

Color appearance: effects of illuminant changes under different surface collections

Karl-Heinz Bäuml

Institut für Psychologie, Universität Regensburg, Regensburg, Germany

Received March 29, 1993; revised manuscript received July 21, 1993; accepted August 5, 1993

A theory about how changes in the illuminant affect the color appearance of objects must specify how the visual system's adjustments to illuminant changes vary with the surface collection in a scene. I report an experiment designed to investigate this issue. The stimuli were CRT simulations of flat matte surfaces rendered under diffuse illumination. For all combinations of 7 daylight illuminants and 12 collections of surface reflectances the subject's achromatic locus was measured on an isoluminant plane in color space. For any surface collection the changes in the achromatic locus could be well approximated by a linear transformation of the illuminant changes. These linear transformations showed relatively small variation with the surface collection. To first approximation, these results suggest that the effect of changes in the illuminant on color appearance can be described linearly and that it can be separated from the surface collection. There was an effect of the surface collection on the achromatic locus. The data rejected the idea that a surface collection's mean reflectance function might capture this effect, ruling out models of color appearance that are based on this kind of averaging assumption.

1. INTRODUCTION

The light that is reflected from an illuminated object is the wavelength-by-wavelength product of illuminant and the object's surface reflectance function. Through this relationship, changes in the illuminant can have a large effect on the reflected light. In fact, the spectral power distribution of natural illuminants shows strong variation both within a day and across days,¹⁻³ thus leading to considerable changes in the reflected light's spectral composition. Despite this variation in the light that reaches our eyes the color appearance of objects varies only moderately with changes in the illuminant. This compensating behavior of our visual system, often referred to as color constancy, has been subject to many experimental studies examining how color appearance depends on the illuminant.⁴⁻⁸

Usually, the visual system's compensation for illuminant changes is examined as a function of changes in the illuminant, ignoring the possible role of a scene's surface collection in this process. *A priori*, however, our system's compensation for illuminant changes may also depend on the surface reflectances in a scene. The way that illuminant changes affect an object's color appearance may therefore vary with the underlying surface collection. The present study addresses this issue. It reports the results from a set of experiments designed to study how the illuminant-induced changes in an object's color appearance vary with the surface collection in a scene. The stimuli were CRT simulations of flat objects rendered under diffuse illumination. For several combinations of illuminant and surface collection the subject's achromatic locus was measured on an isoluminant plane in color space. These measurements were used to quantify appearance changes.

Measuring the achromatic locus under the different viewing contexts establishes equivalent chromatic appear-

ances across contexts. Variations in a subject's locus as a function of changes in the viewing context therefore reflect adjustments of the visual system in response to the context changes. In this sense these loci provide empirical measurements for a theory of color appearance that describes how the visual system adjusts to context changes. While the achromatic locus represents only a first test object for studying the role of viewing context, this kind of measurement permits us to avoid the use of more complex experimental tasks such as memory matching.^{7,9,10} In terms of Hering's opponent-colors theory the achromatic locus equals a color that appears neither reddish nor greenish and neither yellowish nor bluish.^{11,12} When using these absolute perceptual criteria, subjects are able to set the chromatic appearance of a test object reliably,¹³⁻¹⁸ and, as a result, no prior learning of the object's appearance is necessary.

The achromatic objects measured in the present study varied as a function of changes in the illuminant and of changes in the surface collection, thus reflecting adjustments of the visual system to both variables. In order to avoid the direct measurement of the effect of these variables for every combination of illuminant and surface collection, one must identify some theoretical principles to describe how the visual system's adjustments vary with illuminant and surface collection. Two kinds of principle are necessary. First a theory of color appearance must specify the functional form of the transformations that relate the observed changes in the objects, on the one hand, and changes in illuminant, or surface collection, on the other hand.¹⁹ Second, because *a priori* the effects of the two variables on color appearance may interact with each other, a theory of color appearance must specify how the transformation that describes the effect of illuminant changes varies with the surface collection and how the transformation that describes the effect of changes in the surface collection varies with the illuminant. Only in the

most simple case, in which the changes in color appearance that are induced by illuminant changes do not vary with the surface collection and those induced by changes in the surface collection do not vary with the illuminant, does this need of specification become redundant, thus providing a powerful invariance principle in a theory of color appearance.

I started out by examining the data for regularities in the functional form of the mapping between changes in the illuminant and the observed changes in the achromatic object. For any surface collection I found this mapping to be well approximated by a linear transformation. In this sense the data are consistent with Brainard and Wandell's⁷ results concerning illuminant change linearity. I went on to examine to what extent these linear transformations varied with the surface collection in a scene. I found relatively small dependency of the linear transformations on the surface collection. To first approximation, these results suggest that the effect of changes in the illuminant on an object's color appearance can be described linearly and that it can be separated from the surface collection. Finally, I examined whether the effect of the surface collection might be captured by a surface collection's mean reflectance function. The data rejected this idea.

2. METHOD

The methods used in this study were similar to those used by Brainard and Wandell.⁷ The visual stimuli were presented on a CRT monitor. The stimulus was a simulation of an array of flat matte objects rendered under spatially uniform illumination and a test object. This array of illuminated objects represented one of the simulated illuminants and one of the simulated surface collections. All the surface collections consisted of 24 surface reflectances. The stimulus was presented against a large dark background that consisted of two different surface reflectances rendered under the same illuminant as the surface collection. The subjects pressed buttons to adjust the appearance of the test object until it appeared achromatic to him or her. Both the simulated illuminant and the simulated surface collection were fixed during an adjustment process. However, during this process the local positions of a quarter of the stimuli were changed every second, including the test object. This design was chosen to minimize local effects like local adaptation or simultaneous contrast⁸ and to isolate the effect of changing the illuminant from other variables. The randomization was also intended to prevent subjects from using the appearance of a fixed surface as a reference to identify changes in the simulated illuminant.²⁰

A. Visual Display

Figure 1 shows the visual display. It consisted of 25 small foreground regions against a large partitioned background region. The foreground regions consisted of 24 rectangular regions (simulation of illuminated surfaces) and one oval region (test object). The background region consisted of three large rectangular regions. The foreground regions subtended 2.6 vertical by 2.1 horizontal degrees of visual angle, and they were separated by 0.3 vertical and 0.2 horizontal degrees. The whole background region sub-

tended 15.2 vertical by 19.8 horizontal degrees of visual angle. The left-hand and right-hand rectangular regions of the background subtended 15.2 vertical by 2.2 horizontal degrees of visual angle, and the middle rectangular region subtended 15.2 vertical by 15.4 horizontal degrees. Subjects saw the screen without head restraints from a distance of approximately 1 m in an otherwise dark room.

The simulated images were displayed on a computer-controlled color monitor (BARCO Calibrator), using a refresh rate of 60 Hz in noninterlaced 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 of each pixel. There was a software control of the monitor's input signal correcting local variations of the monitor's energy output and nonlinearities in the tube's response function. The luminance of each color channel was measured with a high-precision photometer (Fa. Lichtmesstechnik, Model L1003). The CIE *xy* coordinates of the phosphors were provided by the manufacturer.

To simulate a given surface under a given illuminant, the surface reflectance function and the illuminant spectral power function were multiplied. This product is the spectral power distribution of the reflected light. The CIE *XYZ* tristimulus coordinates of this reflected light were computed, and the values in the monitor frame buffer were appropriately set. The simulation computations were performed prior to the experimental sessions.

