Color appearance of real objects varying in material, hue, and shape

Department of Psychology, Justus-Liebig-University, Giessen, Germany, & SUNY College of Optometry, Graduate Center for Vision Research, New York, NY, USA

Martin Giesel

Karl R. Gegenfurtner

The objects in our environment are made of a wide range of materials. The color appearance of the objects is influenced by many factors, including the geometry of the illumination, the shape of the objects, and the reflectance properties of their materials. Only few studies have investigated the effect of material properties on color perception, mostly with stimuli rendered on a computer screen. Here we set out to investigate color perception for real objects made of different materials. The surface properties of the materials ranged from smooth and glossy to matte and corrugated. We tested objects with similar colors made from different materials and objects made from the same material that differed only in color. Observers matched the color appearance of the objects by adjusting the chromaticity and the luminance of a homogeneous, uniformly colored disk presented on a CRT screen. The observers matched the hue of the objects quite accurately. Chroma matches were modulated by the lightness of the objects. For dark objects, chroma was overestimated, while for light objects it was underestimated. For lightness, observers matched the brightest points of the objects excluding highlights. This is a suitable strategy to compensate for variations in surface geometry and illumination.

Keywords: color vision, color appearance, lightness/brightness perception, 3D surface perception


Introduction

Color matching is the most basic task in color vision, and forms the basis of colorimetry. In the early experiments on color matching, homogeneously colored stimuli were used (e.g., Brown, 1957; Brown & MacAdam, 1949; MacAdam, 1942, 1949; Wright & Pitt, 1934; Wyszecki, 1965, 1972, 1978; Wyszecki & Fielder, 1971a, 1971b; Wyszecki & Stiles, 1980). These experiments were intended to assess the sensitivities of the human visual system to different colors, and elucidate the laws that govern human color vision at the level of the photoreceptors. However, our visual environment and the objects in it contain very few surfaces that are of one uniform color. Moreover, these surfaces are made of a wide range of materials. Different materials exhibit different reflection properties, and these properties introduce variations in the reflected light. Color is an important cue in the perception of materials. While variations in luminance can be caused by both changes in material and changes in illumination, variations in color are highly diagnostic for changes in material (Kingdom, 2008; Shevell & Kingdom, 2008).

The light reflected from an object depends on the geometry and spectral distribution of the incident light, the object’s shape, and the material (or materials) from which the object is made (Dorsey, Rushmeier, & Sillion, 2008). The material determines in which way the incident light is reflected by the object. When dealing with opaque non-uniform objects (dielectrics) — as we mainly do in the following — two types of reflectances occur: surface reflectance and body reflectance (Klinker, 1993). Surface reflectance occurs when light encounters a change in the refractive index at the material surface, and accounts for specular highlights on the object. The specular reflections are influenced by the roughness of the surface, the wavelength of the light, and self-shadowing effects on the surface. Body reflectance occurs within the material body where light is scattered and selectively absorbed, e.g., by pigments, before it leaves the material body. The intensity and color of light is influenced by the distribution of the pigment. When faced with the task of estimating the color of an object, the variety of factors that influence the reflected light has to be taken into account to be able to extract the surface reflectance properties of the object, i.e., the visual system has to disentangle the contributions of illumination, surface geometry, and reflectance based on the information compressed in the two dimensional retinal image. To do this, the visual system might rely on various cues working both on a local and global scale (Kingdom, 2008).
One line of recent research has found that the effects of scene geometry on color appearance are at least partially taken into account (Boyaci, Doerschner, & Maloney, 2004; Boyaci, Maloney, & Hersh, 2003; Doerschner, Boyaci, & Maloney, 2004; Khang & Zaidi, 2004; Ripamonti et al., 2004; Yang & Maloney, 2001).

Bloj, Kersten, and Hurlbert (1999) found that the relative complex interactions between shape and color that are induced by inter-reflections are compensated by the visual system. Their results indicate that color perception strongly interacts with three-dimensional shape perception.

Other studies investigated the effect of the spatial distribution of luminances and chromaticities on the perception of lightness, color, and gloss. It is a common albeit not self-evident observation that observers have no problem assigning a single color name to an object in spite of its variations in chromaticity and luminance, at least if the chromatic variation of the object roughly belongs to the same color category. Kuriki (2004) and Sunaga and Yamashita (2007) investigated how a single color impression is determined from multicolored and two-color textures, respectively. Their results did not support the colorimetric averaging hypothesis which claims that a single color impression is determined by the colorimetric average of the chromaticities of the single elements that constitute the texture. Poirson and Wandell (1993) investigated how color appearance is influenced by spatial patterns by letting observers set color matches between a uniform patch and the individual bars of a square-wave pattern. Their results show that these color matches were not photoreceptor matches but were established at more central neural sites.

Brenner, Granzier, and Smeets (2007) used classification images to examine which parts of a surface are particularly important when observers judge its color. They found that observers tended to report the colors of points close to the fixation point, and concluded that eye movements determine which part of a surface will be given most weight.

Using real objects made from stucco-like material Motoyoshi, Nishida, Sharan, and Adelson (2007) and Sharan, Li, Motoyoshi, Nishida, and Adelson (2008) (but also see Anderson & Kim, 2009) reported a correlation between the skewness of the luminance histogram and the glossiness of their stimuli. With increasing positive skew of the luminance histogram the stimuli were rated to appear glossier. The lightness ratings of the stimuli were negatively correlated with the skewness of the luminance histogram. Lightness ratings decreased with increasing positive skew.

Xiao and Brainard (2008) investigated the effect of material on color perception by measuring the influence of gloss on the color appearance of rendered three-dimensional objects. They found that the effect of material on the color matches was small. Their results also did not support the averaging hypothesis.

