

# The effects of surface gloss and roughness on color constancy for real 3-D objects

Jeroen J. M. Granzier

Department of Psychology, Justus Liebig University,  
Giessen, Germany



Romain Vergne

Department of Psychology, Justus Liebig University,  
Giessen, Germany, Grenoble University, Inria, LJK



Karl R. Gegenfurtner

Department of Psychology, Justus Liebig University,  
Giessen, Germany



**Color constancy denotes the phenomenon that the appearance of an object remains fairly stable under changes in illumination and background color. Most of what we know about color constancy comes from experiments using flat, matte surfaces placed on a single plane under diffuse illumination simulated on a computer monitor. Here we investigate whether material properties (glossiness and roughness) have an effect on color constancy for real objects. Subjects matched the color and brightness of cylinders (painted red, green, or blue) illuminated by simulated daylight (D65) or by a reddish light with a Munsell color book illuminated by a tungsten lamp. The cylinders were either glossy or matte and either smooth or rough. The object was placed in front of a black background or a colored checkerboard. We found that color constancy was significantly higher for the glossy objects compared to the matte objects, and higher for the smooth objects compared to the rough objects. This was independent of the background. We conclude that material properties like glossiness and roughness can have significant effects on color constancy.**

## Introduction

Color vision helps us to reliably and quickly identify objects within a scene (Bramão, Reis, Petersson, & Faísca, 2011; Tanaka, Weiskopf, & Williams, 2001), both for conditions in which the color of the object is diagnostic for that object (i.e., a yellow banana; Tanaka & Presnell, 1999) as for conditions for which color information is not diagnostic for a particular object, for example a yellow sock (Biederman & Ju,

1988; Bramão, Faísca, Petersson, & Reis, 2010; Gegenfurtner & Rieger, 2000; Uttl, Graf, & Santacruz, 2006; Wurm, Legge, Isenberg, & Luebker, 1993). For perceived surface color to be a useful guide to object identity, it should highly correlate with surface reflectance. This is not easy to achieve because the sensory signal that reaches the eye confounds surface reflectance with the illuminant. The effect of the illumination can be so extreme that the light reaching the eye from a “blue” paper in tungsten light can lead to the same photoreceptor activations as that from a “yellow” paper in sunlight (Jameson, 1985). However, under normal visual circumstances, we are perfectly aware whether we are confronted with either a blue or yellow paper (Granzier, Brenner, & Smeets, 2009a). The ability of the visual system to maintain a stable perception of surface color across changes in illumination (and other viewing conditions) is called color constancy. Without this ability, the color appearance of objects would possibly change from moment to moment, making color information fairly useless for object recognition. The degree to which a human observer is color constant depends on many factors, including both low and high level (cognitive) factors (for a recent overview see Smithson, 2005). Lots of groundbreaking work has been done in the last 30 years with the use of flat, matte surfaces simulated on computer monitors (see for an overview Foster, 2011). The degree of color constancy observed in these experiments was quite variable, but generally far from perfect.

Recently, there has been an increased interest in color constancy for complex, real three-dimensional scenes. These studies show that color constancy is

Citation: Granzier, J. J. M., Vergne, R., & Gegenfurtner, K. R. (2014). The effects of surface gloss and roughness on color constancy for real 3-D objects. *Journal of Vision*, 14(2):16, 1–20. <http://www.journalofvision.org/content/14/2/16>, doi: 10.1167/14.2.16.

better in more complex environments compared to simple two-dimensional patterns (e.g., de Almeida, Fiadeiro, & Nascimento, 2004; Bloj & Hurlbert, 2002; Granzier, Brenner, & Smeets, 2009a, 2009b; Brenner, Granzier, & Smeets, 2011; Hansen, Walter, & Gegenfurtner, 2007; Hedrich, Bloj, & Ruppertsberg, 2009; Kraft & Brainard, 1999; Olkkonen, Witzel, Hansen, & Gegenfurtner, 2010; Zaidi & Bostic, 2008). This suggests that a natural environment is important for color constancy. It can also explain why color constancy as investigated in the earlier studies (e.g., Arend & Reeves, 1986; Blackwell & Buchsbaum, 1988; Valberg & Lange-Malecki, 1990) using “Mondrian” stimuli was sometimes quite low. Regrettably, experiments using real 3D objects are rare in color constancy research because the stimuli are difficult to manipulate and it is much more time consuming to perform these experiments, as the experiment is not under automatic control. Therefore, we know very little about how humans achieve color constancy under more realistic viewing conditions.

## Effects of material properties on color perception

Real-world scenes typically consist of objects made of different materials. Our natural environment consists of a multitude of materials and they potentially contribute to estimating the illuminant and supporting a stable appearance of objects. Different material properties like surface roughness and gloss can potentially provide us with cues about the illumination to stabilize our color percept, and of course, the ultimate goal of color perception research is to understand how we perceive the colors of the objects that surround us. These are three-dimensional and made of different materials (such as fur, stone, glass, metal, etc.). The question we want to address here is: How does the color of an object interact with its material properties (like gloss and roughness) under changes in the chromaticity of the illumination?

There are few previous studies on this topic (see Maloney & Brainard, 2010, for a recent overview). Xiao and Brainard (2008) assessed how the presence of specular highlights affects the color appearance of three-dimensional objects rendered on a computer screen. They showed that the visual system is capable of stabilizing color appearance with respect to gloss. Giesel and Gegenfurtner (2010) systematically investigated color perception for real objects made of different materials varying in roughness and gloss from smooth and glossy to matte and corrugated. They show that hue is perceived quite stable across their manipulations, but that saturation and lightness judgments are

systematically affected. However, in their study the illumination was constant so color constancy was not investigated.

Olkkonen and Brainard (2010) studied whether the joint effects of illumination and an object’s shape on the perception of surface reflectance can be predicted from the individual effects in a straightforward manner. They found large interactions between illumination and object shape in their effects on perceived gloss.

Most recently, Xiao, Hurst, MacIntyre, and Brainard (2012) used an achromatic matching task with matte disks, matte spheres, and glossy spheres. In all cases, the test stimuli were viewed in a stereoscopically displayed graphic simulations of three-dimensional scenes, and the authors varied the scene illuminant. Conditions were studied in which all cues were consistent with the simulated illuminant change (consistent-cue conditions) and where local contrast was silenced as a cue (reduced-cue conditions). Color constancy was similar for the three test object types. There was, however, a reliable interaction between test object type and cue condition. In the consistent-cue conditions, constancy tended to be best for the matte disks, while in the reduced-cue conditions constancy was best for the spheres. The authors conclude that the presence of this interaction between an object shape and the presence of information with respect to the illumination presents an important challenge for theorists who seek to generalize models that account for constancy for flat tests to the more general case of three-dimensional objects.

In summary, little information is available with respect to the effects that different materials have on our color percept. Even less information is available when viewing real three-dimensional objects under more complex illumination conditions. In the current study, we focus on two material properties; surface gloss and surface roughness. In contrast to most of the studies mentioned above, we used real physical stimuli in our experiments.

## Experiment 1: The effects of surface gloss on color constancy

### Methods

#### Subjects

Twelve subjects took part in the experiment. All of them (eight females and four males) were students at Giessen University. They had normal color vision as tested with Ishihara color plates (Ishihara, 1969). The subjects were naïve as to the purpose of the experiment. Informed consent was given by all subjects according to the Declaration of Helsinki (World Medical Associa-