B. Illuminants

I used seven experimental illuminants (Fig. 2), which I selected from the CIE daylight locus.²¹ They were typical for natural daylight. All the experimental illuminants were constructed from the three-dimensional linear model of natural daylight proposed by Judd *et al.*³ In fact, the spectra of natural illuminants can be well approximated by an appropriate linear combination of three basis functions that represent the calculated mean spectral distribution and the first two characteristic functions of a large number of measured daylight distributions. The experi-

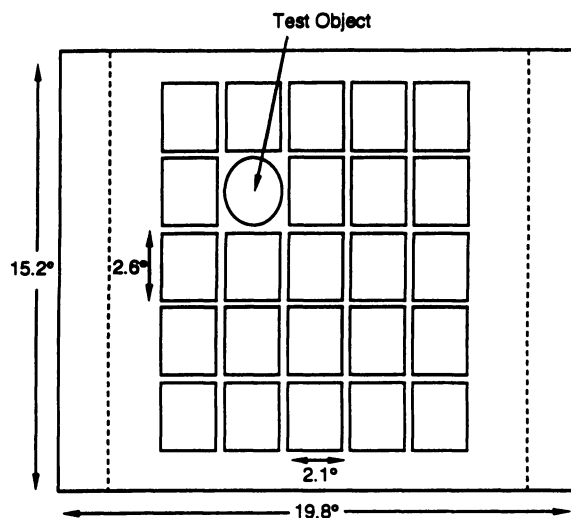


Fig. 1. Visual display. Subjects saw CRT simulations of a collection of 24 flat matte surfaces rendered under a spatially uniform illuminant (rectangular regions) and a test object (oval region). A detailed description of the stimulus is given in Section 2.

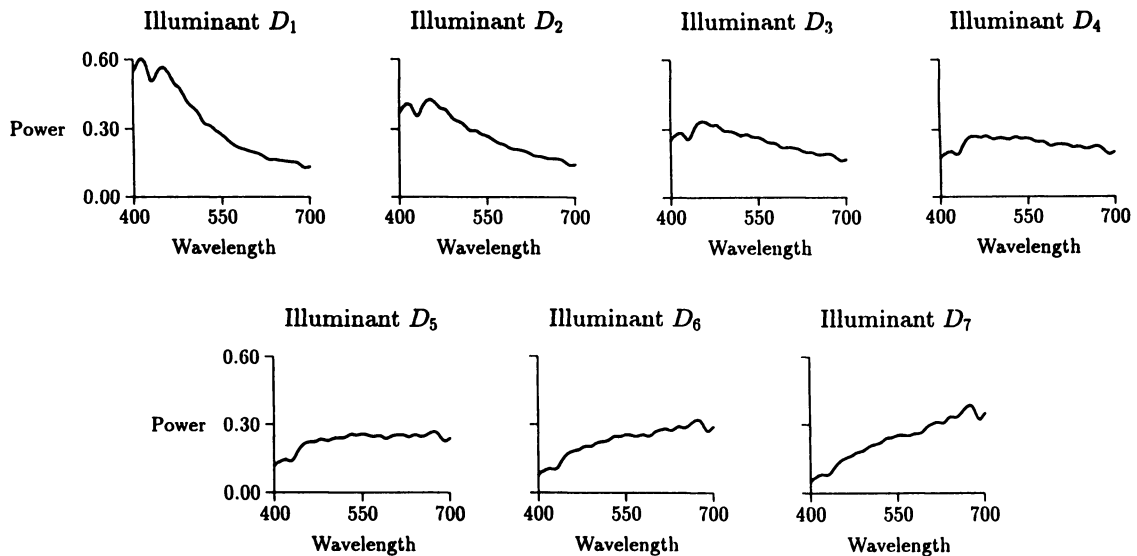


Fig. 2. Experimental illuminants. Each plot shows the spectral power distribution of one experimental illuminant. The standard illuminant in all our conditions was illuminant D_4 . The six other illuminants were used as test illuminants.

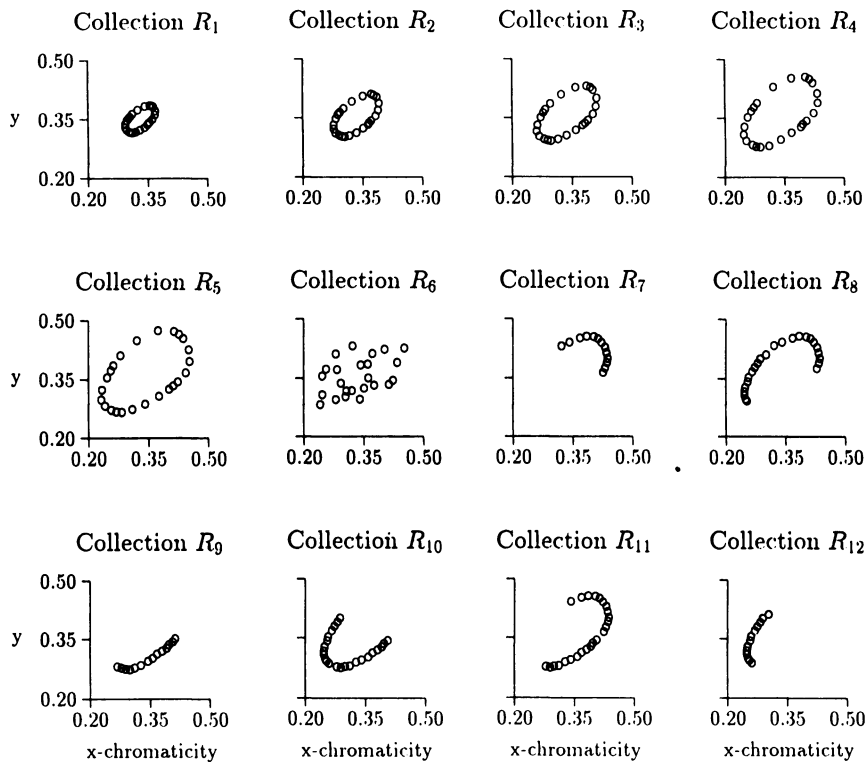


Fig. 3. Experimental surface reflectance collections. Each plot shows for one of the 12 experimental surface collections the CIE xy coordinates of its single surface reflectances when the collections were rendered under the standard illuminant D_4 (compare with Fig. 2). All the surface collections consisted of 24 different surface reflectances, except for collections R_7 , R_9 , and R_{12} , which consisted of 13 or 14 different surfaces.

mental illuminants shared a constant coordinate with respect to the model's first basis function and varied in the coordinates for the second and third basis functions. The spectral properties of the illuminants are provided in Appendix A.

C. Surface Collections

Twelve experimental collections of surface reflectances were constructed, each consisting of twenty-four surface reflectances. All the surface reflectances were approximations of Munsell chips. It is well known that the spec-

tra of Munsell chips can be well approximated by appropriate linear combination of six basis surface reflectance functions.^{22,23} Moreover, when human photoreceptors are taken into account, the first three or four basis functions of the linear model are already sufficient to fit the chips' reflectances closely.²³ I approximated the chips' reflectance functions with a three-dimensional linear model, where the three basis functions represent the first three principal components of the entire data set of Kelly *et al.*²⁴

Figure 3 shows the gamut of all surfaces from all 12 surface collections when they are rendered under experi-

mental illuminant D_4 (compare with Fig. 2). Rendered under this illuminant, the collections R_1 – R_6 spanned the whole gamut of hues and differed only with respect to their surfaces' saturation level(s). The collections R_7 – R_{12} , instead, constituted only a more or less small region of the hue gamut when they were rendered under this illuminant and were roughly constant with respect to their surfaces' saturation. The mean reflectance functions of the surface collections R_1 – R_6 were close to spectrally flat reflectance functions and were approximately identical across collections. The mean reflectance functions of the collections R_7 – R_{12} spanned a considerable range, from a mean reflectance function for which mainly light from the long-wavelength part of the spectrum is reflected (R_7) to a mean reflectance function for which mainly light from the short-wavelength part is reflected (R_{12}). Appendix A provides the spectral properties of the surface collections' mean reflectance functions. When they were rendered under the simulated illuminants, the luminance of the individual illuminated surfaces was roughly constant (50 cd/m^2), both within and across surface collections. The range of variation was approximately 7% in each direction of the luminance scale.

D. Background Surface Collections

For all the simulated collections of surface reflectances the same background surfaces were used. These surfaces were again three-dimensional approximations of Munsell chips. Eight pairs of dark surfaces were used, for which a pair's two surfaces approximated a flat mean reflectance function when they were spectrally superimposed. From this pool of surfaces two distinct surfaces were randomly drawn each second, one surface for both the left-hand and right-hand parts of the background region and the other surface for the middle part of the background region (compare with Fig. 1). This design was chosen to prevent subjects from using the appearance of a fixed background as a reference to identify changes in the simulated illuminant. When they were rendered under one of the experimental illuminants, the luminances of these surfaces were approximately 2 cd/m^2 .