In many applications there is a need to match colors between non-homogeneous surfaces, e.g., in the evaluation of the color of food, hair, teeth, or textiles. A common approach to accomplish this, is to evaluate the similarity in color between different samples by taking colorimetric measurements and then to compute the difference between them in a standardized color space, e.g., the CIELAB (CIE, 1978) color space. However, colorimetry is based on measurements using simple stimuli and does not account for more complex stimuli that vary in chromaticity and luminance. The percept of such a stimulus cannot simply be explained by the percepts evoked by its single elements.

For chromatic discrimination it has been shown that chromatic distributions influence chromatic discrimination thresholds in a systematic way (Giesel, Hansen, & Gegenfurtner, 2009; Hansen, Giesel, & Gegenfurtner, 2008; tePas & Koenderink, 2004; Zaidi, Spehar, & DeBonet, 1998) that cannot be explained on the basis of the single elements of the chromatic texture.

Here, we present exploratory experiments which address two main questions. First, how does a global color percept evolve when an observer is presented with an object that varies in chromaticity and luminance? Second, is this global color percept influenced by the material of the object? As stimuli we used real objects made from different materials presented under controlled illumination conditions. The surface properties of the materials ranged from smooth and glossy (porcelain) to matte and corrugated (crumpled paper). We tested objects with similar colors (red, yellow, green, and blue) made of different materials, and objects made of the same material that differed only in color. We asked observers to adjust the color of a uniform disk presented on a computer screen until its color appearance represented a good match to the color appearance of the object. The observers’ matches were compared to photometric measurements made at various points across the part of the surface of the objects that was visible to the observers.

Our results show that the observers do not simply take the average color across the objects. Rather, observers tended to discount the variations in reflected light induced by the geometry of the objects and matched the light reflected from the materials themselves.

**Methods**

**Apparatus**

**Illumination chamber**

The experiments were conducted in a chamber whose illumination was under computer control (Figure 1). The observers were seated at a distance of 125 cm from the wall in front of the chamber. The monitor was placed 60 cm behind the wall in a tunnel lined with black felt,
and was viewed through an opening of 10.8 × 8.4 degrees of visual angle.

The wall of the chamber was painted with a neutral gray. Two sets of three Osram L36W/66 fluorescence lamps (red, green, and blue) illuminated this wall from two sides. The lamps illuminated about 64 × 45 degrees of visual angle of the viewing field in a fairly homogeneous fashion. Light from the fluorescence lamps could not reach the monitor surface directly. Observers saw only the light emitted from the monitor phosphors and the light reflected from the wall.

The lamps were controlled by a digital-to-analogue converter with a resolution of 12 bits per channel. Even though the relationship between voltage and intensity was almost linear for the lamps, small residual nonlinearities were corrected by using lookup tables. The chromaticities of the three lamp-pairs were measured on the painted wall with the Photo Research PR-650 spectroradiometer (Photo Research Inc., Chatsworth, CA). The paint of the wall was chosen to make the lamp primaries similar to the monitor primaries, resulting in nearly the same gamut for the background illumination and the monitor. The calibration of the setup is described in detail in Rinner and Gegenfurtner (2002). The chromaticity coordinates of the illumination were measured using the RS-3 Reflectance Standard placed at the same location where the stimuli were presented in the experiment. The CIE 1931 coordinates of the illumination were x = 0.2928, y = 0.3282, and Y = 99.4740. The CIE 1931 xyY coordinates of the wall were x = 0.2936, y = 0.3406, and Y = 37.6856.

**Monitor**

The match disk was displayed on a SONY GDM-F520 color CRT monitor. The monitor resolution was set to 1,280 × 1,024 pixels with a refresh rate of 100 Hz noninterlaced. The monitor was controlled by a PC with a color graphics card with 8-bit intensity resolution for each of the three monitor primaries. The nonlinear relationship between voltage output and luminance was linearized by a color look-up table for each primary. To generate the three RGB look-up tables, we measured the luminance of each phosphor at various voltage levels using a Graseby Optronics Model 307 radiometer with a model 265 photometric filter, and a smooth function was used to interpolate between the measured data. The spectrum of each of the three primaries at their maximum intensity was measured with the Photo Research PR-650 spectroradiometer. The obtained spectra were then multiplied with the Judd-revised CIE 1931 color matching functions (Judd, 1951; Wyszecki & Stiles, 1982) to derive CIE xyY coordinates of the monitor phosphors. The xyY coordinates of the monitor primaries at maximum intensity are given by R = (0.6146, 0.3477, 20.6326), G = (0.2828, 0.6056, 59.5495), and B = (0.1555, 0.0831, 8.3094). The xyY coordinates were then used to convert between RGB and DKL color space.

**Stimuli**

Eighteen objects were used as stimuli (Figure 2). The objects belonged to five material categories (paper, sponge, wool, wax, and porcelain) and were chosen to fall into one of four color categories: red, yellow, green, and blue. The stimuli were placed at a distance of 110 cm from the observers. The stimuli subtended between 3 and 8 degrees of visual angle. The stimuli were placed on a table directly below the opening for the monitor. The table...
as well as the part of the wall surrounding the table was covered with black cloth. Figure 3 depicts the chromatic distributions of the objects in CIELAB color space based on spectroradiometric measurements of the objects (see below).

Control stimuli

For the crumpled papers and the wool balls, we also measured the spectral distributions for control stimuli (Figure 4). For the crumpled papers, these were sheets of plain paper identical to those from which the crumpled paper objects were made. For the wool balls, we wrapped the wool tightly around a spool. The papers were presented in an upright position. The control stimuli for the wool balls were presented as shown in Figure 4.