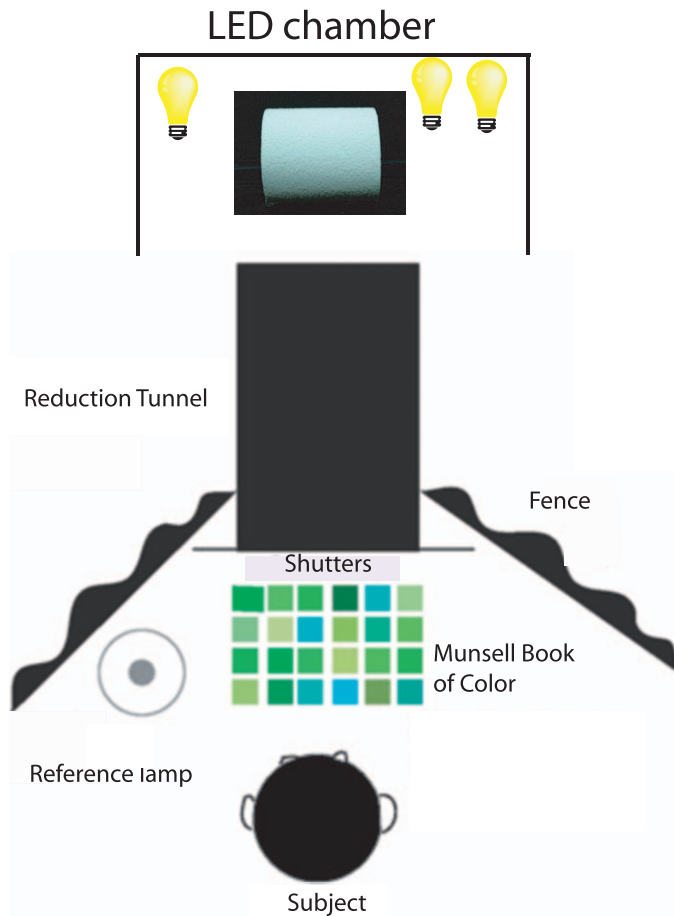


Figure 1. Schematic overview of the experimental set-up. Subjects selected the Munsell sample that best matched the color of the cylinder placed into the centre of the LED chamber.

tion, 2004). The experiments were approved by the local ethics committee.

### Illumination

During the experiment, one of the six objects was placed in a little chamber (LED color viewing light; Just NormLicht, Weilheim/Teck, Germany). The illumination of the chamber was under automatic control by a Dell T3500 computer (Dell, Plano, TX). A metamer of the illuminant D65 was used to illuminate the cylinders (see Figure 3). D65 is intended to represent average daylight and has a correlated color temperature of approximately 6500 K. Subjects' color constancy was tested with a second, artificial reddish illuminant (see Figure 3). The third illuminant was a neutral colored tungsten lamp (see Figure 1, identical to the lamp illuminating the Munsell color book [Munsell Color, Grand Rapids, MI, see <http://munsell.com>]). The lamp illuminating the Munsell book will be referred to as the *reference lamp*. The 1931 CIE xy coordinates of the light from the three lamps as

measured with a Photo Research PR-650 spectroradiometer (Photo Research, Inc., Chatsworth, CA) were (0.311, 0.334), (0.397, 0.294) and (0.458, 0.412) for the D65, the reddish and the reference lamp, respectively. These values were measured as the reflectance of a white reflectance standard in the middle of the experimental scene. The lamps were presented one at a time in random order. The luminance was identical for all three lamps and was  $75 \text{ cd/m}^2$  measured as the light reflected from a white reflectance standard at the center of the experimental scene. The subjects could not see the lamps and did not know how many lamps there were, or their colors. The scene was never illuminated by more than one of the three lamps. Calibration measurements of the reflectance standard were made during the course of the experiments to verify that the luminance and the chromaticity of the lamps remained constant.

### Scenes

We used three-dimensional objects and real lamps to create optimal circumstances for estimating the illuminant's color (see Figure 2). The scene was presented in front of the subjects at a distance of 250 cm with a width of 64 cm and a height of 50 cm. We used two different backgrounds to investigate the effect of local and global color contrast on color constancy. In half of the trials we used a scene with a completely black background, reflecting only about 1% of the light (see Figure 2). We hypothesized that the effect of the different material properties might be quite weak and could be overridden by other potential cues present in the scene (e.g., color contrast or the average color of the scene). We therefore measured color constancy in scenes where the most important cue for estimating the illuminant's color would be the interaction between the illumination and the objects themselves with other cues for color constancy mostly silenced. In the other half of the trials, we placed a multicolored checkerboard (see Figure 4) behind the cylinders. The colored papers of the checkerboard would reasonably provide subjects with more information concerning the illumination (i.e., the white paper in the checkerboard would reflect most of the color of the lamp, thereby providing direct information with respect to the illumination). Moreover, as the cylinders were placed in front of the checkerboard, the local color contrast also provides direct information about the color of the object.

Figure 4 shows the checkerboard that we used during the experiments. Each color is numbered. The CIE xyY values of the checkerboard colors under the reference lamp from the number 1 onwards are: (0.448; 0.398;  $31 \text{ cd/m}^2$ ), (0.455; 0.413;  $12 \text{ cd/m}^2$ ), (0.350; 0.421;  $7 \text{ cd/m}^2$ ), (0.503; 0.443;  $30 \text{ cd/m}^2$ ), (0.368; 0.521;  $7 \text{ cd/m}^2$ ), (0.444; 0.476;  $4 \text{ cd/m}^2$ ), (0.646; 0.330;  $7 \text{ cd/m}^2$ ),

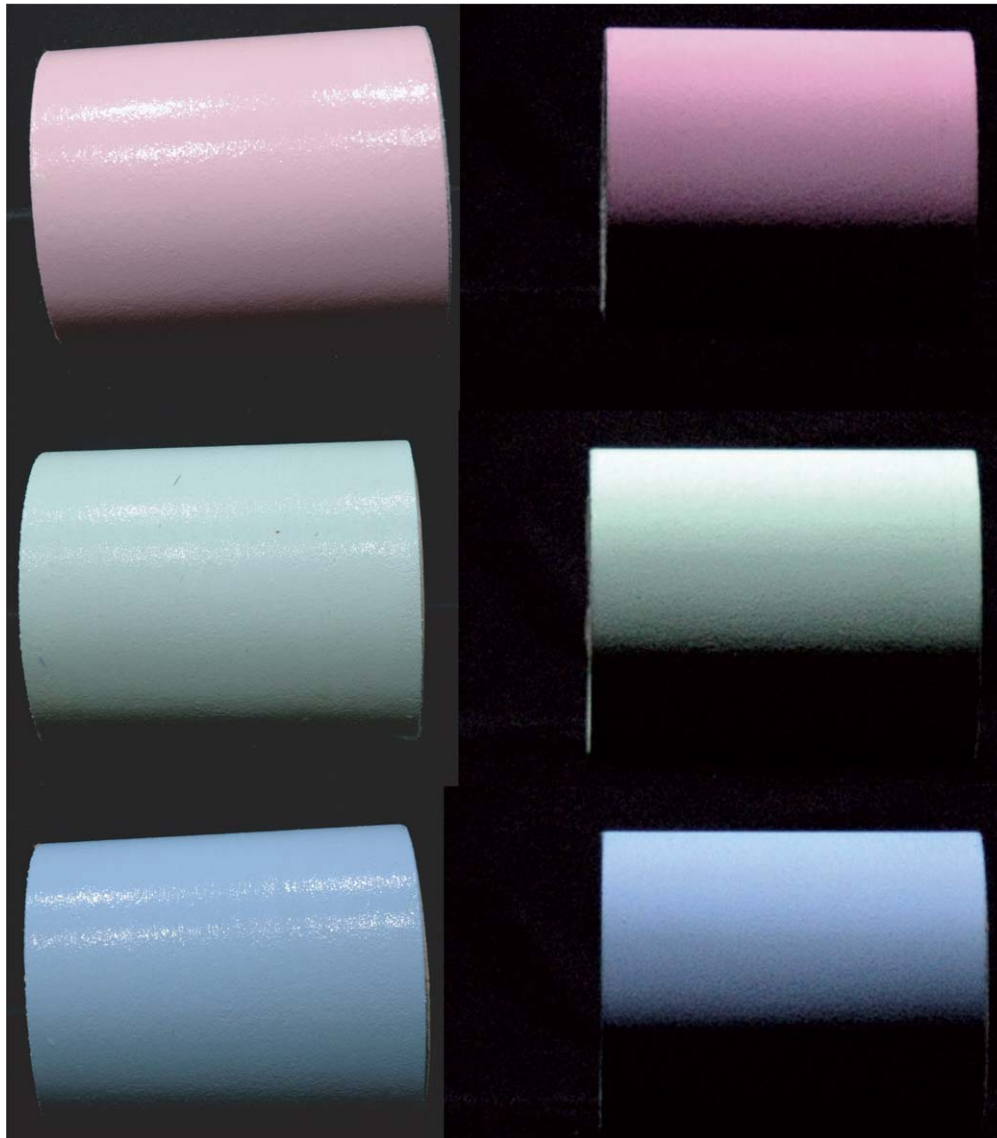


Figure 2. The stimuli used for Experiment 1 are shown. The picture on the left shows the glossy cylinders from each color group while the cylinders on the right shows the matte versions. The photographs were taken by illuminating the objects under lamp D65.