E. Subjects

Two subjects took part in this study. They had normal color vision. They were partly informed of the goals of the experiment.

F. Procedure

In each experimental session three combinations of surface collection and illuminant were presented to a subject. For each combination a subject made four settings during a session. Each combination of surface collection and illuminant was presented in two sessions to a subject, resulting in 8 settings for each of the 84 combinations (12 surface collections \times 7 illuminants). A subject began an experimental session with 5 min of dark adaptation. Then a collection of simulated surfaces rendered under a simulated illuminant was presented to the subject. The test object was not yet visible. After 4 min of adaptation to these illuminated surfaces the test object appeared, and the subjects made four settings for this condition, readapting between the settings for 30 s. After these four settings the next combination of surface collection and

illuminant was presented. Exactly the same procedure was used as that for the first combination. The subjects made adjustments by pressing the buttons of a mouse. They were instructed to adjust a test object that appeared neither reddish nor greenish and neither yellowish nor bluish, i.e., achromatic to them. In dependency of these button presses they moved the test object on an isoluminant plane in color space of approximately 50 cd/m^2 . Each adjustment that a subject made was recorded in CIE XYZ coordinates.

G. Data Analysis

I report tests of linear models, that is, models of the form $\mathbf{t} = \mathbf{Q}\mathbf{c}$, where \mathbf{c} is a vector representing a change in the viewing context (either a change in the linear coordinates of the simulated illuminants or a change in the linear coordinates of the simulated surface collection's mean reflectance functions), \mathbf{t} is a vector representing the difference between the subjects' settings under two different viewing contexts, and \mathbf{Q} is a matrix that maps \mathbf{c} into \mathbf{t} . For each model transformation \mathbf{Q} , I require an error measure to choose a best-fitting transformation. I evaluate the size of difference between observed and predicted settings relative to an estimate of the objects' covariances.

For each achromatic object there are eight repeated measurements adjusted by a subject. Based on these replications I estimate for each object i its covariance matrix Δ_i . I minimize the differences between the observed and predicted settings by minimizing the error measure

$$\sum_i \sum_j \mathbf{e}_{ij}^t \Delta_i^{-1} \mathbf{e}_{ij},$$

where \mathbf{e}_{ij} denotes the difference between the observed and predicted settings for object i under replication j . This error measure is equivalent to transforming the model deviations into a coordinate frame in which the errors from the three coordinates are independent and have unit variance and to measuring the Euclidean distance in that frame. I used the iterative search procedure PRAXIS²⁵ to perform the error minimizations.

In the present context the use of this error measure includes the statistical problem that, in general, eight replications will not provide a very exact estimation of a covariance. I also evaluated the models by using other error measures. Specifically, I minimized the differences between the observed and predicted settings by using one single global covariance matrix estimated for all the objects simultaneously and by using the CIE LUV metric space.²¹ The conclusions drawn about the quality of the models did not depend on the choice of error measure.

3. RESULTS

Figure 4 shows the effect of illuminant changes on the two subjects' achromatic locus. The effect is shown for all 12 experimental surface collections using CIE xy chromaticity coordinates. The 7 diamonds in each of the 12 plots represent the xy coordinates of the seven experimental illuminants. The big circle in each plot represents the xy coordinates of the respective surface collection's mean reflectance function when it is rendered under

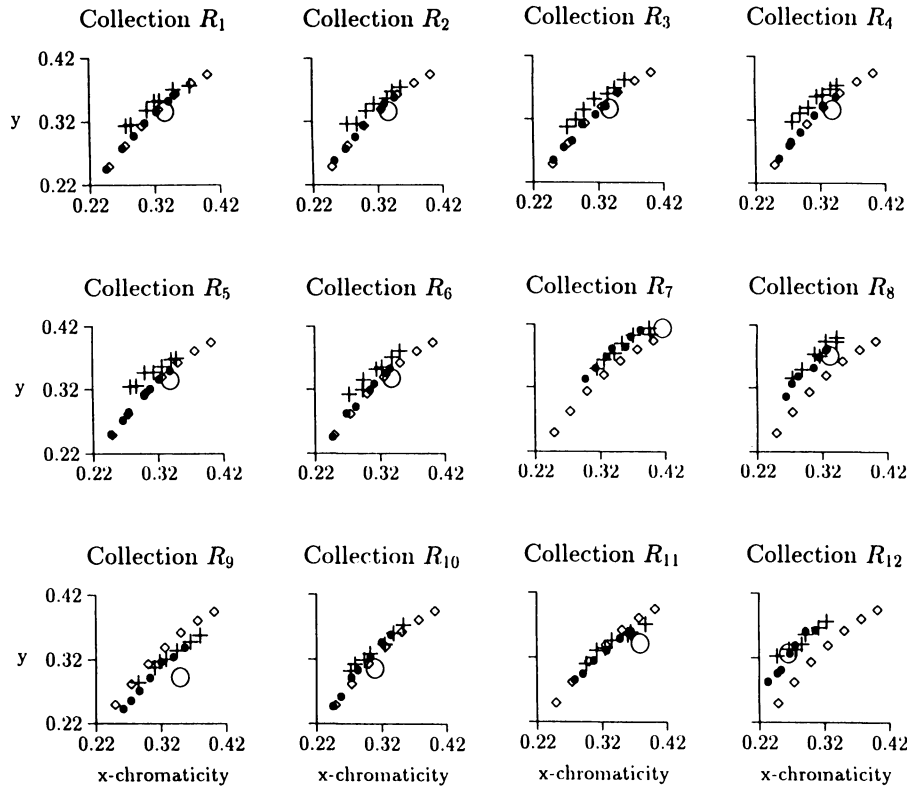


Fig. 4. Test objects adjusted by the two subjects. Each plot shows the effect of illuminant changes on the test object for one experimental surface collection. Each plot shows the seven test objects' mean CIE xy coordinates [subject MP (+), subject AH (●)], the xy coordinates of the seven experimental illuminants (◇), and the xy coordinate of the respective experimental surface collection's mean reflectance function when it is rendered under spectrally flat illumination (○).

a spectrally flat illuminant. The change of location of this circle as a function of surface collection reflects the changes in the mean reflectance function across surface collections.

Each plot also shows the mean xy coordinates of the seven achromatic loci adjusted by the two subjects under the seven illuminants. In general, the two subjects' data show a regular pattern in that, loosely speaking, their achromatic points shift from more bluish locations in the diagram to more yellowish locations as the illuminant changes from bluish to yellowish locations. This pattern holds for both subjects, although their measurements in absolute coordinates differ remarkably. This effect of changes in illuminant demonstrates the well-known tendency of the human visual system to color-constant performance for the present experimental paradigm.

The figure also reveals an effect of the surface collection on the achromatic locus. The locations of the achromatic points observed under the single surface collections vary considerably as a function of surface collection. This effect shows up, in particular, for pairs of surface collections with major differences in their mean reflectance functions, while for pairs of surface collections with approximately identical mean reflectance functions (R_1 – R_6) the variation appears fairly small. As a result, there is an effect of both illuminant and surface collection on the achromatic locus.

A. Illuminant Changes

I held one of the illuminants (D_4) fixed. I refer to it as the standard illuminant in the following. Under each of

the 12 experimental surface collections I also held the mean test object fixed, which the subject adjusted as achromatic under the standard illuminant. For each surface collection I refer to it as the standard object in the following. Illuminants were represented by the degree to which their coordinates differed from those of the standard illuminant; as coordinates I used the illuminant weights with respect to the three basis functions of the Judd *et al.*³ model (see Table 2 in Appendix A, below). These differences are called illuminant changes. Test objects were represented by the degree to which their XYZ coordinates differed from the XYZ coordinates of the respective standard object. These differences are called changes in the test objects. Because the test objects were equiluminant (see Section 2), they differed only in their X and Z coordinates, sharing a constant Y coordinate (see Section 1). As a result, I describe the test objects as two-dimensional vectors in the following.

