Finlayson, Drew and Funt (1994). In fact, the assumption of a scaling of the responses of these sharpened sensors provides an equally good account of the data as the assumption of a scaling of the responses of the three cone classes. The data, however, can distinguish between the cone-based version of the von Kries principle and opponent versions of the principle. All opponent spaces that were examined provided a poorer fit to the matches than the cone space.

4.3. Influence of surface collection

In Experiment 2 the illuminant adjustment of both appearance and surface matches across three different surface collections were compared. Both the appearance and the surface matches were influenced by the surface collection in the image, and this influence was comparable between the two types of matching tasks, both in pattern and in size. Thus, in both types of tasks, the adjustment to an illuminant change was not only a function of illumination but also a function of the image surfaces. This dependence on the image surfaces, however, was not very large. The error that was introduced by ignoring the role of surface collection was comparable to the increase in error that is introduced when the von Kries model is accepted as, as good a description of the adjustment as the general linear model.

The conclusion that in simultaneous color constancy image surfaces do not have a major impact on the illuminant adjustment, however, must remain preliminary. First, there is a huge number of possible variations in surface collections and the present study just investigates a few. Second, since the same surfaces were used on both sides of the display—as is usually done in studies on simultaneous color constancy—any effect of the surface collection common to both illuminants would not be seen in an asymmetric matching paradigm. In this sense, the present experiment just makes a start and further studies will be necessary to draw more firm conclusions on the role of image surfaces.

4.4. Surface matching: perceptual or conscious reasoning?

In general, appearance matching is a fairly easy task for subjects and they do not report any problems with it (but see Bäuml & Wandell, 1996, or Brainard, Brunt & Speigle 1997, for exceptions). At least for some subjects, surface matching seems to be more challenging, and from time to time a subject may even struggle with this task. The question that arises is whether surface matching is still a perceptual task, or whether it involves a substantial amount of conscious reasoning. In the present experiments there was no evidence that surface matching was largely based on reasoning. Most subjects found the surface matching task very natural and just seemed to perceive that a certain color was the 'right' one, in order for the matching field to represent the same surface as shown in the test field. This result is noticeable, since seven of the eight subjects who set surface matches in the two experiments were naive and had never before set any types of color matches. In previous studies many of the subjects were informed or so-called experienced subjects (Arend & Reeves, 1986; Cornelissen & Brenner, 1995).

Surface matching is usually examined in situations in which the simulated surfaces are the same in the left and right Mondrians, a feature which is also used for the subjects' instructions (see Section 3.1). But, is surface matching still perceptual, when the left and right Mondrians differ in their surfaces? This issue was addressed in informal observations by asking two fresh naive subjects to set surface matches under such conditions. In the left Mondrian, the standard collection rendered under the standard illuminant and in the right Mondrian the red collection rendered under the bright yellow or the dark blue test illuminant was presented. I used the same test surfaces as in Experiment 2. Both subjects found the task extremely difficult. They reported that they were searching for a reliable criterion to set the matches but failed to find one. The same subjects showed no problems in setting surface matches when the display's left and right Mondrians were identical. This observation may suggest that surface matching is no longer perceptual (if possible at all) in situations in which not only the illuminant but also the surfaces vary within the image. If this observation generalized to other subjects as well, we would need to think in more detail about the connection between the surface matching task and seeing in natural scenes. Indeed, while there will be common context across illuminant changes in some cases, in others there will not. Discovering exactly what stimulus conditions support surface matching and what conditions do not might become an important task for future research.

Acknowledgements

I thank D. Brainard, F. Cornelissen, D. Foster, K. Gegenfurtner, A. Reeves, and B. Wandell for their comments on a previous version of the paper and/or useful discussions. I also thank T. Engelschalk and S. Söllner for their help in the experimental work. Preliminary versions of this research were presented at the 1996 meeting of the Experimental Psychology Society in London, England, and at the 1996 conference Color vision: From Genes to Perception in Tübingen, Germany.

Appendix A. Experimental illuminants and surfaces

This Appendix consists of two tables. Table A1 specifies each of the five experimental illuminants. Table A2 specifies, separately for each of the three experimental surface collections, the 24 surfaces which surrounded the test surface and matching field within the visual display.