(0.599; 0.381; 14  $\text{cd/m}^2$ ), (0.555; 0.389; 3  $\text{cd/m}^2$ ), (0.462; 0.410; 8.5  $\text{cd/m}^2$ ), (0.540; 0.433; 18  $\text{cd/m}^2$ ), (0.628; 0.334; 2  $\text{cd/m}^2$ ), (0.461; 0.413; 17  $\text{cd/m}^2$ ), (0.459; 0.415; 3.5  $\text{cd/m}^2$ ), (0.357; 0.359, 1.5  $\text{cd/m}^2$ ). The sizes of the checks were 6 cm  $\times$  6 cm. The black rim surrounding the checks was 2 cm wide.

### Test objects

The test objects were six cylinders (10 cm wide  $\times$  8 cm high) made out of hard carton (see Figure 2). Sandpapers of different grits were purchased in a local hardware store and glued onto the cylinders. By using sandpapers of different roughness we could vary the roughness of the stimuli (see Experiment 2). For Experiment 1 we used very smooth sandpaper, since we

were only interested in the effects of surface gloss on constancy. The cylinders with the glued sandpaper were subsequently painted in either a matte greenish, a matte bluish, or a matte reddish color using Design Color Mix paint (Design Color Mix, Ruhl farben, Ober-Ramstadt, Germany). We used a glossy paint with the same colors for the other three cylinders, resulting in a total of six objects. To exaggerate the visual appearance of gloss for the glossy objects, we added transparent oil (regular carpet protection oil) on top of the colored paint (see Figure 2).

We have chosen to use these particular shades of blue, green, and red as they are quite unsaturated (see Figure 2), which introduces large shifts in color appearance by changing the illuminants' color. The 1931 CIE xy color coordinates of the light reflected by the blue, green, and red cylinders under the reference

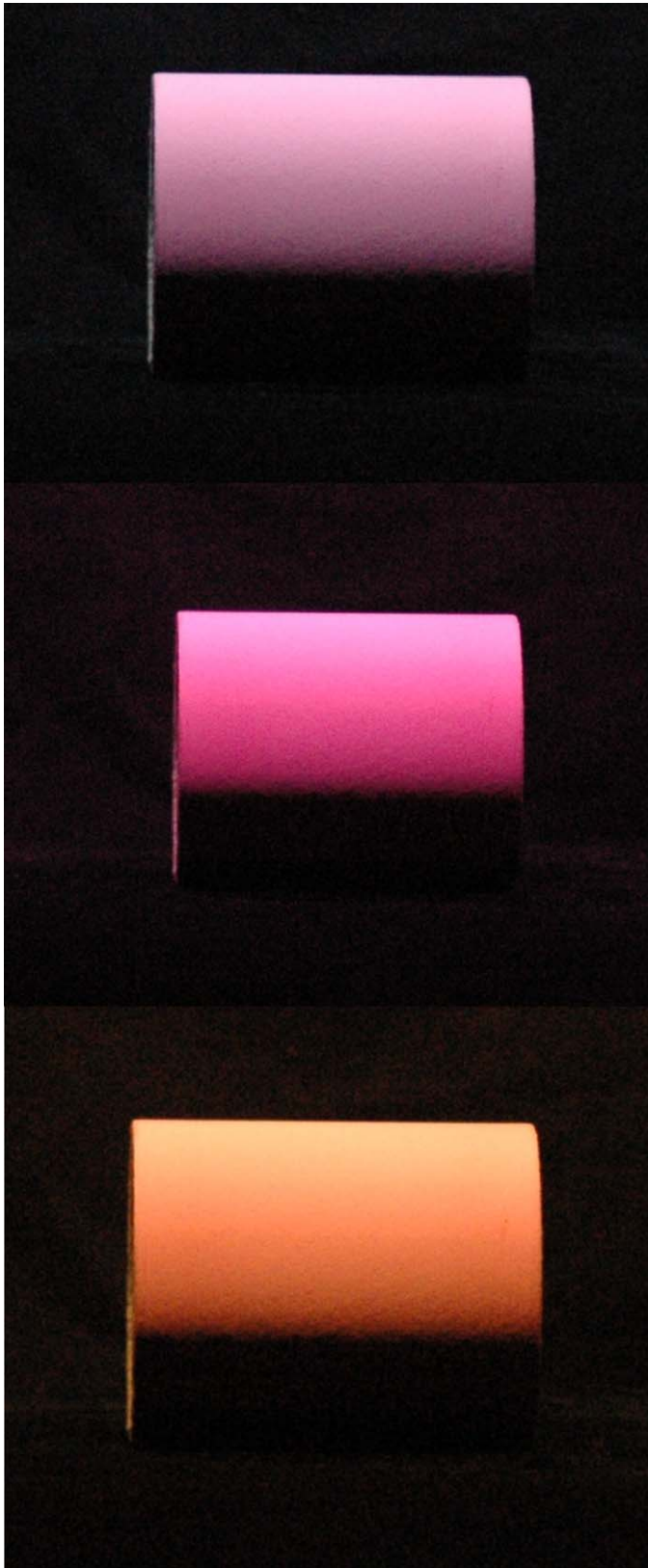


Figure 3. The three photographs show the same object (red cylinder) illuminated by the three lamps used during the experiment.

lamp (the one illuminating the Munsell book) were (0.405; 0.4), (0.437; 0.44), and (0.508; 0.386), respectively. One of the cylinders was placed in the middle of the scene, always at the same location and with the same orientation with respect to the observer. Please note that a given color (i.e., red) painted with either the glossy or the matte paint resulted in equal chromaticities (see Figure 2), so that by comparing the constancy results for the matte red cylinder with the glossy red cylinder, only the effect of gloss is studied. However, obviously the specular part of the glossy objects reflected in part the chromaticity of the illumination so if subjects would use the chromaticity of the specular reflection of the cylinder to match the object's color, subjects would obtain low amounts of color constancy. The six cylinders were each illuminated by each of the three lamps, giving a total of 18 combinations of surface and illumination. Each combination was presented twice, once with the black and once with the checkerboard background, resulting in a total of 36 matches for each subject. The black and checkerboard backgrounds were presented in different blocks, with the order of blocks randomized between subjects. Changing the background between the trials would have been prohibitively time consuming.

### **Procedure**

We took precautions to avoid providing observers with any information about the illuminant other than that available by looking through the reduction tunnel (see below). First, observers were seated in front of the reduction tunnel (see Figure 1), so that essentially no light from the experimental chamber scattered to their eyes from the walls or floor of the experimental room. Second, observers entered and left the experimental room under incandescent room illumination and with the experimental illuminants turned off. When the observers were first seated, a shutter prevented them from seeing the contents of the experimental chamber. Subsequently, the room lights were turned off and the experimental illuminant was set. The shutter was opened so that the observer could view the stimulus and a lamp illuminating the Munsell book (see below) was switched on. This way, the subjects could adapt while instructions were given to them. This took about five minutes. Once the instructions were given and subjects were confident that they understood the task, one of the cylinders was placed in the middle of the scene by the experimenter and one of the experimental lamps was switched on. Subjects viewed the scene through a black reduction tunnel (see Figure 1) with a built-in shutter. Once the experimenter had placed a cylinder into the scene and one of the lamps was switched on, subjects were instructed to raise the shutter and make their color match. In the first session