I analyzed the data by examining three issues. First, to what extent can illuminant changes, on the one hand, and the illuminant-induced changes in the test objects, on the other hand, be related by a linear transformation? Brainard and Wandell⁷ provided empirical evidence that the functional form of this mapping might be well approximated by a linear transformation, referred to as illuminant change linearity in the following. Second, based on illuminant change linearity, to what extent do the single linear transformations vary as a function of the collection of surface reflectances? Third, to what extent does the description of the data improve when a more general affine linear transformation is used to relate illuminant

changes and changes in the test objects?²⁶ I refer to this mapping as affine illuminant change linearity in the following.

I represent the illuminants D_i by vectors \mathbf{d}_i and represent the achromatic object adjusted under illuminant D_i and surface collection R_j by \mathbf{t}_{ij} ($i = 1, \dots, 7; j = 1, \dots, 12$). Illuminant change linearity then assumes that there exists a 2×2 matrix \mathbf{M}_j for each of the 12 surface collections R_j that transforms the illuminant changes $\Delta \mathbf{d}_i$ ($= \mathbf{d}_i - \mathbf{d}_4$) into the test object changes $\Delta \mathbf{t}_{ij}$ ($= \mathbf{t}_{ij} - \mathbf{t}_{4j}$):

$$\Delta \mathbf{t}_{ij} = \mathbf{M}_j \Delta \mathbf{d}_i.$$

Because the 12 matrices \mathbf{M}_j are free to vary as a function of surface collection, this linear model requires 48 (12×4) parameters in order to predict the 144 ($12 \times 6 \times 2$) measured coordinates. Recall that the standard objects of the single surface collections were held fixed and thus were not subject to prediction. More generally, affine illuminant change linearity assumes that, for each of the 12 surface collections, there exists a 2×2 matrix \mathbf{M}_j and a 2×1 vector \mathbf{n}_j that transform the illuminant changes $\Delta \mathbf{d}_i$ into the test object changes $\Delta \mathbf{t}_{ij}$:

$$\Delta \mathbf{t}_{ij} = \mathbf{M}_j \Delta \mathbf{d}_i + \mathbf{n}_j.$$

Because both the 12 matrices \mathbf{M}_j and the 12 vectors \mathbf{n}_j are free to vary as a function of surface collection, this affine linear model requires 72 (12×6) parameters to predict the 144 measured coordinates.

Were the visual system's adjustments to illuminant changes linear and not varying with the surface collection, the 12 matrices \mathbf{M}_j fitted for the 12 surface collections should not differ reliably from one another; i.e., the restriction

$$\mathbf{M}_1 = \mathbf{M}_2 = \dots = \mathbf{M}_{12}$$

should hold. As a result, the data should be equally well describable by just one 2×2 matrix \mathbf{M} that transformed the illuminant changes into the test object changes simultaneously for all the surface collections. This more restrictive linear model requires only four parameters to predict the 144 measured coordinates. Similarly, were the visual system's adjustments to illuminant changes affine linear and not varying with the surface collection, both the 12 matrices \mathbf{M}_j and the 12 vectors \mathbf{n}_j fitted for the 12 surface collections should not differ reliably from one another; i.e., the restrictions

$$\mathbf{M}_1 = \mathbf{M}_2 = \dots = \mathbf{M}_{12}$$

$$\mathbf{n}_1 = \mathbf{n}_2 = \dots = \mathbf{n}_{12}$$

should hold. This more restrictive affine linear model requires six parameters to predict the measured coordinates.

The fit of the models was compared with the subjects' precision and a no-model prediction representing the effect to be explained. I evaluated the fit of the models by using the differences between observed and predicted test objects as error measures. To evaluate the subjects' precision, I used as error measures the variation of the repeated settings. Finally, to evaluate the size of the effect, I used the differences of the single settings from the respective standard objects as error measures. In all these

cases I used the objects' covariance matrices for the analysis (see Section 2).

Figure 5 shows the results. The first and sixth bars of each figure reveal the effect of illuminant changes on color appearance in the present experiment. Changing the illuminant had a considerable effect on the test objects, demonstrating the visual system's adjustments to illuminant changes. The second and third bars show the quality of fit of the (weak) affine linear and the (weak) linear model, respectively, predicting the changes in the test ob-

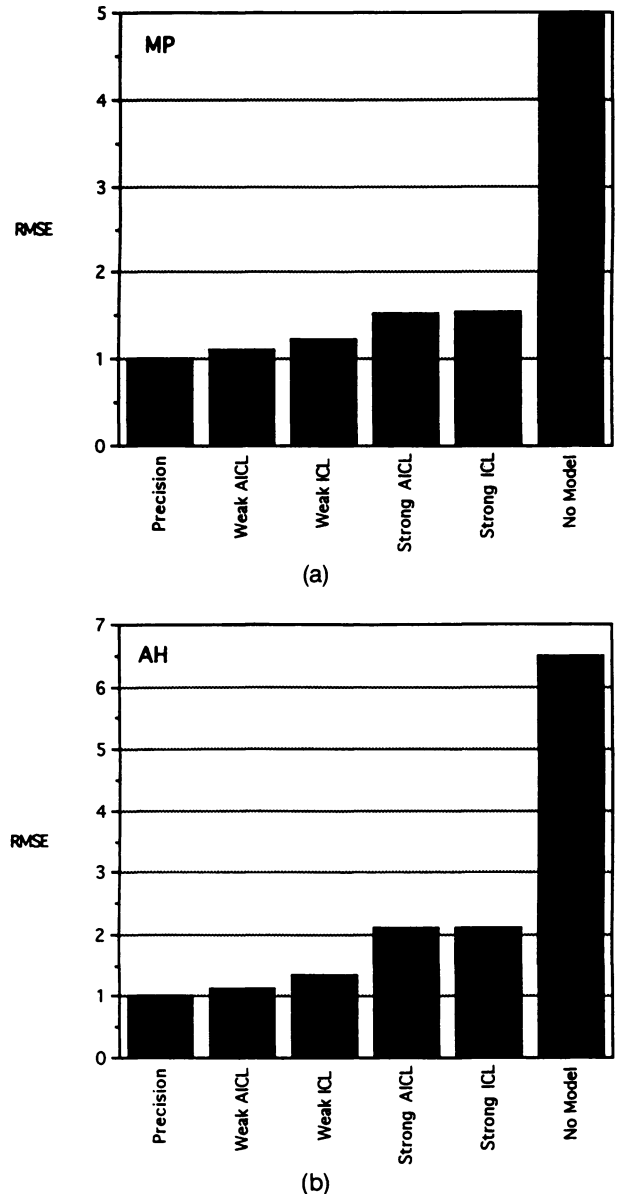


Fig. 5. Illuminant changes: quality of model fit. Comparison among a subject's precision, the size of the effect (no model), and the quality of fit for affine illuminant change linearity (AICL) and illuminant change linearity (ICL). The 144 measured coordinates were described for the weak affine linear model (Weak AICL) by 12 matrices and 12 vectors with 72 parameters, for the weak linear model (Weak ICL) by 12 matrices with 48 parameters, for the strong affine linear model (Strong AICL) by only 1 matrix and 1 vector with 6 parameters, and for the strong linear model (Strong ICL) by only 1 matrix with 4 parameters. The error measurements were root-mean-square errors (RMSE).