Spectroradiometric measurements

The spectral measurements of the stimuli were made at various points across the part of the object that was visible for the observers during the experiment using the Photo Research PR-650 spectroradiometer. The illumination conditions for these measurements were the same as in the experiment. The measurements were made from the location from which the observers saw the stimuli. The PR-650 measures the optical radiation that passes through a 1° aperture. Measurements were taken across the objects’ surface in 0.5° steps. Depending on the size of the object, this resulted in 40 to 100 points that were measured. From the spectra of these points we computed CIE chromaticity coordinates and luminances. We chose objects for which all measured points were inside the gamut of the monitor on which the matches were made. However, some stimuli, especially the red objects, were close to the gamut border.

Luminance estimation from digital images

Since the spectroradiometer only provides a coarse measurements of the objects, due to the integration of the radiation passing through the 1° aperture, we took photographs of the stimuli presented under the same conditions as in the experiment, and estimated the luminance of each pixel of the images. We used a standard digital camera (Nikon D70). The whitepoint of the camera was set to the coordinates of the illumination. The photographs were taken from the observers’ point of view.

To calibrate the camera, we first took a photograph of the Macbeth ColorChecker, and measured the radiation of each patch of the ColorChecker with the spectroradiometer. The RGB values of each patch were computed by averaging the RGB values of a 100 × 100 pixel cutout from each patch.

Since here we were only interested in estimating the luminance distribution of the objects, we computed a weighted average across the R, G, and B values of the six gray scale patches of the ColorChecker, resulting in one intensity value \( I \) for each of the six patches.

\[
I = w_R R + w_G G + w_B B.
\]  

(1)

We fitted a 5th degree polynomial to model the functional relationship between the values of \( I \) and the luminance measured for the six patches.

\[
Y_{est} = p_1 I^5 + p_2 I^4 + p_3 I^3 + p_4 I^2 + p_5 I + p_6.
\]  

(2)

To find the optimal combination of the RGB values, Equations 1 and 2 with \( w \) and \( p \) determined based on the fit to the six gray scale patches were used to predict the luminance values of all 24 patches of the ColorChecker. The weights \( w \) that minimized the squared error between the predicted and measured luminance values of all 24 patches of the ColorChecker were determined by using the Matlab (The MathWorks, Inc., Natick, MA) function fmincon. The resulting weights were \([w_R, w_G, w_B] = [0.1914, 0.7906, 0.0180]\). The mean absolute deviation of the predicted and measured luminance values was 2.6 cd/m² (range: 0.1–11.2 cd/m²). The mean error in CIELAB was \( \Delta L^* = 3 \). The weights \( w \) were used to compute gray scale versions of the images of the objects. The luminance of each pixel of the images was then estimated by applying Equation 2.

Procedure

Color adjustment method

The software for the matching procedure and the stimulus presentation was programmed in Matlab using the Psychophysics toolbox extensions (Brainard, 1997;
The matches were made on a uniform disk that subtended 3 degrees of visual angle. The background of the monitor was set to black.

To change the color of the matching disk, the observers could adjust the chromaticity and luminance of the disk along the two chromatic and the achromatic axes of the DKL color space (Derrington, Krauskopf, & Lennie, 1984; Krauskopf, Williams, & Heeley, 1982) by pressing corresponding buttons on a computer keyboard. The DKL color space is a second stage cone-opponent color space, which reflects the preferences of retinal ganglion cells and LGN neurons. It is spanned by an achromatic luminance axis, the L + M axis, and two chromatic axes, the L − M axis and S = (L + M) axis. The L + M axis is determined by the sum of the signals generated by the long wavelength cones (L-cones) and the middle wavelength cones (M-cones). The L − M axis is determined by the differences in the signals generated by the L-cones and the M-cones. Along the L − M axis the L- and M-cone excitations co-vary at a constant sum, while the S-cone excitation does not change. Colors along the L − M axis vary between reddish and bluish-greenish. The S = (L + M) axis is determined by the differences in the signals generated by the short wavelength cones (S-cones) and the sum of the L- and M-cones. Along the S = (L + M) axis only the excitation of the S-cones changes and colors vary

Figure 3. Chromatic distributions of the stimuli based on spectroradiometric measurements. The center plot shows the chromatic distributions of all stimuli in the chromatic plane of the CIELAB color space. The colors indicate the color of the objects. The surrounding plots show the same chromatic distributions of the stimuli separately for the four object colors (right: red; top: yellow; left: green; bottom: blue). For better visibility, the different colors in the surrounding plots indicate the different materials (blue: paper; cyan: sponge; magenta: wool; green: wax; porcelain: red.)
between yellow-greenish and purple. The axes of the DKL color space were arbitrarily scaled from $-1$ to 1, where $z1$ corresponds to the maximum contrast achievable on the monitor used. Although the DKL color space is a physiologically defined space, here we made no use of this feature but rather employed it as one of various possible implementations of an opponent color space. We have used this color space in previous experiments, and have found that observers quickly became familiar with this space, and had no problem to navigate in it.

**Experiment**

One experimental session consisted of three consecutive matches to each of the 18 stimuli. For each observer, the stimuli were presented in randomized order. The initial chromaticity and luminance of the match disk were randomized on every trial. The observers were instructed to adjust the color and the luminance by pressing the appropriate keys until they had the impression that the color appearance of the match disk was identical to the color appearance of the objects. No further explanation was given to the observers. Especially, we did not want to suggest any kind of strategy to our observers, e.g., to discount the illumination.