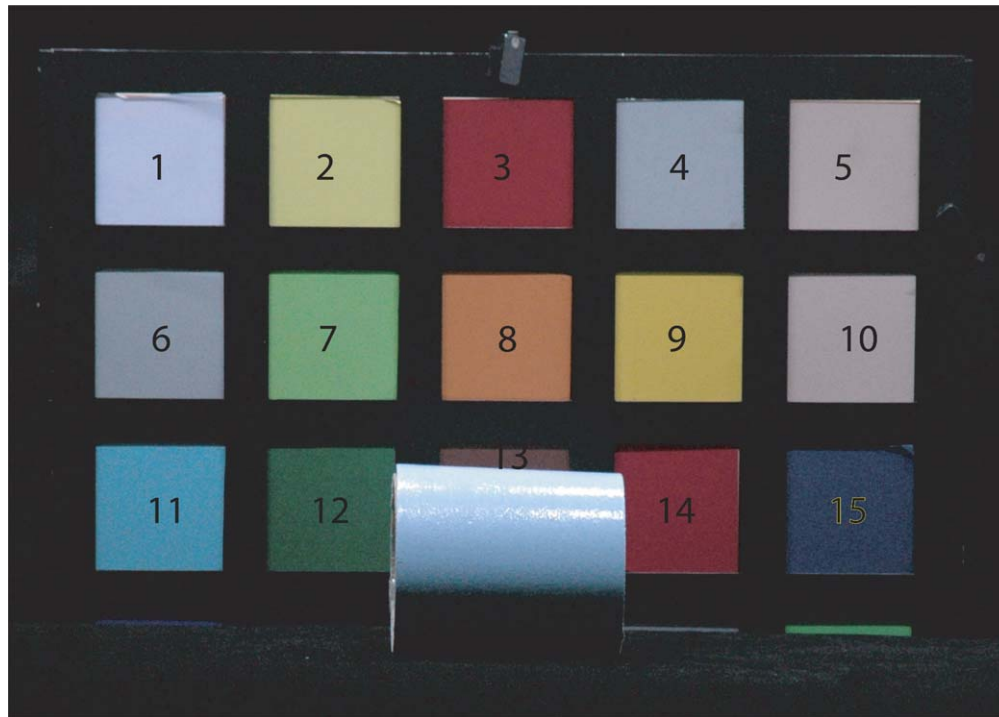


Figure 4. The blue glossy cylinder with the color checkerboard as background under D65 illumination is shown. The checkerboard was used in half of the trials to investigate the influence of local and global color contrast on color constancy. The checks are numbered for identifying each single check more easily. See the main text for references for the CIE color values for each color check.

they were encouraged to familiarize themselves with the contents of the Munsell book. They read out the Munsell number of the color that they thought matched the color of the cylinder in the scene. The experimenter confirmed whether this was indeed the color that the subject intended. If this was the case, the experimenter wrote down the number of the Munsell color and the subject was instructed to lower the shutter. Shutters were lowered so that subjects were unable to use the colors of the clothing of the experimenter as a reference as he changed the cylinder in the experimental scene and to prevent subjects from seeing the chromaticity of the illumination changing and to use this information in obtaining color constancy. Once the experimenter had changed the object and had switched on a new lamp, the subject could raise the shutters again to make his or her next color match. The order of the lamps and cylinders was randomized. Subjects could take as long as they wanted to make the color match. Each session took about 90 min for each subject.

### Color matching

Subjects selected the sample of the Munsell book that best matched the surface of one of six cylinders. The number of cylinders was unknown to the subjects and only one was visible at a time (see Figure 2). The whole Munsell color space was available for the subject

to choose from. Observers could take as long as they wanted to make their color match and they were not explicitly instructed on how they should make their color match or which strategy would be good for them to use. No feedback was given with respect to the quality of their color matches. Subjects were subsequently asked to rate the quality of their color matches on a scale from 1 (extremely poor color match) to 7 (perfect color match). We pursued two strategies for data analysis. First, we will analyze only the color matches with ratings between the range of 4 and 7, since consideration of the poor quality matches might introduce a bias. We reasoned that if observers state that the quality of the color match was poor, these color matches would not say much about observers' color perception per se as there is no good agreement between the chosen Munsell color and the way they perceive the objects' color. For Experiment 1, 85% of the trials were of high quality while for Experiment 2, 90% of the color matches were high quality. This shows that only a small part of the trials was excluded for data analysis. Table 1 (Experiment 1) and Table 3 (Experiment 2) provide an overview of the number of trials (in %) per condition that were used for data analysis. Second, to have an indication of whether the exclusion of the low quality matches (between '1-3') might have an effect on the color constancy results, we additionally analyzed the complete data set.

	Good trials (%) matte Munsell book	Good trials (%) glossy Munsell book
Blue glossy object	92%	94%
Blue matte object	85%	88%
Green glossy object	90%	83%
Green matte object	81%	83%
Red glossy object	83%	90%
Red matte object	73%	77%

Table 1. The percentage of trials that were used for data analysis, shown separately for each object (rows) and version of the Munsell color book (columns).

We also performed the experiments with both versions of the Munsell book: the matte version, with 1,270 different colored matte papers, and the glossy version, with 1,600 glossy papers of different colors. We did this for two reasons. First, we wanted to explore whether subjects are better able to match glossy objects with the glossy version of the Munsell book than the matte version, and vice versa. Second, we wanted to have a measure for the consistency of our observers when performing this task (Granzier & Gegenfurtner, 2012). Therefore, we performed the experiment twice; in one session subjects used the matte version of the Munsell book throughout the session, while in the other session they used the glossy version. The order in which subjects used the two versions of the Munsell book was randomized. As indicated above, observers could find a suitable color match in most trials.

### Data analysis

The first step was to measure the 1931 CIE xy color coordinates of each of the chosen Munsell samples when illuminated by the reference lamp. If subjects had perfect color constancy, they would choose the same color chip to match the cylinders independently of the illumination (D65 illumination and the reddish lamp). If subjects would have poor amounts of color constancy, they would choose Munsell chips with different colors when the same cylinder is viewed under different illuminations. A complete absence of color constancy would mean that the difference (a vector in CIE color space) between subjects' color matches corresponds to the difference of the same object measured under the two illuminations (shift in chromaticity caused by the illumination). A negative color constancy percentage would mean that the difference in CIE color space between subjects' color matches when viewing the same object under the two different lamps (i.e., the *perceptual shift*) is even larger than the measured difference between the CIE values of the same object under the two illuminations (i.e., the

*physical color shift*). This principle is explained in Figure 5. The filled black diamond represents the CIE coordinates for the color match when seeing a given cylinder under lamp 1 (D65). The open diamond represents the CIE color values for the color match when the same cylinder is viewed under lamp 2 (the reddish illuminant). To obtain a measure of color constancy, we projected these coordinates onto the line that is defined by the effect of the illumination change. Only the component of the difference in that direction is considered. We subsequently computed the difference between the projected values of the color matches (represented as the perceptual shift). This difference is the shift in color perception for this particular subject caused by the change in illumination (i.e., failure in color constancy). As the open and filled diamonds do not lie on top of each other we can already conclude that this subject did not have perfect color constancy for this object. We subsequently measured the *shift caused by the lamps*. We did this by computing the difference (in CIE color space) between the light reaching the subjects' eyes under lamp 1 (solid triangle) and lamp 2 (open triangle). The larger this difference is, the larger the illumination changes the light reaching the eyes. A color constancy index *CC* was then computed by the following equation:

$$CC = 1 - \frac{\text{Perceptual shift}}{\text{Illumination shift}} \quad (1)$$

We then multiplied the resulting quantity by 100 to obtain percentages. Thus, if the perceptual shift is equal to the shift caused by the illumination, the result would be zero. An index of 1 would mean that the perceptual shift is zero. This would indicate perfect color constancy.

To test whether our results depended on the choice of the color space used, we also calculated the color constancy indices in the perceptually more uniform color space  $u^*v^*$ . Since all statistical tests yielded equivalent results when performed in both color spaces, and because the overall correlation between constancy indices in both spaces was high ( $\alpha = 0.981$ ), we report results only for CIE xy space here.

Repeated-measures analyses of variance were used to evaluate the influence of the factor *reflectance* (containing two levels: glossy objects vs. matte objects), the factor *color* (having three levels: red objects, green objects, and blue objects), the factor *Munsell book* (containing two levels: The glossy Munsell book vs. the matte Munsell book), and the factor *background* (containing two levels: checkerboard background vs. black background) on the amount of color constancy. Paired samples *t* tests were used to determine whether color constancy was significantly different for the glossy objects compared to the matte objects (or for the smooth vs. the rough object explored in Experiment 2).