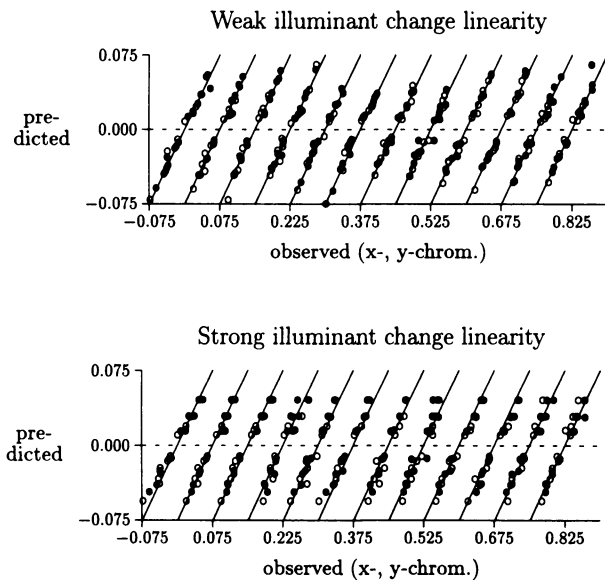


Fig. 6. Illuminant changes: scatter plot comparing the observed mean changes in the test objects with the models' predictions of these changes. The quality of fit of the two linear models (Weak ICL and Strong ICL) is shown with CIE xy coordinates. The 12 data sets in each plot relate to the 12 experimental surface collections (R_1 – R_{12}). The sets are laterally shifted by 0.075, 0.150, 0.225, ... units relative to the origin of the coordinate system. The two subjects' data are merged. If linearity held perfectly, all 24 coordinate points for each surface collection would fall on a diagonal line [x chromaticity (\bullet), y chromaticity (\circ)].

jects by means of 12 individual affine linear and linear transformations, respectively. The errors resulting from the models' predictions were close to the subjects' precision, reflecting a good fit to the data. The fourth and fifth bars show the quality of fit of the (strong) affine linear and the (strong) linear model, respectively, predicting the changes in the test objects by means of only one affine linear and linear transformation, respectively. A comparison of quality of fit between the two models' weak and strong forms shows to what extent the description of the data deteriorates when the same transformation for all the surface collections is used instead of the 12 individual transformations that are permitted to vary with the surface collection. The models' weak and strong forms are nested, and thus the respective strong form fits the data worse than the respective weak form. The deterioration in fit, however, was fairly moderate. A comparison of quality of fit between the affine linear and the linear model reveals to what extent the fit to data improves when one allows for an additional affine term in the linear transformation. Because the affine linear and the linear models are nested, the affine linear model fits the data better. The fit to data, however, did not improve in a substantial way. As a result, the affine linear and the linear models compare well in their quality of fit, with only moderate differences between the models' weak and strong forms.²⁷

For each of the 12 surface collections Fig. 6 compares the observed mean changes in the test objects with the predictions of these changes for the two linear models. The two subjects' data are merged. If linearity held perfectly, all 24 coordinate points for each surface collection (6 illuminant changes \times 2 color coordinates \times 2 subjects)

would fall on a diagonal line. The weak linear model fits the data well. The strong linear model is worse, though it still fits the data in a reasonable way. This pattern of results reflects the conclusions already drawn from Fig. 5.

B. Changes in the Surface Collection

Figure 4 showed an effect of a given surface collection on the achromatic locus. A theoretically attractive way to account for these variations is the idea of capturing the effect of surface collection by the collections' mean reflectance functions. This idea assumes a well-defined function that maps changes in the mean reflectance function into changes in the test objects. The empirical condition for this assumption to hold is that surface collections with an equal mean reflectance function induce equal test objects when the collections are rendered under the same illuminant.

The present experiment permits one to examine the empirical soundness of this condition in some detail. For each of the seven experimental illuminants I examined to what extent the approximately identical mean reflectance functions of the six surface collections R_1 – R_6 resulted in the same locations for the achromatic points. For one of the two subjects (MP) Fig. 7 shows for each of the seven illuminants the mean xy locations of the achromatic points observed under these six collections, together with their 95% uncertainty ellipsoids. The computation of the ellipsoids is based on the assumption that repeated measurements distribute normally about a center.^{21,28}

The achromatic points show some variation as a function of surface collection. I used a multivariate extension of the F -ratio test in simple analysis of variance^{29,30} to test the significance in the differences among the six achromatic points. Table 1 shows the F values and the corresponding p values for the two subjects. For subject MP, for four of the seven illuminants the achromatic points differ reliably. For subject AH, for six of the seven experimental illuminants the achromatic points differ reliably from one another. In all these cases the six surface collections did not induce the same test objects, contrary to the idea that the mean reflectance function might capture the effect of surface collection on color appearance.

The assumption that the effect of surface collection can be completely captured by the collections' mean reflectance functions is highly restrictive in nature and correspondingly quite easy to reject statistically. Although it is at best a simplification, this assumption might nonetheless be useful for practical purposes. To address this point, I examined to what extent the changes in color appearance that are induced by changes in the surface collections can be approximated by a linear or affine linear function of changes in the collections' mean reflectance functions.

I held the mean reflectance function of surface collection R_1 fixed. I refer to it as the standard mean reflectance function in the following. Under each of the seven experimental illuminants the mean test object that the subject adjusted as achromatic under surface collection R_1 was also held fixed. For each illuminant it is referred to as the standard object in the following. The mean reflectance functions of the single surface collections were represented by the extent to which their coordinates differed from those of the standard mean reflectance function; as

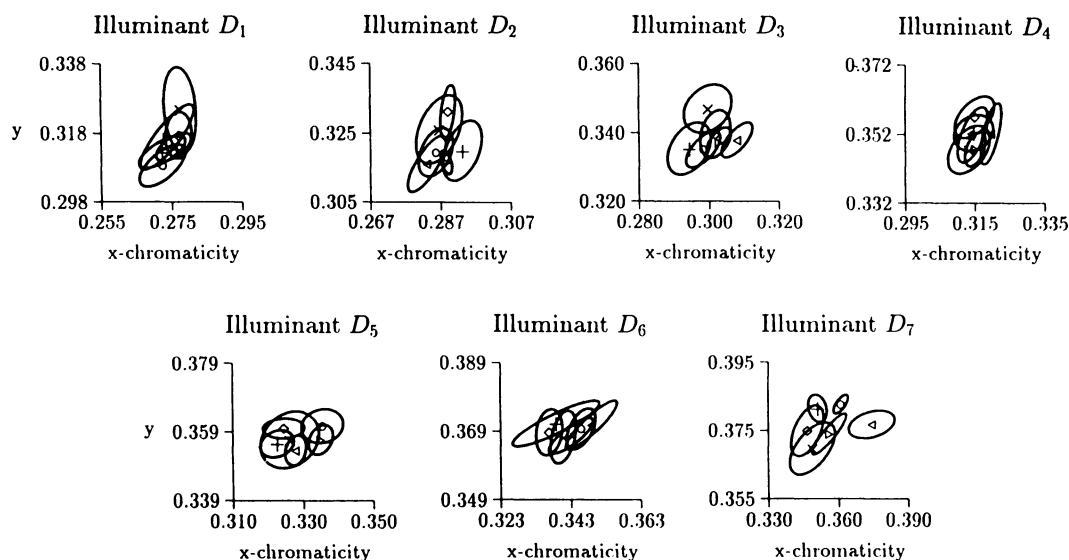


Fig. 7. Changes in the surface collection: role of a collection's mean reflectance function. Each plot shows for one of the seven experimental illuminants the CIE xy locations of the mean test objects measured under the experimental surface collections R_1 – R_6 . These six collections' mean reflectance functions are approximately identical (see Section 2). Each plot also shows the 95% uncertainty ellipsoids of these objects. The computation of the ellipsoids is based on the assumption that repeated measurements distribute normally about a center. The data are shown for subject MP [R_1 (\triangleleft), R_2 (\triangleright), R_3 (\circ), R_4 (\diamond), R_5 (\times), R_6 ($+$)].

Table 1. Mean Reflectance Function as an Indicator of the Effect of Surface Collection on Color Appearance^a

	Subject MP	Subject AH
D_1	1.356 ($p = 0.21496$)	2.770* ($p = 0.00527$)
D_2	3.186* ($p = 0.00165$)	1.790 ($p = 0.07494$)
D_3	3.758* ($p = 0.00033$)	3.738* ($p = 0.00035$)
D_4	1.377 ($p = 0.20503$)	4.699* ($p = 0.00003$)
D_5	3.020* ($p = 0.00262$)	9.574* ($p < 0.00001$)
D_6	1.710 ($p = 0.09184$)	7.223* ($p < 0.00001$)
D_7	7.491* ($p < 0.00001$)	3.078* ($p = 0.00223$)

^aThis table shows for each of the seven experimental illuminants the results from F tests conducted to test whether the CIE xy locations of the mean test objects are identical to the precision specified by the sample ellipsoids. The sample ellipsoids examined were estimated from the test objects adjusted under the surface collections R_1 – R_6 (see Section 3). Asterisks denote significant values.

coordinates I used the reflectance functions' weights with respect to the three basis functions of the data of Kelly *et al.*²⁴ (see Table 3 in Appendix A, below). These differences are called changes in the mean reflectance function. Test surfaces were represented by the extent to which their XYZ coordinates differed from those of the respective standard objects. These differences are called changes in the test objects.