Table A1
Experimental illuminants

Illuminant	d_0	d_1	d_2	x	y	L
T_1	3.846×10^{-2}	1.113×10^{-1}	6.303×10^{-2}	0.249	0.249	29.76
T_2	6.538×10^{-2}	7.838×10^{-2}	-1.415×10^{-2}	0.274	0.282	48.48
S	8.077×10^{-2}	-4.986×10^{-2}	-3.500×10^{-2}	0.326	0.339	57.84
T_3	5.385×10^{-2}	-7.730×10^{-2}	1.199×10^{-1}	0.377	0.380	38.71
T_4	9.615×10^{-2}	-1.565×10^{-1}	4.328×10^{-1}	0.402	0.394	70.00

All five illuminants stem from the CIE daylight locus. They were constructed from a three-dimensional linear model of natural daylights (see Method section). This table gives the illuminants' weights for the three basis functions and the corresponding CIE x, y coordinates and luminance values.
S, standard illuminant; T_i, test illuminant i; d_k, weight for basis function k; L, luminance.

Table A2
Experimental surface collections

Surface	Standard			Bright			Red		
	x	y	L	x	y	L	x	y	L
(1, 1)	0.295	0.374	4.10	0.319	0.316	23.46	0.432	0.311	12.28
(1, 2)	0.340	0.267	4.24	0.282	0.336	26.27	0.425	0.377	11.06
(1, 3)	0.444	0.355	4.32	0.259	0.374	12.26	0.408	0.251	4.35
(1, 4)	0.308	0.265	11.29	0.351	0.321	20.06	0.385	0.326	2.33
(1, 5)	0.321	0.413	11.80	0.281	0.315	26.07	0.417	0.263	8.11
(2, 1)	0.311	0.267	2.04	0.337	0.369	24.75	0.454	0.271	4.34
(2, 2)	0.374	0.293	3.92	0.307	0.275	18.32	0.414	0.276	2.44
(2, 3)	0.344	0.298	12.40	0.306	0.352	27.31	0.456	0.294	7.91
(2, 4)	0.337	0.369	24.75	0.400	0.393	18.30	0.367	0.407	25.81
(2, 5)	0.337	0.204	3.79	0.270	0.320	25.58	0.443	0.387	8.26
(3, 1)	0.304	0.248	12.18	0.375	0.326	13.13	0.350	0.254	8.16
(3, 2)	0.330	0.471	18.81	0.367	0.407	25.81	0.466	0.365	12.06
(3, 4)	0.266	0.278	2.03	0.283	0.298	20.69	0.388	0.362	7.35
(3, 5)	0.260	0.229	11.60	0.240	0.285	13.43	0.457	0.332	7.27
(4, 1)	0.433	0.311	12.28	0.321	0.413	11.80	0.426	0.321	2.23
(4, 2)	0.319	0.316	23.76	0.326	0.339	17.55	0.407	0.342	2.66
(4, 3)	0.323	0.334	3.77	0.466	0.365	12.06	0.356	0.236	7.74
(4, 4)	0.375	0.326	13.12	0.294	0.387	18.24	0.415	0.333	7.20
(4, 5)	0.279	0.260	12.37	0.291	0.279	12.64	0.347	0.269	8.18
(5, 1)	0.390	0.280	3.88	0.270	0.320	25.58	0.424	0.300	4.16
(5, 2)	0.454	0.271	4.34	0.297	0.311	19.78	0.463	0.327	4.84
(5, 3)	0.262	0.296	12.33	0.338	0.295	19.12	0.392	0.364	4.07
(5, 4)	0.447	0.381	4.56	0.340	0.330	18.39	0.406	0.269	7.94
(5, 5)	0.270	0.320	25.58	0.283	0.298	20.69	0.419	0.308	7.40

The table shows, separately for each of the three surface collections, the CIE x, y chromaticity coordinates and the luminance values (L) of the 24 surfaces which surrounded the test surface and matching field within the visual display.
In Experiment 1 only the standard collection was used; in Experiment 2 the standard, the bright, and the red collection were used.
For each single surface, the first number specifies the row and the second number the column in which the surface was presented within the 5×5 array of experimental surfaces (compare Fig. 1).
The coordinates of the surfaces that result when they are rendered under the standard illuminant are shown.

References

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

Arend, L., & Goldstein, R. (1987). Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America A*, *4*, 2281–2285.

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

Bäuml, K. H. (1994). Color appearance: effects of illuminant changes under different surface collections. *Journal of the Optical Society of America A*, *11*, 531–543.

Bäuml, K. H. (1995). Illuminant changes under different surface collections: examining some principles of color appearance. *Journal of the Optical Society of America A*, *12*, 261–271.