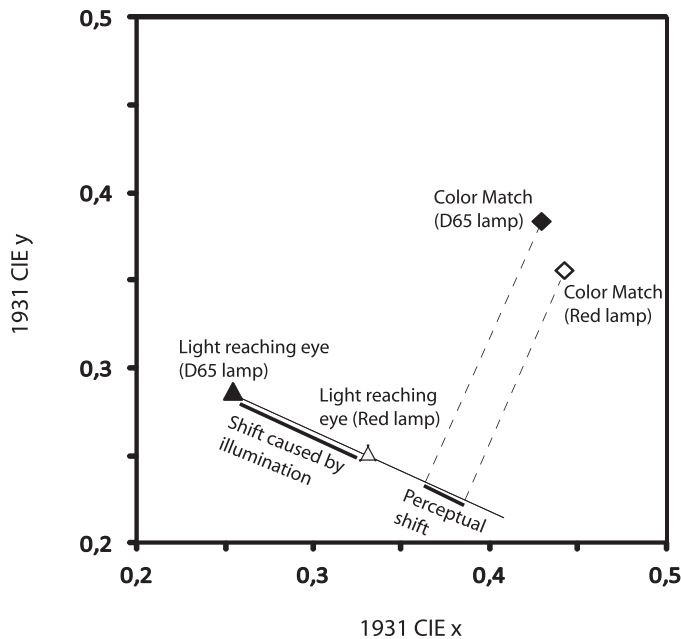


Figure 5. A schematic representation of how the color constancy index was computed. Part of the CIE color space is shown. Solid and open diamonds show the color matches for a given object under the two illuminations used. We projected these CIE values on a line (represented by the dashed lines in the figure) that lies in the direction of the shift caused by the illumination. The distance between the projected values represents the shift in color perception caused by the illumination (failure of color constancy). We then computed the distance in CIE color space between the measured values of the same object under the two different illuminations (represented by the solid and open triangles). The distance between the open and solid triangles represents the “physical” shift (shift in the light reaching the eyes) caused by the two experimental illuminations. If the distance of the perceptual shift and the distance in CIE color space caused by the shift of the lamps is the same, color constancy is absent (0%). If however subjects choose the same Munsell chip when the same cylinder is presented under different illuminations, color constancy is perfect (a color constancy index of 100%).

To study the effects that different material properties can have on constancy performance, for each color of cylinder we calculated the difference in color constancy (in %) between the performances obtained for the glossy object and the matte object. Positive difference scores indicate superior performance with the glossy objects. Negative difference scores indicate better constancy for the matte objects.

Finally, a categorical color constancy index was used (see Olkkonen et al., 2010; Troost & de Weert, 1991) to have a second test of whether the potential effects of the material properties depended on the way color constancy was computed.

This index is independent of any metric properties of color spaces and simply reflects whether the same Munsell hue category was chosen for a given object under the two different illuminations. The categorical color constancy index was set to 1 for a particular observer and object, if the object was given the same hue category under the two illuminants. Otherwise, the index was set to zero. A  $\chi^2$  test was used to test for significant differences between the categorical color constancy indices for the glossy and matte objects.

## Results for exclusion of poor color matches

### Effects of background and Munsell color book

The major reason for running these experiments with two different backgrounds, black and checkerboard, was to obtain a range of color constancy performances to avoid potential floor and ceiling effects. Also, we wanted to see whether the background had any effects on subjects’ color constancy performance and whether the effects of material properties on color constancy would depend on the background. In order to analyze the influence of the background on color constancy performance, we separated the trials with a black background from trials with the checkerboard background and calculated the color constancy indices (across both versions of the Munsell book) for each subject, for each of the six objects used. A repeated measures analysis of variance showed that there was a significant difference in the amount of color constancy between the black background compared to the checkerboard background,  $F(1, 60) = 11.613$ ,  $p < 0.001$ . Subjects had an average color constancy of about 68% (averaged across the six different objects), with the checkerboard compared to a constancy percentage of 59% with the black background.

This result shows (as expected) that local and global color contrast did provide subjects extra information about the illumination. However, the crucial aspect is whether the two backgrounds produced differences in the effects of the glossy versus the matte cylinders on color constancy performance. In other words, if there exist differences in the amount of color constancy for glossy objects compared to matte objects, we are interested in whether these differences might be more pronounced when tested with a black background compared to when tested with a checkerboard. In order to test this, we computed difference scores for each subject for each color group (blue, green, and red), separately calculated for when subjects were tested with either a black background or when tested with a checkerboard. Repeated measures of analysis of variance revealed no significant difference,  $F(1, 28) = 0.813$ ,  $p = \text{ns}$ , between the difference scores (the effects of glossy objects on color constancy) between both backgrounds. These results show that the potential



effects of the material property gloss on color constancy are not significantly different between both backgrounds. As we only had one trial per condition (object  $\times$  illumination  $\times$  background  $\times$  version of the Munsell book) we will therefore group the data for both background conditions to increase the statistical power and test whether there are any significant differences between the glossy versus matte objects. Moreover, missing trials (because of lowly rated color matches) would make a repeated measures design including the factor “background” too complex.

A second factor that we wanted to investigate is whether the use of either the glossy or matte Munsell book made any difference in subjects’ color constancy performance. In order to study this, we analyzed the data separately for both versions of the Munsell book and also analyzed the data separately for the glossy and the matte cylinders. Of the trials, 85% could be included in the data analysis (were rated between 4 and 7) when observers were matching the cylinders with the glossy version of the Munsell book. When the matte version of the Munsell book was used for making the color matches, 84% of the data could be used.

Color constancy was higher on average when objects were matched with the matte version of the Munsell book compared to when matched with the glossy version (67% vs. 54%, respectively). This effect was independent of which reflectance (matte or glossy object) was tested. These results can also be observed in Figure 6, which shows the average color constancy indices shown separately for each combination of Munsell book (matte version vs. glossy version) and object (glossy vs. matte).

Repeated measures of analysis of variance revealed a significant main effect of the factor Munsell book on color constancy,  $F(1, 49) = 174.254$ ,  $p < 0.01$ . A paired  $t$  test revealed that color constancy was significantly higher when the matte version of the Munsell book was used compared to when the glossy version was used,  $t_{12} = -2.904$ ,  $p < 0.01$ . This latter result might be explained by the fact that fewer colors are contained in the matte Munsell book. Therefore, the chance that observers chose a very similar color sample when the same object is viewed under different illuminations is higher for the matte version of the Munsell book. The correlation between the amount of color constancy between the session with the matte and glossy versions of the Munsell book was moderate at 0.30. This could probably be explained by the fact that the color samples of the glossy Munsell book are spread out in two separate books while the color samples of the matte Munsell book are collected in a single book. Perhaps observers did not always look in both books to find an appropriate sample, which might have caused the differences in the sample chosen between both sessions. Higher correlations were obtained for the blue