I analyzed to what extent changes in the mean reflectance functions, on the one hand, and the collection-induced changes in the test objects, on the other hand, could be related by a linear or affine linear transformation. I refer to these properties as reflectance change linearity or affine reflectance change linearity, respectively, in the following. I represent the mean reflectance functions of the collections R_j by vectors \mathbf{r}_j and represent the achromatic locus adjusted under collection R_j and illuminant D_i by \mathbf{t}_{ji} ($j = 1, \dots, 12; i = 1, \dots, 7$). Reflectance change linearity then assumes that there exists a 2×3 matrix \mathbf{P}_i for each of the seven illuminants D_i that transforms the changes in the mean reflectance functions $\Delta \mathbf{r}_j$

($=\mathbf{r}_j - \mathbf{r}_1$) into the test object changes $\Delta \mathbf{t}_{ji}$ ($=\mathbf{t}_{ji} - \mathbf{t}_{1i}$):

$$\Delta \mathbf{t}_{ji} = \mathbf{P}_i \Delta \mathbf{r}_j.$$

Because the seven matrices \mathbf{P}_i are free to vary as a function of illuminant, this linear model requires 42 (7×6) parameters for prediction of the 154 ($11 \times 7 \times 2$) measured coordinates. Recall that the standard objects of the single illuminants were held fixed and thus were not subject to prediction. Affine reflectance change linearity assumes that, for each of the seven illuminants D_i , there exist a 2×3 matrix \mathbf{P}_i and a 2×1 vector \mathbf{q}_i that transforms the changes in the mean reflectance functions $\Delta \mathbf{r}_j$ into the test object changes $\Delta \mathbf{t}_{ji}$:

$$\Delta \mathbf{t}_{ji} = \mathbf{P}_i \Delta \mathbf{r}_j + \mathbf{q}_i.$$

Because the seven matrices \mathbf{P}_i and the seven vectors \mathbf{q}_i are free to vary as a function of illuminant, this affine linear model requires 56 (7×8) parameters to predict the 154 measured coordinates. The fitting was done in an analogous way to that in Subsection 3.A for illuminant changes.

Figure 8 shows the results. As a comparison of the first and sixth bars in each figure shows, changes in the surface collection had a considerable effect on the observed test objects, demonstrating the visual system's adjustments to changes in the surface collection. The second and third bars show the quality of fit of the (weak) affine linear and the (weak) linear model, respectively, predicting the changes in the test objects by means of seven affine linear and linear transformations, respectively, that are permitted to vary with the illuminant. Both models led to rather poor predictions of the test objects. The fourth and fifth bars show the quality of fit of the (strong) affine linear and the (strong) linear model, respectively, predicting the changes in the test objects by means of only one affine linear and linear transformation, respectively, that does not depend on the illuminant. Again, the two models' weak and strong forms are nested, as are the affine

linear and the linear models. Thus the affine linear model leads to better predictions than those of the linear model, and the two models' strong forms leads to worse predictions than their weak forms. The fit to data, however, did not improve in a substantial way when the more general affine linear transformations were used. Also, the deterioration in fit with the models' respective strong forms was fairly small and thus reflects the fairly small interplay of illuminant and surface collection already found in Subsection 3.A for illuminant changes. As a re-

sult, there is not much benefit in using the mean reflectance function to predict the collection-induced changes in color appearance, either by means of a linear transformation or by means of a more general affine linear transformation.

4. DISCUSSION

A. Illuminant Changes

For any collection of surface reflectances, illuminant changes induced changes in the achromatic object that could be well approximated by a linear transformation. Brainard and Wandell⁷ called this regularity an illuminant change linearity. They found strong evidence for it in an experimental paradigm similar to the present one, using a large collection of different surfaces with a yellowish mean reflectance function. The present study generalizes their results to surface collections with a much smaller number of different surfaces and with quite different mean reflectance functions. The present study also extends the range of illuminant changes used by Brainard and Wandell. While these investigators used variations in the coordinates of the first and second basis functions of the three-dimensional model of Judd *et al.*,³ I used variations in the coordinates of the model's second and third basis functions. Taken together, these results suggest that over a wide range of surface collections the effect of illuminant changes on color appearance can be described in a linear way. A comparison of quality of fit of the linear rule with quality of fit of the affine linear rule confirmed this suggestion. The description of the present data did not improve in a substantial way when the more general model was used.

Illuminant change linearity is a considerable theoretical simplification. It implies two restrictive principles, homogeneity and additivity. Homogeneity means that changes in the test objects are proportional to changes in the illuminant. Additivity means that the change in the test object induced by the sum of two illuminants is simply the sum of the changes induced by the two illuminants separately. These two principles gain special practical significance because of another regularity. As Judd *et al.* showed, daylight can be well approximated by a small-dimensional linear model. Thus, in effect, one must measure an object's change in color appearance only for a fairly small number of illuminant changes. Based on these measurements, linearity then provides a simple rule to predict an object's change in color appearance for all linear combinations of these illuminant changes.

The linear transformations describing the effect of illuminant changes on color appearance showed relatively small dependency on a given surface collection. This result means that knowing the effect of illuminant changes for one surface collection led to reasonable predictions on this effect for the other surface collections. In this sense the data are roughly consistent with the view that the changes in an object's color appearance that are induced by changes in illumination do not vary with the collection of surface reflectances in a scene. Even if it is only an approximation, this view constitutes a simple and useful first-order model to describe the interplay between the effect of illuminant and the effect of surface collection in color appearance.

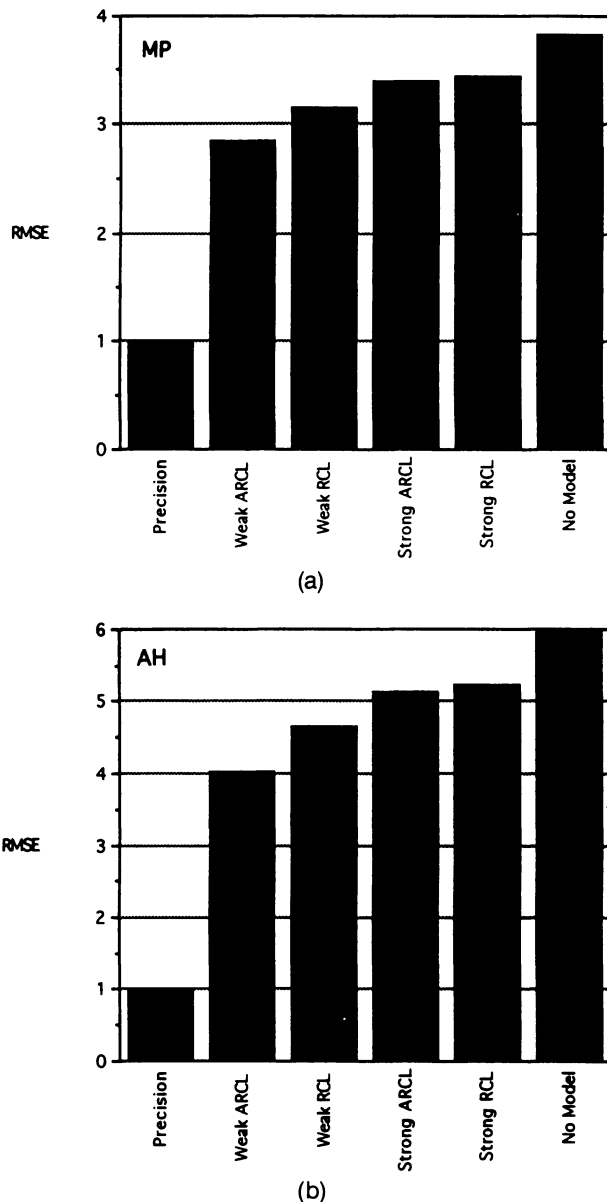


Fig. 8. Changes in the surface collection: quality of model fit. Comparison among a subject's precision, the size of the effect (no model), and the quality of prediction for affine reflectance change linearity (ARCL) and reflectance change linearity (RCL). The 154 measured coordinates were described for the weak affine linear model (Weak ARCL) by 7 matrices and 7 vectors with 56 parameters, for the weak linear model (Weak RCL) by 7 matrices with 42 parameters, for the strong affine linear model (Strong ARCL) by only 1 matrix and 1 vector with 8 parameters, and for the strong linear model (Strong RCL) by only 1 matrix with 6 parameters. The error measurements were root-mean-square errors (RMSE).