The observers completed three practice trials with a test object before starting with the experiment. During the experiment, they had unlimited time to make a setting. When the observers tried to make settings that would have been out of the gamut of the monitor they were warned by a message appearing on the screen. The RGB values of the final setting were recorded after the observer had indicated that she/he was satisfied with the match by pressing a key. After that the next trial started. After three matches to one stimulus, the next stimulus was placed on the table by the experimenter. The whole session was repeated on another day. One session took between 60 to 90 minutes.

**Observers**

13 paid observers (6 female, 7 male; age range 20–28 years) participated in the experiment. All observers were university students, and were naive as to the purpose of the experiment. All observers had normal or corrected-to-normal vision and normal color vision as tested with Ishihara plates (Ishihara, 1973).

**Data processing**

After each session, the match disk was re-displayed on the monitor with its color set to the recorded RGB values of the observers’ matches to each stimulus. The disk was then measured with the spectroradiometer from the observer’s viewpoint, and the chromaticity coordinates and the luminance of the match disk were computed from the measured spectra.

Since in the following we present the data in the CIELAB color space we used the CIE 1931 color matching functions to convert from spectra to chromaticity coordinates, although the CIE 1964 color matching functions for the 10° observer might have been the more appropriate choice. However, we also did the analysis on the basis of the CIE 1964 color matching functions and found no large differences in the deviations of the observers’ matches from the values measured for the objects between the two color matching functions. For the transformation to CIELAB, we used the coordinates of the illumination as whitepoint. The results are presented in the polar variant of the CIELAB coordinates, i.e., as lightness ($L^*$), chroma ($C^*$), and hue ($h^*$). When in the following we speak of the objects’ lightness, chroma, or hue we refer to the lightness, chroma, and hue values derived from the spectroradiometric measurement of the objects.

For the data analysis, we averaged the observers’ data across repetitions and sessions. For the hue data, circular statistics were used. The deviations of the lightness and chroma matches from the mean object values were analyzed using repeated-measures multi-variate analysis of variance (MANOVA) with object color and object material as independent variables, and the lightness and chroma shifts of the matches as dependent variables. The MANOVA showed a highly significant multivariate effect for both the object color and the object material. We analyzed the data in more detail using repeated-measures analysis of variance (ANOVA) with the factors object color (red, yellow, green, and blue) and object material (paper, sponge, wool, and wax). Values are presented as means $\pm$ standard errors of the mean. Since the objects made from porcelain (mugs) were only represented in two of the four levels of the first factor, they were excluded from this analysis. Post-hoc contrasts were carried out using Fisher’s LSD testing procedure. In the statistical analysis, the significance level was set to $\alpha = 0.05$. 

Figure 4. Control stimuli for the crumpled papers (bottom row) and the wool balls (third row).
Results

Figure 5 shows the color matching ellipses for the 18 objects in the chromatic plane of the CIELAB color space. Ellipses were determined by computing the eigenvectors of the covariance matrix of the mean matches of the observers. In the following, we first present the results for the chroma and hue matches, and then the results for the lightness matches. The results in Figures 6 and 10 are presented in separate figures for each of the four object colors. The object color is indicated by the background color of the figure (from top to bottom: red, yellow, green, and blue). Same materials are located in the same column. The material porcelain was only represented in object colors yellow and blue. For each stimulus, we show the measures derived from the spectroradiometric measurements of the objects, and the results of the matching experiment side by side. For the spectroradiometric measurements (black), the filled circles represent the average across all measurements, the vertical lines represent the range of the measurements. For the matches (red), the open circles show the individual observers’ matches averaged across repetitions and sessions. The horizontal lines indicate the mean match averaged across observers. The bar plots show the mean distance between the mean object measure (filled black circle), and the individual observers’ matches (open red circles). The error bars denote the standard error of the mean difference.

Chroma and hue matches

Figure 6 shows the results for the chroma matches. The repeated-measures ANOVA showed a significant main effect of object color \((F(3, 36) = 24.40, p < 0.001)\) and object material \((F(3, 36) = 11.50, p < 0.001)\) on the mean distance between the chroma of the object and the chroma of the matches. There also was a significant interaction between the two factors \((F(9, 108) = 34.97, p < 0.001)\).

The largest shifts in chroma occurred for the red objects \((15.34)\). Post-hoc tests showed that the magnitude of these shifts was significantly larger than the shifts for the yellow \((6.90.62)\), green \((8.10.62)\), and blue objects \((8.90.73)\) which were not significantly different from each other with respect to the magnitude of the shift. In addition, there were differences between the object colors in the direction of the chroma shift. For the red objects, the mean chroma matches were shifted to higher values, i.e., they were more saturated than the mean object chroma. A similar shift of the matches to higher chroma was found for the blue objects. The deviations of the matches for the yellow objects from the mean object chroma were smallest. Contrary to the red and blue objects, the matches for the yellow objects tended to be shifted to lower chroma values. The matches for the green objects showed shifts in different directions. The paper and the wool were matched with higher chroma values, while the matches for the candle were close to the mean measured values, and the matches for the sponge exhibited a displacement to lower chroma.

With respect to the influence of object material on the chroma matches, the largest shifts away from the mean object chroma were found for the materials wax \((11.80.54)\) and wool \((11.30.87)\). Post-hoc tests showed that both shifts were significantly larger than the shifts for the materials paper \((8.50.86)\) and sponge \((8.00.55)\). With respect to the direction of the shift away from the measured value no consistent pattern for the different materials emerged.