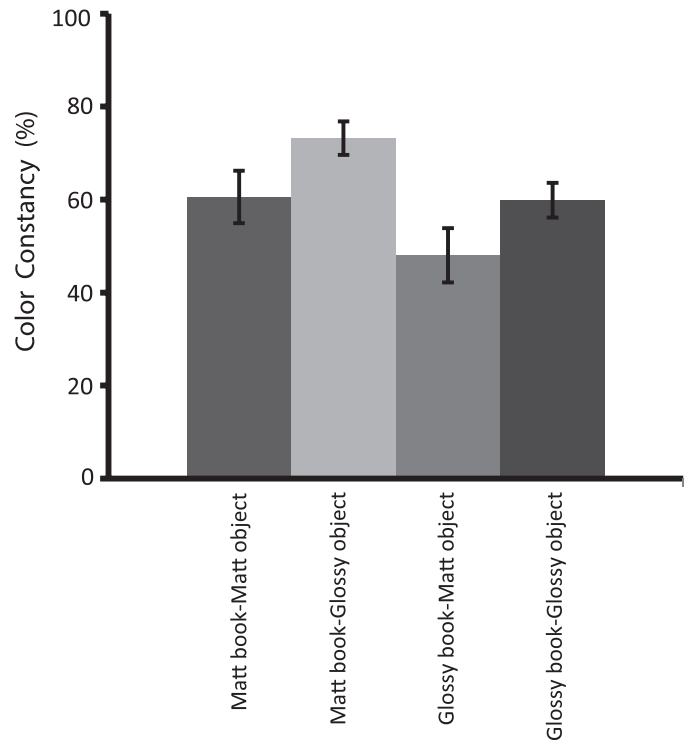
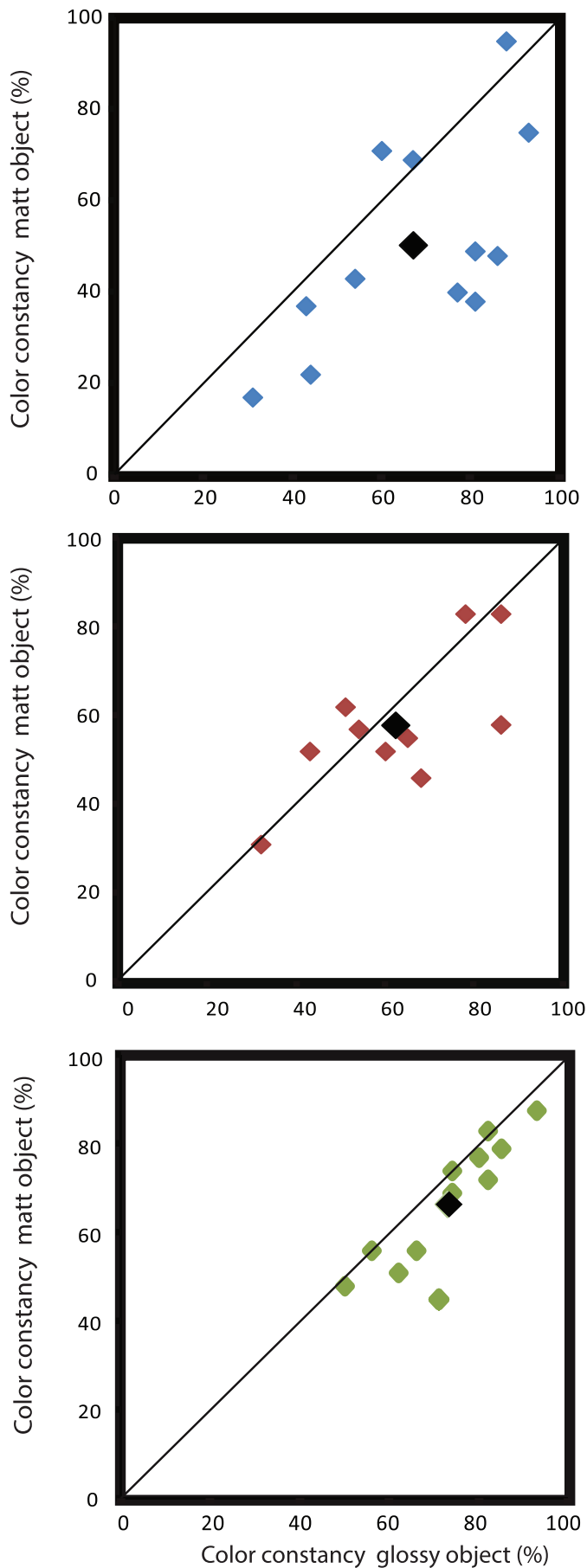


Figure 6. Results of Experiment 1 (exclusion of below-average quality color matches). Represented are the average (with the standard error) color constancy indices in percentages (shown on the y-axis) as a function of each combination of object (matte vs. glossy) and version of the Munsell book (glossy vs. matte), as shown on the x-axis.

gloss ( $r = 0.45$ ) and the green matte ( $r = 0.62$ ) cylinders when tested in both sessions. Thus, a direct comparison between both versions of the Munsell book must be considered with some caution. However, most important for our current discussion is whether there is any difference in both versions of the book in detecting differences in color constancy between the glossy and the matte cylinders. For that reason, we calculated the difference in constancy performance for the glossy and the matte cylinders. This was done separately for both versions of the Munsell book. A paired  $t$  test revealed no significant difference in the difference scores for either the matte or the glossy version of the Munsell book,  $t_{21} = -0.420$ ,  $p = \text{ns}$ .

This result shows two things. First, it shows that color matches are not better when glossy objects are tested with the glossy version or matte objects are matched with the matte version of the Munsell book. Second, the chance of detecting a significant difference in constancy between the glossy and matte objects is independent of which version of the book we use. The quality of the color matches was not different between the four combinations of the reflectance of an object (glossy vs. matte) and the two versions of the Munsell book (glossy vs. matte). The average quality of the



color match given by observers was 5 for all combinations.

As we were unable to find any significant difference in the difference scores between both versions of the Munsell book, we grouped the data for both versions to increase the power of our statistical analysis.

### Comparison between glossy and matte objects

Figure 7 shows the results for Experiment 1. Each graph shows the percentage of color constancy for the glossy objects (shown on the x-axis) as a function of the average color constancy percentage for the matte objects (shown on the y-axis). Each panel shows the results separately for the blue, red, and green cylinder (green data points indicate data for the green cylinders, red ones for the red cylinders, etc.). Each data point represents the data for a single observer. Thus, if there would be no difference in color constancy performance for our subjects, all symbols would lie on the unity line. Figure 7 shows that this is not the case; most of the symbols lie beneath the unity line indicating that overall subjects' color constancy performance was better when tested with the glossy objects compared to the matte objects. The big black diamond represents the average constancy percentage across subjects.

From this figure we can conclude other important factors. First of all, looking at the overall color constancy performance, we can see that there are very large differences in the amount of color constancy that subjects achieve. For example, looking at the top panel, the most extreme differences in color constancy are the ones in which one subject obtained an overall constancy percentage of 30% and 15% for the glossy and matte blue objects, respectively. On the other hand, another subject tested with the blue cylinders obtained on average a constancy percentage of 85% and 95% for the glossy and matte objects, respectively. It can also be observed in Figure 7 that there are substantial differences in the amount of color constancy that

Figure 7. Results of Experiment 1 (exclusion of below-average quality color matches). The color constancy indices (%) are shown for the glossy objects (x-axis) plotted against the color constancy indices for the matte objects (y-axis). Each data point represents the average color constancy performance for a single subject. The results are plotted separately for the three color groups, each shown in a separate panel. The diagonal line represents the unity line. If there was no difference between color constancy performance for the glossy versus the matte objects, all data points would lie on the unity line. The panels show that, for most subjects, the dots align to the right side of the unity line indicating superior performance for the glossy objects. The big black diamond represents the average constancy percentage across subjects.

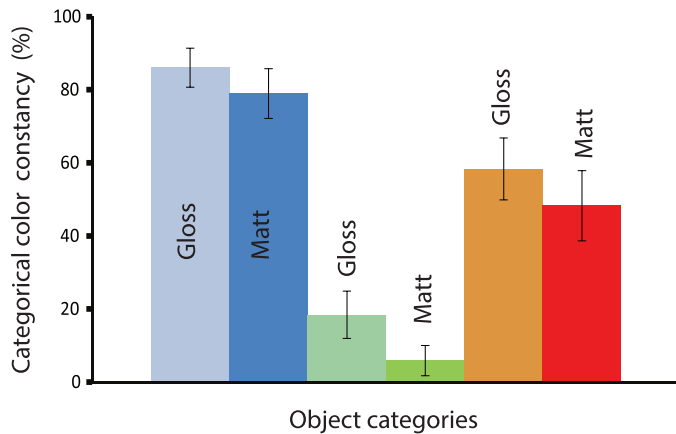


Figure 8. Results of Experiment 1 (exclusion of below-average quality color matches). The average data for the categorical color constancy index (%) are shown, ranging from zero (observers gave consistent different color categories for the same object under the different illuminants) to 100 (observers gave consistently the same color name to the same object under different illuminants). The data are shown separately for each color group (represented in the typical color), and for the gloss and matte objects separately.

subjects achieve. These large between-subject differences in color constancy performances have been found in other studies as well, but are typically neither discussed nor explained (Granzier, Toscani, & Gegenfurtner, 2012). Moreover, the color constancy performance of our subjects lies well within the dynamic range, indicating that the task at hand was neither too difficult nor too easy.

Finally, looking at Figure 7, the data seem to indicate that the superior performance for the glossy objects is largest when tested with the blue cylinders, compared to the red and green cylinders.