Bäuml, K. H. (1997). Sensory and perceptual mechanisms of color constancy, Universität Regensburg. (Unpublished habilitation).

Bäuml, K. H., & Wandell, B. A. (1996). Color appearance of mixture gratings. *Vision Research*, *36*, 2849–2864.

Brainard, D. H. (1998). Color constancy in the nearly natural image. 2. Achromatic loci. *Journal of the Optical Society of America A*, *15*, 307–325.

Brainard, D. H., & Wandell, B. A. (1992). Asymmetric color-matching: how color appearance depends on the illuminant. *Journal of the Optical Society of America A*, *9*, 1433–1448.

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.

Chaparro, A., Stromeyer, C. F. III, Chen, G., & Kronauer, R. E. (1995). Human cones appear to adapt at low light levels: measurements on the red-green detection mechanism. *Vision Research*, *35*, 3103–3118.

Chichilnisky, E. J., & Wandell, B. A. (1995). Photoreceptor sensitivity changes explain color appearance shifts induced by large uniform backgrounds in dichoptic matching. *Vision Research*, *35*, 239–254.

Cornelissen, F. W., & Brenner, E. (1995). Simultaneous colour constancy revisited: an analysis of viewing strategies. *Vision Research*, *17*, 2431–2448.

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

Fairchild, M. D., & Berns, R. S. (1993). Image color-appearance specification through extension of CIELAB. *Color Research & Application*, *18*, 178–190.

Fairchild, M. D., & Reniff, L. (1995). Time course of chromatic adaptation for color-appearance judgments. *Journal of the Optical Society of America A*, *12*, 824–833.

Finlayson, G. D., Drew, M. S., & Funt, B. V. (1994). Spectral sharpening: sensor transformations for improved color constancy. *Journal of the Optical Society of America A*, *11*, 1553–1563.

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*, *121*, 115–121.

Gegenfurtner, K. R. (1992). PRAXIS: Brent's algorithm for function minimization. *Behavior Research Methods, Instruments Computers*, *24*, 560–564.

Hunt, R. W. G. (1950). The effects of daylight and tugsten-light adaptation on color perception. *Journal of the Optical Society of America*, *40*, 362–371.

Irtel, H. (1997). PXL: a library for psychological experiments on IBM PC type computers. *Spatial Vision*, *10*, 467–469.

Jameson, D., & Hurvich, L. M. (1955). Some quantitative aspects of an opponent colors theory. III. Changes in brightness, saturation, and hue with chromatic adaptation. *Journal of the Optical Society of America*, *45*, 546–552.

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

Kaiser, P. K., & Boynton, R. M. (1996). *Human color vision*, 2nd. Washington: Optical Society of America.

Kelly, K. L., Gibson, K. S., & Nickerson, D. (1943). Tristimulus specification of the munsell book of color from spectrophotometric measurements. *Journal of the Optical Society of America*, *33*, 355–376.

Kries von, J. (1905). Die Gesichtsempfindungen. In W. Nagel, *Handbuch der physiologie des menschen*, Vol. 3. Braunschweig: Vieweg, 109–279.

Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America A*, *3*, 1673–1683.

Nascimento, S. M. C., & Foster, D. H. (1998). Detecting natural changes in cone-excitation ratios in simple and complex coloured images, *Proceedings of the Royal Society of London Series B* (in press).

Poirson, A. B., & Wandell, B. A. (1993). The appearance of colored patterns: pattern-color separability. *Journal of the Optical Society of America A*, *10*, 2458–2470.

Shevell, S. K. (1978). The dual role of chromatic backgrounds in color perception. *Vision Research*, *18*, 1649–1661.

Smith, V., & Pokorny, J. (1975). Spectral sensitivity of the foveal cone photopigments between 400 and 700 nm. *Vision Research*, *15*, 161–171.

Troost, J. M., & deWeert, C. M. (1991). Naming versus matching in color constancy. *Perception & Psychophysics*, *50*, 591–601.

Walraven, J. (1976). Discounting the background-the missing link in the explanation of chromatic adaptation. *Vision Research*, *16*, 289–295.

Wandell, B. A. (1995). *Foundations of vision*. Sunderland: Sinauer.

Werner, J., & Walraven, J. (1982). Effect of chromatic adaptation on the achromatic locus: the role of contrast, luminance and background color. *Vision Research*, *22*, 929–943.

Wyszecki, G., & Stiles, W. S. (1982). *Color science*, 2nd. New York: Wiley.

Zeki, S. (1993). *A vision of the brain*. Cambridge: Blackwell.