This can be more clearly seen in Figure 7 where the mean chroma matches are plotted against the mean chroma of the objects. The symbols for all yellow objects are placed below the line. This indicates that the mean chroma matches for the yellow objects were lower than the mean chroma of the objects. All red and blue objects are located above the line corresponding to a shift of the mean matches to higher chroma compared to the mean chroma of the objects. Only for the green objects differences in the direction of the chroma matches were found. The green paper and wool were matched with higher chroma values whereas the mean match for the candle was close to the mean chroma.

![Figure 5](https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933478/ on 11/23/2018)
of the object, and the match for the green sponge was shifted to lower chroma values.

These deviations of the chroma matches seem rather unsystematic at first sight. However, when plotted as a function of the lightness of the objects (Figure 8) it can be seen that there is a high negative correlation ($r = -0.80$) between these two variables. Darker objects tended to appear more and lighter objects less saturated. There is no such correlation between the lightness and chroma of the objects based on the spectroradiometric measurements ($r = 0.12$). Since the observers' matches were also measured with the spectroradiometer these results cannot result from calibration or color space related issues.

As can be seen in Figure 3 there was not much variation in the hue of the objects since their chromatic distributions lie approximately on a line originating from the origin. There also was little variation in the observers' hue matches. Figure 9 shows the mean hue matches plotted against the mean hue of the objects.

In general, the mean hue matches were close to the mean hue of the objects. The mean hue shifts of the matches away from the mean object hue were small (red: $4.3^\circ$; yellow: $4.9^\circ$; green: $8.3^\circ$; blue: $7.9^\circ$) with the largest deviations being around $12^\circ$ for the green and blue wool, and the green sponge. The matches for all green objects and for four of the five blue objects were shifted to larger hue angles. There was no effect of material on the hue matches. There also is no indication for a shift of the hue matches toward unique hues supposing that the unique hue angles in CIELCh are approximately at $24^\circ$, $90^\circ$, $162^\circ$, and $246^\circ$ for red, yellow, green, and blue, respectively.

Lightness matches

There was larger intra- and interindividual variability in the luminance settings. Figure 10 shows the results for the lightness matches. The format of the figure is the same as in Figure 6.
The repeated-measures ANOVA showed a significant main effect of object color \( (F(3, 36) = 26.25, p < 0.001) \), and object material \( (F(3, 36) = 63.27, p < 0.001) \) on the mean distance between the lightness of the objects and the lightness of the matches. There also was a significant interaction between these two factors \( (F(9, 108) = 8.77, p < 0.001) \).

The magnitude of the shift away from the mean lightness of the objects was largest for the green \( (12.6 \pm 0.66) \) and blue objects \( (12.5 \pm 1.11) \). Post-hoc tests reveal that these shifts were significantly larger than the shifts found for the red \( (7.1 \pm 0.45) \) and yellow objects \( (7.0 \pm 0.44) \). Regarding the direction of shift, the matches for the red objects were shifted to higher lightness values, except for the candle. For the yellow objects, the matches were close to the mean measured values, except for the wool. All blue and the green objects except for the green candle were shifted to higher lightness values.

With respect to the material of the object, the shift of the matches away from the mean lightness of the objects was largest for the wool objects \( (17.2 \pm 1.00) \), followed by the papers \( (10.1 \pm 0.94) \). The shifts were smallest for the candles \( (5.3 \pm 0.48) \), and the sponges \( (6.6 \pm 0.40) \). Post-hoc test showed that there were no significant differences between the candles and the sponges. The most interesting features of the results for the lightness matches are the large deviations of the lightness matches from the mean lightness of the wool and paper objects. For those objects, the matches were shifted to distinctly higher values than the mean lightness of the objects. This can also be seen in Figure 11 where the mean lightness matches are plotted against the mean lightness of the objects. Most matches were shifted to higher lightness values compared to the mean lightness of the object. The red, yellow and green candle were close to the mean lightness while the blue candle was shifted to higher values. The mean lightness match for the yellow mug was shifted to higher lightness values compared to the mean object lightness while the mean lightness match for the blue mug was shifted to values lower than the mean object lightness. This is in accordance with the direction of the chroma shifts for the two mugs.

**Control stimuli**

To check whether observers discarded variations in the lightness of the objects that were due to the surface geometry of the objects, we compared the matches the observers made to the wool balls and crumpled papers with the lightness measured for the control stimuli, i.e., for the plain papers, and the tightly coiled wool (see Figure 4). Figure 12 shows the lightness of the objects used in the matching experiment (black), and the matches to these objects (red) that have already been shown in Figure 10 together with the lightness of the control objects (blue).

For all papers the mean lightness of the matches is closer to the mean lightness of the control objects than to the lightness of the objects to which the matches were actually made. However, for the green and blue paper the difference is still quite large. For the wool objects, the results are inconsistent. For the yellow wool, the mean lightness of the corresponding control object is close to the mean match made to the yellow wool ball. The mean lightness for the red, green, and blue wool control objects were close to or even below the mean lightness of the objects to which the matches were made. The reason why the lightness of the control object for the green wool ball is below the lightness of the wool ball might be the tightly

Figure 10. Lightness matches (red) and lightness of the objects derived from the spectroradiometric measurements (black) separately for the four object colors (from top to bottom: red, yellow, green, and blue). The filled black circles indicate the mean and the vertical line the range of the lightness of the objects. The horizontal red lines show the mean lightness match averaged across observers. Open red circles represent the individual observers’ matches show the mean lightness match averaged across repetitions and sessions. The bars show the mean distance between the mean lightness of the objects and the individual observers’ lightness matches. Filled bars indicate objects for which the mean lightness match was shifted to lower values compared to the mean lightness of the object. Error bars denote the standard error of the mean difference. Same materials are located in the same column.
Consequently, the spectroradiometric measurement became more homogeneous. The yellow wool object differed from the other wool balls, in that it was more loosely assembled compared to the relatively tightly coiled red, green, and blue wool balls, and therefore it contained more variations in lightness.