Repeated measures analysis of variance revealed that there was a significant effect of the factor *reflectance* showing that there is a statistical difference between constancy performance when tested with the glossy objects compared to when tested with the matte objects,  $F(1, 35) = 337.339$ ,  $p < 0.01$ . Paired samples two-tailed  $t$  tests (Bonferroni corrected) showed that color constancy was significantly higher ( $t_{37} = 4.149$ ,  $p < 0.01$ ) for the glossy objects compared to the matte objects. The paired  $t$  test revealed a significant difference between the glossy and matte objects for the blue cylinders ( $t_{11} = 3.183$ ,  $p < 0.01$ ) and the green cylinders ( $t_{11} = -2.705$ ,  $p < 0.01$ ) but not for the red color cylinders ( $t_{11} = 1.637$ ,  $p = \text{ns}$ ). A Fisher's least significant difference (LSD) post hoc test showed no significant differences in color constancy between the color groups when the reflectance (either glossy or matte) was the same.

As indicated above, we also computed a categorical color constancy index. The results from this analysis can

be seen in Figure 8, which shows the average categorical color constancy indices in percentages and the standard error for each color category (indicated in the representative color) and for each reflectance separately. From this figure, we can conclude that on average observers had higher categorical color constancy indices when tested with the glossy object compared to when tested with the matte objects. This latter result indicates that observers named the color of the cylinder more consistently under the different illuminants for the glossy object compared to the (identically colored) matte cylinder. Although this effect seems to be independent of the color group, the effect seems to be more pronounced for the green and red color groups. However, a  $\chi^2$  test did not reveal an overall significant difference between the glossy and the matte objects ( $p = 0.14$ ). Moreover, no significant difference in categorical color constancy could be detected for each separate color group between the glossy and the matte objects. One has to keep in mind that the categorical color constancy index is much weaker than our other index, in that it only checks whether the same hue category is assigned to an object under both illuminations. Any more subtle differences will be lost.

In short, the results as shown above indicate that color constancy is significantly better when tested with glossy objects compared to when tested with matte objects. Although the categorical color constancy indices were overall higher for the glossy objects compared to the matte objects, no statistical difference could be detected between both reflectances. Therefore, the superior constancy found for the glossy objects seems to be dependent on the way in which color constancy is computed (color constancy index vs. categorical color constancy index).

## Results for inclusion of poor color matches

In the previous analysis we excluded observers' color matches for analysis when they indicated that the color match was below average quality (a score of 1–3). It might have been the case that the superior color constancy performance found for the glossy objects compared to when tested with the matte objects was the result of introducing a bias in the results by leaving out these trials. In order to test whether this might have been the case, we repeated the same analysis but now including all trials. One observer was excluded from analysis as this observer failed to complete all trials. Therefore, for this analysis we had 11 observers.

### Effects of background and Munsell color book

Subjects had an average color constancy of 68% (averaged across the six different objects) with the

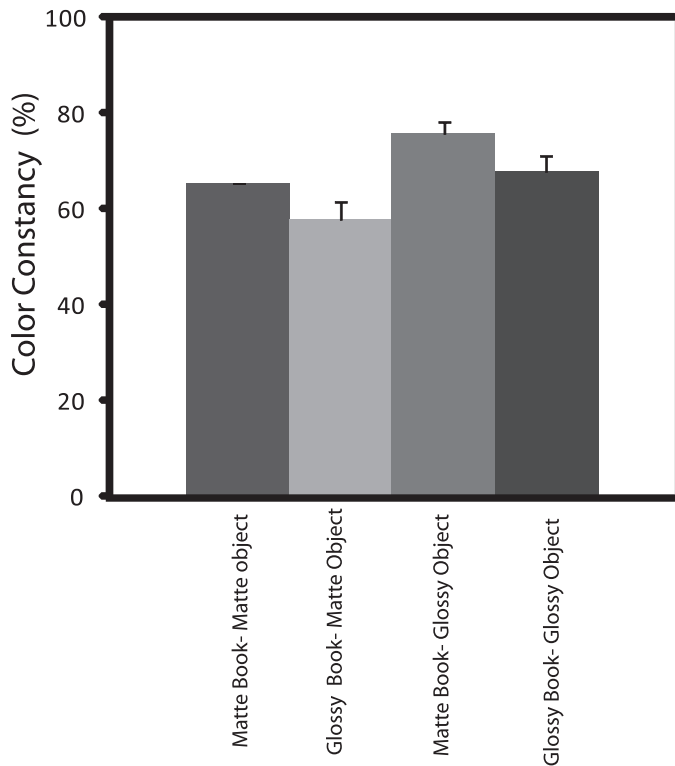


Figure 9. Results of Experiment 1 (inclusion of all color matches). The average (with the standard error) color constancy indices are represented in percentages (y-axis) as a function of each combination of object (matte vs. glossy) and version of the Munsell book (glossy vs. matte), as shown on the x-axis.

checkerboard compared to 65% with the black background, showing that local and global color contrast did only slightly provide subjects with additional information about the illumination when including all trials for analysis. The two numbers are not significantly different,  $F(1, 10) = 0.591$ ,  $p = ns$ . We made sure that there was no significant interaction with any of the potential experimental effects of material properties. There was indeed no effect of background on the difference in constancy between glossy and matte cylinders,  $F(1, 10) = 0.937$ ,  $p = ns$ .

A second experimental variation was the measurement of using both versions of the Munsell book. As indicated above, it would not be unexpected for observers to be able to find better matches for the matte object in the matte version of the Munsell book, and better matches for the glossy objects in the glossy version of the Munsell book. Interestingly, this was not the case.

When comparing color constancy for the two different versions of the Munsell book, there was a small improvement for the matte version over the glossy version (70% vs. 62%), but this was independent of the surface properties of the objects. Figure 9 shows the average color constancy indices for each combination of reflectance of the object and version of the

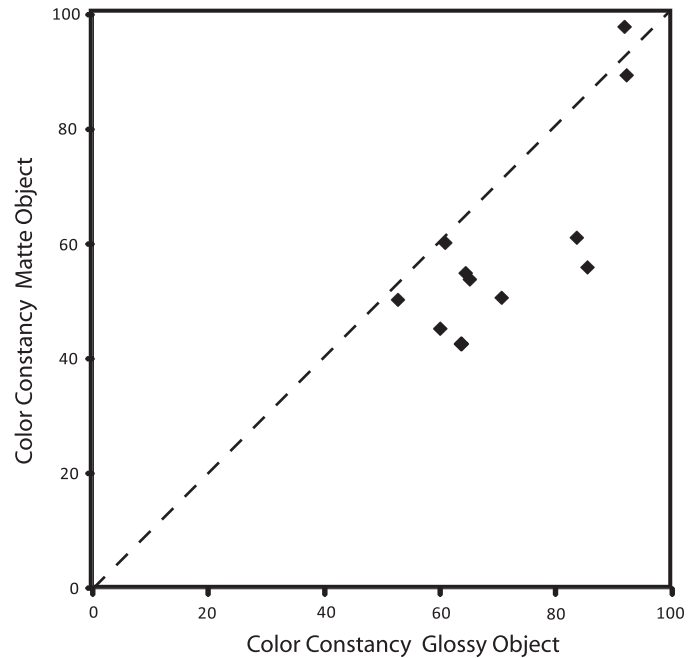


Figure 10. Results of Experiment 1 (inclusion of all color matches). The color constancy indices (%) for the glossy objects (x-axis) are plotted against the color constancy indices for the matte objects (y-axis). Each data point represents the average color constancy performance for a single subject. The data are averaged across color groups. The dashed line represents the unity line. If there was no difference between color constancy performance for the glossy versus the matte objects, all data points would lie on the unity line. The panels show that for most subjects, the data points align to the right side of the unity line indicating superior performance for the glossy objects.

Munsell book. Repeated measures of analysis of variance showed no significant main effect of the factor “Munsell book” on color constancy,  $F(1, 10) = 4.290$ ,  $p = ns$ . Moreover, no interaction could be detected between the factor Munsell book and the factor surface,  $F(1, 10) = 0.019$ ,  $p = ns$ .

The crucial aspect of this finding for our study is the lack of an interaction between the surface properties of the objects and the Munsell book. This lack of interaction indicates that our constancy measures are valid, independently of the properties of the surfaces used to measure them. Thus, we can measure color constancy and the effects of the objects’ surface properties independently of which Munsell book we use for testing, and independently of the particular background of the scenes. This result is identical to the results of the analysis when excluding the below-average rated color matches (see above).