The extent to which the independence of the illuminant effect from a scene's surface collection holds may be a function of several features in a set of surface collections. The present experiment suggests that at least two features are not a major source for dependency: variation with respect to the hue gamut spanned by the surface collections and variation with respect to the saturation of the collections' reflectance functions (see Section 2). In fact, variation in these two factors induced considerable effects in the subjects' settings without showing a major interaction with the illuminant effect. The same may not be true for surface collections that vary largely with respect to their surface lightness. The collections used in the present study hardly varied with respect to this feature.

B. Changes in the Surface Collection

Changes in the surface collection induced changes in the achromatic object. At first this effect showed some rough correlation with changes in the collections' mean reflectance functions. More detailed analyses, however, revealed that the mean reflectance function could not capture the effect of surface collection: equal mean reflectance functions, when they were rendered under the same illuminant, did not induce the same achromatic object. This finding rejects the idea of a well-defined function that transforms changes in the mean reflectance function into changes in an object's color appearance to account for the effect of surface collection. As a result, it challenges models of color appearance that use simple mean assumptions to account for changes in the surface collection. This holds true both within a theoretical framework in which the image intensities are separated into illuminant and surface collection and within the theoretical framework of image intensities confounding illuminant and surface collection. Concerning the latter theoretical framework, this result is consistent with a recent finding by Fuchs.³¹ Using two combinations of illuminant and surface collection that shared the same average image intensities, she found objects' color appearance to be different for the two combinations. *A priori*, however, this finding might have resulted because of differently strong, separate effects of changes in the illuminant and of changes in the surface collection, thus not ruling out the mean assumption for a framework that separates the image intensities into their illuminant and surface collection components.

The effect of surface collection shown in the present experiment might have at least two sources. First, it might have been driven by local effects of nearby surfaces. Local effects, often called simultaneous contrast, are well known in the literature.⁸ However, in general, local effects are assumed to occur mainly at the border of nearby surfaces. In this experiment I separated test object and surrounding surfaces both vertically and horizontally in order to minimize simultaneous contrast. As a result, local effects should have been rather small in the present experiment.

Second, the effect of surface collection might reflect an effect of the surface collection on the visual system's estimation of the illuminant. Different surface collections might induce different absolute adjustments to the same illuminant. Indeed, from computational models of color

constancy it is well known that the estimation of the illuminant can strongly depend on the surface collection.³²⁻³⁴ This holds particularly if the number of surfaces in an image is only moderate or if the surfaces span only a rather small gamut. This suggestion is not in contradiction to the illuminant change linearity finding reported in Subsection 4.B. Illuminant change linearity highly restricts the visual system's adjustments to illuminant changes. However, it is completely silent on the system's absolute adjustments to illuminants and therefore permits different absolute adjustments to the same illuminant in different images.

C. Measuring Color Appearance

Measuring the achromatic locus on an isoluminant plane generally provides only a two-dimensional measurement of color appearance. Under the conditions met in the present experiment these measurements were implicitly regarded as full, three-dimensional measurements of color appearance. Indeed, there is good empirical evidence that the lightness of an object is maintained under chromatic changes in the viewing context if object and context are equiluminant.^{17,35} These conditions are fairly well met in the present experiment. In this sense the achromatic loci measured under the different viewing contexts constituted asymmetric color matches between a specified achromatic standard object under a standard context and matching achromatic test objects under test contexts.³⁶ In fact, the two subjects did not report any major failures of this assumption; the test objects appeared approximately equiluminant to them across varying contexts. Conjectures against this assumption should have led to deterioration in the fit of the models.

D. From Achromatic Objects to Chromatic Objects

The question arises of how the reported results for the achromatic object might generalize to other, chromatic test objects. If illuminant changes linearly changed the relative sensitivity of the different cone classes^{4,7,15,17} or some other putative postreceptor mechanisms,^{37,38} the mapping of matching standard and test objects could be described by a diagonal linear transformation. As a result, knowing the matching test object for only one standard object would be sufficient to fix the whole mapping for a given illuminant change. Knowing how illuminant changes affect the color appearance of the achromatic object would then be equivalent to knowing how any test object is affected by these illuminant changes. In this case the results reported in the present study for the achromatic object could be generalized to other chromatic objects. On the other hand, if more general models than the diagonal linear model were necessary to account for illuminant changes,^{5,6,16,39,40} such a generalization would be at least premature.

5. CONCLUSIONS

The present experiment reveals an effect of changes in illuminant and of changes in the surface collection on the color appearance of an achromatic object. The effect of illuminant changes could be well approximated by a linear transformation that mapped changes in illuminants into changes in the achromatic objects. This mapping was

Table 2. Experimental Illuminants^a

	d_0	d_1	d_2	x	y	L
D_1	2.5×10^{-3}	7.236×10^{-3}	4.097×10^{-3}	0.249	0.249	193.4
D_2	2.5×10^{-3}	2.997×10^{-3}	-5.410×10^{-4}	0.274	0.282	185.4
D_3	2.5×10^{-3}	2.782×10^{-4}	-1.907×10^{-3}	0.300	0.313	181.0
D_4	2.5×10^{-3}	-1.543×10^{-3}	-1.083×10^{-3}	0.326	0.339	179.0
D_5	2.5×10^{-3}	-2.777×10^{-3}	1.462×10^{-3}	0.351	0.362	178.7
D_6	2.5×10^{-3}	-3.589×10^{-3}	5.565×10^{-3}	0.377	0.380	179.7
D_7	2.5×10^{-3}	-4.069×10^{-3}	1.125×10^{-2}	0.402	0.394	182.0

^aAll seven illuminants stem from the CIE daylight locus. They were constructed from a three-dimensional linear model of natural daylight (see Section 2). This table gives the illuminant weights for the three basis functions, the corresponding CIE xy coordinates, and the luminance (D_i = daylight number i , d_k = weight for basis function $k + 1$, x = x chromaticity, y = y chromaticity, L = luminance).

Table 3. Mean Reflectance Functions of the Experimental Surface Collections^a

	r_1	r_2	r_3	x	y
R_1	-1.484	-0.216	-0.033	0.336	0.336
R_2	-1.483	-0.207	-0.030	0.337	0.337
R_3	-1.484	-0.196	-0.029	0.338	0.337
R_4	-1.491	-0.193	-0.035	0.339	0.336
R_5	-1.501	-0.186	-0.044	0.339	0.335
R_6	-1.478	-0.202	-0.024	0.338	0.338
R_7	-1.346	0.217	0.214	0.416	0.414
R_8	-1.274	-0.281	0.168	0.332	0.371
R_9	-1.847	-0.029	-0.364	0.349	0.293
R_{10}	-1.569	-0.396	-0.165	0.309	0.306
R_{11}	-1.619	0.115	-0.093	0.379	0.342
R_{12}	-1.254	-0.757	0.073	0.265	0.327

^aAll the experimental reflectance functions can be expressed by a three-dimensional linear model (see Section 2). This table gives for each surface collection's mean reflectance function the weights for the three basis functions. Also, the CIE xy coordinates of these mean reflectance functions are given when they are rendered under spectrally flat illumination. (R_j = mean reflectance function of surface collection number j , r_k = weight for basis function k , x = x chromaticity, y = y chromaticity.)

only moderately affected by a given surface collection. To first approximation, this finding suggests that the effect of changes in the illuminant on color appearance can be described linearly and that it can be separated from the surface collection.

I did not find any regularity to account for the effect of changes in surface collection on an object's color appearance. Instead, the main result of the present study is a negative one. The mean reflectance function cannot account for changes in color appearance that are induced by changes in the surface collection. This result challenges models of color appearance that use simple mean assumptions to account for changes in the surface collection, both within the theoretical framework of image intensities confounding illuminant and surface collection and within a theoretical framework in which the image intensities are separated into illuminant and surface collection.