**Luminance estimation from images**

Since neither the mean lightness nor the range of lightness of the objects themselves nor the mean lightness of the control objects seem to be a good predictor for the observers’ lightness matches, we investigated whether there are lightness values on the objects surfaces that were not captured by the spectroradiometric measurements but might be closer to the observers’ lightness matches. To do this, we estimated the luminance of each pixel of the digital images of the objects following the procedure described in the Methods section.

Figures 13 and 14 compare the lightness derived from the spectroradiometric measurements with the lightness distribution estimated from the digital images of the wool and paper objects (third column). The means of the measured and estimated lightness distribution were close together. As expected, the distribution of the estimated lightness values has a wider range than the luminance distribution measured with the spectroradiometer. For the wool and paper objects, the mean lightness of the matches is always close to the highest lightness values of the estimated distribution. This might indicate that observers’ made their lightness matches to the brighter elements of the objects. In the first column of Figures 13 and 14 the
Figure 13. Lightness estimated from digital photographs of the wool objects. Left column: Grayscale images of the objects. Colored areas indicate the 15% of the pixels that were closest to the mean lightness match. Middle column: Histogram of the estimated lightness distribution. Right column: Comparison of the mean and range of lightness estimated from the digital images (blue), the mean and range of the lightness computed from the measurements with the spectroradiometer (black), the mean lightness match averaged across observers (red filled circles), and the mean matches of the observers averaged across repetitions and sessions (red open circles).
Figure 14. Lightness estimated from digital photographs of the paper objects. Left column: Grayscale images of the objects. Colored areas indicate the 15% of the pixels that were closest to the mean lightness match. Middle column: Histogram of the estimated lightness distribution. Right column: Comparison of the mean and range of lightness estimated from the digital images (blue), the mean and range of the lightness computed from the measurements with the spectroradiometer (black), the mean lightness match averaged across observers (red filled circles), and the mean matches of the observers averaged across repetitions and sessions (red open circles).
colored areas on the surface of the objects indicate the 15% of the pixels that were closest to the mean lightness match for the object.

In the images for the yellow wool and the papers it can be seen that the pixels closest to the mean lightness match were not the pixels with the highest lightness but were located around the lightest regions of the objects. The observers seem to have chosen the brightest pixels that did not belong to any of the highlights. This can be seen particularly clearly for the blue mug which had the most pronounced highlights of all our objects (Figure 15). For their matches, observers chose the brightest parts of the object that were directly neighboring the highlight regions.

The role of highlights was also investigated in the study of Motoyoshi et al. (2007). They found that increasing positive skew of the luminance histogram was correlated both with increasing specular intensity and with decreasing lightness ratings. Some of our objects contained distinct specular reflections: the candles except the blue one—and most strongly the two mugs. In accordance with Motoyoshi et al. (2007), the luminance histograms of these objects, determined from the luminance values estimated from the images, had a positive skew. However, a positive skew was also found for all other red objects. There was a tendency that higher skewness was correlated with smaller deviations from the mean lightness of the objects but the lightness matches made to those objects that appeared more glossy did not follow a common trend. The mean lightness matches for the yellow mug, the red, yellow, and green candle were close to the mean lightness of the objects with a slight trend towards lower lightness values for some of these objects while the blue mug showed a shift in the opposite direction towards higher lightness values (compare the filled bars in Figure 10).

Based on the lightness values estimated from the digital images, we computed a compensation index (CI) assuming that the maximum lightness—outside the highlight regions—is probably the best estimate of the real lightness of an object. The index is defined as

$$CI = (L_{\text{MATCH}} - L_{\text{MEAN}})/(L_{\text{MAX}} - L_{\text{MEAN}}),$$

where \(L_{\text{MATCH}}\) is given by an observer’s mean lightness match to an object, and \(L_{\text{MAX}}\) and \(L_{\text{MEAN}}\) refer to the maximum and the mean of the lightness distribution of the object based on the lightness values estimated from the digital images of the objects. For objects containing highlights (candles, mugs), the highest lightness value that did not belong to any of the highlights was chosen. The index indicates to which degree the observers compensated over shape and material. Figure 16 shows the median index collapsed over the different objects plotted separately for the 13 observers (left), and the median index collapsed over the observers plotted separately for the different stimuli (right). The left plot of Figure 16 shows that for most observers the median indices were similar. The overall mean is 0.63 indicating a shift of the matches towards the maximal lightness determined for the objects. The right part of Figure 16 shows that the observers were quite consistent in their matches for the different objects.

**Discussion**

We investigated the color appearance of real objects by asking observers to adjust a uniform disk presented on a monitor to have the same color appearance as a real object presented next to the disk. Our goal was, first, to investigate which chromaticity and luminance observers choose when presented with an object that varies in both,
and second, to determine whether the observers’ choice is influenced by the material of the object. In the following, we first discuss the results for the lightness matches, and then the results for the chroma and hue matches.

**Lightness matches**

Variations in lightness can be due to changes in material or due to variations in illumination caused by the geometry of the light source or the object and surface reflectance (Kingdom, 2008). The lightness matches showed larger variability compared to the chroma and hue matches. The deviations between the mean lightness matches and the mean lightness of the objects were significantly smaller for red and yellow objects than for green and blue objects. There was a significant difference in the magnitude of the shifts away from the mean object lightness between different materials. This difference seems not to be due to the materials but due to the different surface structures of these objects. The wool and the paper objects had corrugated surfaces while the surfaces of the candles, mugs and sponges were relatively smooth. Independent of the object color, the mean lightness matches for the objects with corrugated surfaces (wool and paper) were shifted to higher lightness values. These shifts were less pronounced for the yellow objects but surprisingly large for the green and blue objects.

If observers matched the lightness of the materials rather than the lightness of the objects, e.g., tried to discard the stronger variations in lightness due to the crumpling of the papers, then their lightness matches should be closer to the lightness of the material when it is presented without the variations introduced by the geometry of the objects. At least for the papers, we found that the mean lightness matches were closer to the mean lightness measured for the plain sheets of paper than to the crumpled papers to which the match was made, although for the green paper the difference was still large. For the wool balls, this effect was only found for the yellow wool ball while for the red, green, and blue wool balls the opposite effect can be seen. Here, it is important to note that the yellow wool object was different from the other wool balls and contained stronger variations in lightness. The spectroradiometric measurements of the objects provided only a coarse measure of the actual luminance distribution of the objects. To get a more accurate estimate of the range of luminance, we estimated the luminance from images of the objects. Although the method we employed was a very basic calibration of the camera, it seems to provide reasonable estimates. This is indicated by

---

**Figure 16.** Compensation index plotted separately for the 13 observers (left), and compensation index plotted separately for the 18 stimuli (right). The red horizontal lines indicate the median across objects (left) and across observers (right), respectively. The blue box indicates the lower and upper quartile of the distribution while the whiskers demarcate extreme values within 1.5 times the interquartile range. The red plus signs indicate values outside that range.
the fact that the mean luminance of the spectroradiometric measurements and the mean of the estimated luminance distributions were close. For the wool balls, we found that the mean lightness matches were close to the maximum lightness of the estimated values. As already hypothesized by Kuriki (2004), it is possible that the color matches to multicolored stimuli do not involve any kind of averaging but are determined by the most salient colors, e.g., the brightest elements of a stimulus. However, we also found evidence that the choice of bright regions is constrained to areas that are not part of specular reflections. For the objects used in our study, this is an excellent strategy to extract information about their reflectance properties since the highlights convey information about the illumination rather than about the object. The regions apparently chosen by the observers corresponded to the lightest parts of the objects that carried information about the object color. As proposed by Gilchrist et al. (1999) the highest luminance in a display might serve as an anchor or standard for judging the lightness of a surface.

Chroma and hue matches

Most of the chromatic variation of our stimuli was in chroma while there was only little variation in hue. One important source of variations in chroma are interreflections between different surfaces of an object (Bloj et al., 1999). Such variations were clearly present for the wool balls, the crumpled papers, and for the sponges with a coarser surface structure. Another cause for variations in chroma are variations in pigmentation as they can be seen in the case of the blue candle. Variations in hue mostly come from differences in pigmentation (Kingdom, 2008; Shevell & Kingdom, 2008). Since our objects did not vary in material there was only little variation in hue.

For the chroma matches, we found shifts to more saturated colors for some object colors (red and blue). Only the magnitude of the shifts for the red objects was significantly different from the shifts for the other colors. For yellow objects, the shifts were smallest but in the direction of less saturated colors. For green objects, shifts in both directions were found. The shifts were significantly larger for the objects made from wool and wax compared to sponge and paper objects but the direction of the shifts was heterogeneous and depended on the object color. We found a strong correlation between the mean lightness of the objects and the chroma shift of the matches. The matches for objects that had lower lightness were more saturated than the matches to objects with higher lightness.

The question how a global color percept is generated when presented with chromatically variegated textures was addressed by Kuriki (2004) and Sunaga and Yamashita (2007) using multicolored and two-color textures, respectively. In these studies two hypotheses were tested. The first hypothesis, the colorimetric averaging hypothesis, claims that a single color impression is determined by the colorimetric average of the chromaticities of the single elements that constitute the texture. The second hypothesis, the color appearance hypothesis, predicts that the global color impression is determined by the color appearance of the single elements, i.e., by their hue, saturation, and brightness.

The stimuli in Kuriki (2004) were isoluminant chromatic mosaic patterns that were sampled from circular chromatic distributions with different radii centered around different colors. Matches were made on a uniform patch. Although the patterns were isoluminant, in addition to the chromaticities the observers could also adjust the luminance to account for differences in brightness. The observers were explicitly asked to adjust the color of the matching pattern in such a way that it represented the average color of the textures. The results showed that with increasing range of the chromatic distribution of the mosaic patterns, the matches of the observers deviated more from the photometric average. The matches were near the color of the elements of the mosaics that had the highest saturation at least as long as the colors of the texture belonged roughly to one color category.

In the study by Sunaga and Yamashita (2007) textures consisting of two colors were used as stimuli. The colors of the textures had the same unique hue and were matched to have the same brightness while they differed in saturation. The matches to these textures were made on a homogeneously colored patch presented adjacent to the test stimuli. The colors of this patch were changed by adjustments along the three dimensions of the CIELUV (CIE, 1978) color space.

Sunaga and Yamashita (2007) also found that the chromaticities of the matches had a higher saturation than the colorimetric average. Since their stimuli consisted of elements that had equal brightness, they ruled out that these shifts to more saturated color were caused by the brighter appearance of more saturated colors (Burns, Elsner, Pokorny, & Smith, 1984; Burns, Smith, Pokorny, & Elsner, 1982). The authors argued that their results speak against the averaging hypotheses and support the color appearance hypothesis. With respect to the chroma matches, our results do not support the averaging hypothesis either.