APPENDIX A

This appendix consists of two tables. They specify the experimental stimuli. Table 2 specifies each of the seven simulated illuminants. Table 3 specifies the mean reflectance function of each of the 12 simulated surface collections.

ACKNOWLEDGMENTS

I thank K. R. Gegenfurtner, B. A. Wandell, and the two anonymous reviewers for their helpful comments. This study was supported by the Deutscher Akademischer Austauschdienst/North Atlantic Treaty Organization.

REFERENCES AND NOTES

1. S. T. Henderson and D. Hodgkiss, "The spectral energy distribution of light," *Br. J. Appl. Phys.* **14**, 125-131 (1964).
2. H. R. Condit and F. Grum, "Spectral energy distribution of daylight," *J. Opt. Soc. Am.* **54**, 937-944 (1964).
3. D. B. Judd, D. L. MacAdam, and G. W. Wyszecki, "Spectral distribution of typical daylight as a function of correlated color temperature," *J. Opt. Soc. Am.* **54**, 1031-1040 (1964).
4. J. v. Kries, "Die Gesichtsempfindungen," in *Handbuch der Physiologie des Menschen*, W. Nagel, ed. (Vieweg-Verlag, Braunschweig, 1905), Vol. 3.
5. R. W. Burnham, R. M. Evans, and S. M. Newell, "Prediction of color appearance with different adaptation illuminations," *J. Opt. Soc. Am.* **47**, 35-42 (1957).
6. D. Jameson and L. M. Hurvich, "Color adaptation: sensitivity, contrast, after-images," in *Handbook of Sensory Physiology*, L. M. Hurvich and D. Jameson, eds. (Springer-Verlag, Berlin, 1972), Vol. VII/4.
7. D. H. Brainard and B. A. Wandell, "Asymmetric color-matching: how color appearance depends on the illuminant," *J. Opt. Soc. Am. A* **9**, 1433-1448 (1992).
8. G. W. Wyszecki, "Color appearance," in *Handbook of Perception and Human Performance*, K. R. Boff, L. Kaufman, and J. P. Thomas, eds. (Wiley, New York, 1986), Vol. I.
9. H. Helson, D. Judd, and M. Warren, "Object-color changes from daylight to incandescent filament illumination," *Illum. Eng.* **47**, 221-233 (1952).
10. Both the method of memory matching and the measurement of subjects' achromatic locus permit us to control a subject's state of adaptation experimentally. Brainard and Wandell⁷ discuss respective problems when using some other experimental methods.
11. E. Hering, *Zur Lehre vom Lichtsinn* (Carl Gerold's Sohn, Vienna, 1878).
12. D. H. Krantz, "Color measurement and color theory: II. Opponent-colors theory," *J. Math. Psychol.* **12**, 304-327 (1975).
13. J. Larimer, D. H. Krantz, and C. C. Cicerone, "Opponent process additivity—I. Red/green equilibria," *Vision Res.* **14**, 1127-1140 (1974).
14. J. Larimer, D. H. Krantz, and C. C. Cicerone, "Opponent process additivity—II. Yellow/blue equilibria and nonlinear models," *Vision Res.* **15**, 723-731 (1975).
15. J. Walraven, "Discounting the background—the missing link in the explanation of chromatic adaptation," *Vision Res.* **16**, 289-295 (1976).
16. S. K. Shevell, "The dual role of chromatic backgrounds in color perception," *Vision Res.* **18**, 1649-1661 (1978).
17. J. Werner and J. Walraven, "Effect of chromatic adaptation

- on the achromatic locus: the role of contrast, luminance and background color," *Vision Res.* **22**, 929-943 (1982).
18. K.-H. Bäuml, "A ratio principle for a red/green and a yellow/blue channel?" *Percept. Psychophys.* **53**, 338-344 (1993).
 19. D. H. Krantz, "A theory of context effects based on cross-context matching," *J. Math. Psychol.* **5**, 1-48 (1968).
 20. L. Arend and A. Reeves, "Simultaneous color constancy," *J. Opt. Soc. Am. A* **3**, 1743-1751 (1986).
 21. G. W. Wyszecki and W. S. Stiles, *Color Science*, 2nd ed. (Wiley, New York, 1982).
 22. J. Cohen, "Dependency of the spectral reflectance curves of the Munsell color chips," *Psychonom. Sci.* **1**, 369-370 (1964).
 23. L. T. Maloney, "Evaluation of linear models of surface spectral reflectance with small numbers of parameters," *J. Opt. Soc. Am. A* **3**, 1673-1683 (1986).
 24. K. L. Kelly, K. S. Gibson, and D. Nickerson, "Tristimulus specification of the *Munsell Book of Color* from spectrophotometric measurements," *J. Opt. Soc. Am.* **33**, 355-376 (1943).
 25. K. R. Gegenfurtner, "PRAXIS: Brent's algorithm for function minimization," *Behav. Res. Methods Instrum. Comput.* **24**, 560-564 (1992).
 26. Krantz¹⁹ presents a rationale for the use of affine linear transformations in any cross-context matching situation.
 27. Because the illuminant changes seem to induce color appearance changes mostly along the yellow-blue dimension, I tested the idea that a one-dimensional transformation would be able to explain the data. In fact, when we restricted the two rows of the matrices M_j , or M , to be linear dependent, the fit of the two linear models did not deteriorate severely. For the weak linear model the root-mean-square error for a one-dimensional model is 1.644 for subject AH (1.341 for a two-dimensional model) and 1.373 for subject MP (1.225 for a two-dimensional model). For the strong linear model the root-mean-square error for a one-dimensional model is 2.166 for subject AH (2.110 for a two-dimensional model) and 1.642 for subject MP (1.549 for a two-dimensional model).
 28. W. R. J. Brown, "Statistics of color-matching data," *J. Opt. Soc. Am.* **42**, 252-256 (1952).
 29. S. S. Wilks, "Certain generalizations in the analysis of variance," *Biometrika* **24**, 471-494 (1932).
 30. T. W. Anderson, *An Introduction to Multivariate Statistical Analysis* (Wiley, New York, 1958).
 31. H. Fuchs, "Color constancy in non-grey average surrounds," submitted to *J. Opt. Soc. Am. A*.
 32. G. Buchsbaum, "A spatial processor model for object color perception," *J. Franklin Inst.* **310**, 1-26 (1980).
 33. L. T. Maloney and B. A. Wandell, "Color constancy: a method for recovering surface spectral reflectance," *J. Opt. Soc. Am. A* **1**, 29-33 (1986).
 34. D. H. Brainard and B. A. Wandell, "A bilinear model of the illuminant's effect on color appearance," in *Computational Models of Visual Processing*, J. A. Movshon and M. S. Landy, eds. (MIT Press, Cambridge, Mass., 1991).
 35. M. P. Lucassen and J. Walraven, "Quantifying color constancy: evidence for nonlinear processing of cone-specific contrast," *Vision Res.* **33**, 739-759 (1993).
 36. The condition that test object and context are exactly equi-luminant does not seem to be critical in the present experiment. For 3 of the 84 combinations of illuminant and surface collection and one subject (MP), I measured the achromatic locus also on an isoluminant plane of 30 and 70 cd/m^2 . There was no substantial effect of the luminance level on the measured xy coordinates, suggesting a scalar invariance, or Bezold-Brücke invariance, of the achromatic object for that luminance range. Thus the pattern of results reported in this paper would not have changed if the luminance level of the test objects had not been 50 cm^2/m^2 but had been anywhere between 30 and 70 cd/m^2 .
 37. J. Krauskopf, D. R. Williams, and D. W. Heeley, "Cardinal directions of color space," *Vision Res.* **22**, 1123-1131 (1982).
 38. J. Krauskopf and K. R. Gegenfurtner, "Color discrimination and adaptation," *Vision Res.* **32**, 2165-2175 (1992).
 39. C. C. Cicerone, D. H. Krantz, and J. Larimer, "Opponent-process additivity—III. Effect of moderate chromatic adaptation," *Vision Res.* **15**, 1125-1135 (1975).
 40. R. W. C. Hunt, "A model of colour vision for predicting color appearance in various viewing conditions," *Color Res. Appl.* **12**, 297-314 (1987).