The matches to the red objects were in general the matches that were most difficult for most observers. They frequently reported that they were not satisfied with these matches. This is also indicated by the fact that the matches for most red objects were close to or at the gamut border, especially for the candle and the sponge. Although based on the spectroradiometric measurements, the colors of all objects lay within the gamut of the monitor it is possible that the objects contained colors not captured by the spectroradiometer that were outside of the gamut of the monitor and to which the observers tried to establish a match.

Xiao and Brainard (2008) investigated the effect of material on color perception by measuring the influence of gloss on the color perception of rendered three dimensional objects. The observers were presented with stereo image pairs of a rendered room. In the first experiment, the
observers had to match the color of a sphere with diffuse reflection properties to the color of test spheres with five different glossiness values. The color of the test sphere was either purple or yellow-green. The observers changed the color of the match sphere by adjusting the color of the match sphere along the three dimensions of the CIELAB color space. They found that these matches became darker and more saturated with increasing glossiness of the test spheres. This effect was more pronounced for the purple sphere than for the green-yellow sphere. These results do not support the averaging hypothesis because the computed LMS average of the test spheres became lighter and less saturated with increasing glossiness. Overall, the effect of material on the color matches was small.

We found no easily interpretable effect of material on the saturation matches. While there was a significant difference in the magnitude of the matches for the wool balls and candles on the one hand and the sponges and papers on the other hand, it is not clear what features of the objects caused the differences between these two groups. Moreover, the directions of the shifts were not the same within these groups. Given that the candles—with the exception of the blue candle which had a rougher surface and wider variation in saturation than the other candles—and the mugs were the most glossy objects among our stimuli, the direction of the shift away from the mean chroma values of the objects seems to depend more on the object color than on the glossiness of the objects, e.g., the mean chroma match for the yellow mug was shifted to lower chroma values like all yellow objects, while the chroma match for the blue mug was shifted to higher chroma values in the same way as all other blue objects.

Regarding the hue matches, Sunaga and Yamashita (2007) found a shift of the color matches towards the unique loci line. In our experiment, hue matches were quite accurate. There was no systematic shift towards the unique hues. The mean hue angles of the red and yellow objects were close to the respective unique hue angle. The mean hue matches for these objects were close to the mean hue of the objects. The green objects covered a wider range of hue angles, approximately centered around the unique angle for green. The hue matches showed slight shifts to larger hue angles compared to the hue angles of the objects. The mean hues of the blue objects clustered around a hue angle of 270°, and were thus further away from their unique hue angle than the other objects. However, the matches for the blue objects also showed a slight shift towards larger hue angles.

Most of the chroma and lightness shifts away from the mean object chroma and lightness, respectively, correspond to ΔE differences that are around 10 or larger. ΔE = 1 is generally considered to represent a threshold difference under optimal conditions. Thus, the differences between the mean object color and the match color would be readily visible for most objects if both the mean object color and the match color were presented as homogeneous surfaces.

### Task

We gave our observers very little instructions on how to do the task. The important role of the instruction in such tasks is well known from studies on color constancy (e.g., Arend & Reeves, 1986; Bloj & Hurlbert, 2002; Bäuml, 1999). Here, we did not want to put additional constraints on the observers’ matches. In general, observers did the experiment without asking for more detailed instructions. They seemed to have had no problem with the task, although they reported that they were not satisfied with some matches. Some observers indicated that they were aware of the inherent problem of this matching task, i.e., to reduce the chromatic and luminance distribution of the objects to a single color. It is quite possible that different observers employed different strategies. This might make it problematic to average across the matches of the observers.

Another source of error might be in the asymmetric matching task itself in which a real object has to be matched on a monitor screen. Granzier, Brenner, and Smeets (2009) compared the accuracy of color matches for different combinations of stimulus presentation (presentation as real papers, or as colored patches on a monitor) and matching procedure (adjustment on a monitor, or selection of real colored samples). They found that the deviations between reference color and match were largest in the condition in which the reference colors were presented as real papers and the matches were made on a monitor.

The usage of real, natural or man-made stimuli has always the disadvantage that it allows only a limited control of the properties of the stimuli. Certainly, not all the variations between our stimuli were captured by the factors color and material. Objects in one material group did not only vary in color, e.g., the threads of the different wool balls differed in their thickness; the papers were not crumpled in the same way, and the sponges had different surface structures. On the other hand, while using rendered objects as stimuli allows systematic variation of certain features, this usually comes at the cost of reducing the complexity of the stimuli, and this might also be a reduction of potentially important cues (Koenderink, 1999; Robilotto & Zaidi, 2004). Therefore, we think that it might be a good starting point for the investigation of the color perception of real objects to actually start with these objects, and to derive further research questions based on the outcome of these experiments which then can be investigated in a more systematic way.

### Conclusion

Using a color matching task, we investigated the color perception of real objects presented under controlled illumination conditions. Our results indicate that the mean
color of an object is not a good predictor of its color appearance. We found evidence indicating that the observers’ lightness matches were determined by the brighter colors of the objects. Differences in material had only little effect on the matches, indicating that observers are capable of discounting shape and texture variations.

**Acknowledgments**

This research was supported by the German Science Foundation grant Ge 879/9. We thank the two anonymous reviewers for their useful suggestions, and Roland Fleming for helpful discussions and comments. We also thank our observers, and Lena Vorobyova for help with the experiments and spectroradiometric measurements. We would like to thank Johanna Gegenfurtner for making the yellow woolen ball.

Commercial relationships: none.
Corresponding author: Martin Giesel.
Email: mgiesel@sunyopt.edu.
Address: SUNY College of Optometry, Graduate Center for Vision Research, 33 West 42nd St., New York, NY 10036, USA.

**References**