#### Comparison between glossy and matte objects

Figure 10 shows the results for Experiment 1. The panel shows the percentage of color constancy for the

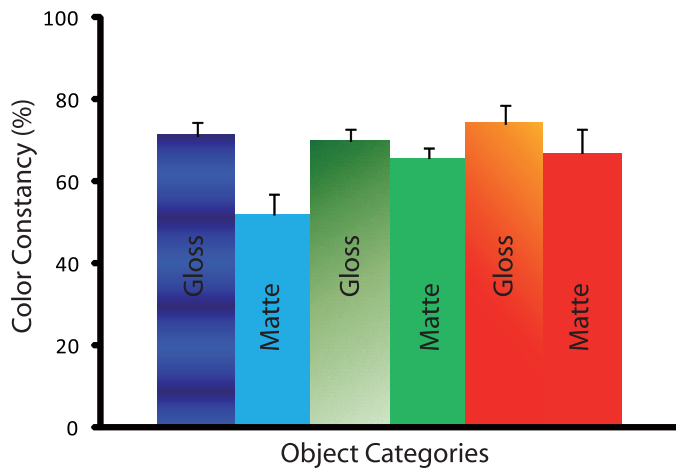


Figure 11. Results of Experiment 1 (inclusion of all color matches). The average data for the categorical color constancy index (%) are shown ranging from zero (observers gave consistent different color categories for the same object under the different illuminants) to 100 (observers gave consistently the same color name to the same object under different illuminants). The data are shown separately for each color group (represented in the typical color) and for the gloss and matte objects separately.

glossy objects on the x-axis and the average color constancy percentage for the matte objects on the y-axis. The results are averaged across the three color groups. Each data point represents the data for a single observer. Thus, if there were no difference in color constancy performance between matte and glossy objects, all symbols would lie on the unity line. Figure 10 shows that this is not the case; most of the symbols lie beneath the unity line, indicating that overall subjects' color constancy performance was better when tested with the glossy objects compared to the matte objects. Important for the purpose of our experiment, the color constancy performance of our subjects lies well within the dynamic range, indicating that the task at hand was neither too difficult nor too easy.

Repeated measures of analysis of variance revealed that there was a significant effect of the factor "reflectance" showing that color constancy was significantly higher when tested with the glossy objects compared to when tested with the matte objects,  $F(1, 10) = 8.708$ ,  $p = 0.015$ . A significant interaction between the factors "surface" and "color" on the amount of color constancy could be detected,  $F(2, 20) = 4.084$ ,  $p = 0.033$ .

A paired  $t$  test revealed a significant difference between the glossy and matte objects for the blue cylinders,  $t(10) = 3.183$ ,  $p < 0.001$ , and only a trend for the red color cylinders  $t(10) = 1.637$ ,  $p = 0.079$ , but not for the green cylinders  $t(10) = -2.705$ ,  $p = \text{ns}$ . A Fisher's least significant difference (LSD) post hoc test showed no significant differences in color constancy between

the color groups when the reflectance (either gloss or matte) was the same.

The results for the categorical color constancy indices are shown in Figure 11 for each of the three color groups (shown in the representative color) and also presented separately for either the glossy or the matte object. We conclude that for all color categories the average categorical color constancy indices are higher for the glossy objects compared to the matte objects. A  $\chi^2$  test was performed to investigate whether the distributions of categorical variables differ significantly from one another. Overall, the  $\chi^2$  test showed no significant difference between the glossy and the matte objects in the categorical color constancy distributions ( $p = 0.21$ ). However, the green color category showed a significant superior performance in categorical color constancy for the glossy objects compared to the matte objects ( $p = 0.013$ ). No significant difference could be detected between the glossy and the matte objects for the other color groups.

## Discussion

From the data of Experiment 1 we can conclude that subjects' color constancy performance is significantly better when tested with glossy objects compared to when tested with matte objects. On average, subjects' color constancy improved from 58% to 67% when tested with the glossy objects compared to when tested with the matte objects. This is a large effect considering the fact that in most color constancy studies subjects' performance lies between 20% and 80% (see Foster, 2011). When tested with the glossy objects compared to when using the matte objects, this superior color constancy performance was independent of whether we included the below-average quality color matches or not.

Testing subjects with either the matte version of the Munsell book or with the glossy version did not seem to make a large difference in detecting differences in constancy between glossy and matte objects. One would assume that, when one has to match the color of a glossy object, using glossy samples would improve performance than when tested with matte samples and vice versa. Probably this effect (if there is any) is too small to be reliably tested (although we used a relatively large subject sample). This could indicate that subjects can estimate surface reflectance of an object and at the same time ignore the material property of gloss.

Also, the effect of glossy objects on color constancy was independent from the background (black background vs. checkerboard). This latter result did also not depend on whether the color matches with ratings between 1 and 3 were either excluded or included from analysis.

However, different results were obtained between the three color groups in the effects of gloss on color

constancy performance depending on the data analysis; when excluding the below-average rated color matches, superior performance of the glossy objects on the color constancy index could be detected for the blue and green color categories, but not for the red color category. When including all data for analysis, significantly higher color constancy indices were obtained for the blue and red color categories, but not for the green color category. These results, in our opinion, are very important and stress the importance of letting observers rate the quality of their color percept. Although probably the same general conclusion can be drawn when including poorly rated color matches (if the effects of interest or the statistical power is big enough), the results of other experimental factors might highly depend on the choice of whether the poor color matches were either excluded or included from data analysis.

These differences in results can be explained easily if observers find it more difficult to find a suitable color match for some colors but not for others. This situation might arise when the gamut is too limited when viewing a given object under a particular illuminant. The inclusion of these color matches, when, in fact, the gamut is too limited to find a suitable color match, could potentially introduce a bias in the results. Obviously similar gamut problems might occur when using a CRT to match colors under different illuminants.

A similar conclusion can be drawn by looking at the different ways in which color constancy is computed; either by using a color constancy index or a categorical color constancy index. The same general conclusion can be drawn from both analyses; glossy objects show, to some extent, superior color constancy performance compared to the matte object category. However, both ways of computing color constancy show different effects with respect to color categories. The biggest effects between the glossy and matte objects are found for the green and red color categories when excluding the poorly rated color matches, while the biggest differences in constancy between the glossy and matte objects can be found for the blue and red color categories when the poorly rated color matches are included for analysis. That the color constancy indices are highly dependent on how (or by which formula) color constancy is computed, has been discussed in greater detail elsewhere (Foster, 2011). More consideration should be given to both types of methodological issues.

## Experiment 2: The effects of surface roughness on color constancy

The same 12 subjects and the same set-up and analysis as in Experiment 1 were used for Experiment 2, with the exception of the visual stimuli. For this

	Black background (%)	Checkerboard (%)
Rough objects	53%	61%
Smooth objects	59%	67%

Table 2. The color constancy indices (%) shown separately for the rough and smooth objects (rows) and the two background conditions (columns).

experiment, we wanted to test whether subjects obtain different amounts of color constancy for smooth compared to rough objects. The stimuli are shown in Figure 12. The left column shows the objects with the rough surface, while on the right, the cylinders for the smooth surface are depicted.

For the smooth objects, the same objects that were used in Experiment 1 (matte objects category) were also used in Experiment 2. For the rough objects, we used the roughest available sandpaper (CAMI grit designation of 80) and glued it on identically shaped cylinders. The cylinders were painted in identical colors as the smooth objects. Thus, only surface roughness was varied in this experiment. In total we had six different stimuli, a rough red object, a rough green object, and a rough blue object, and three corresponding smooth objects. Once again, both object categories were tested for both the black and the checkerboard background. As we were unable to find any significant differences in the difference scores between using the glossy and the matte versions of the Munsell book, subjects were only tested with the matte version in this experiment. Six objects were shown under the three illuminations and two different backgrounds. This resulted in a total of 36 trials per subject. Each session took about 90 minutes.

## Results for exclusion of poor color matches

### Effects of background

Once again we tested whether there would be a difference in color constancy performance when the objects were either placed in front of a black background or whether they were presented against the checkerboard. No systematic differences in constancy performance could be detected between both background conditions (see Table 2). Therefore, we grouped the data of both background conditions to analyze the data.

### Comparison between rough and smooth objects

Table 3 shows an overview of the percentages of trials that were included for data analysis for Experiment 2. This table shows that only a small fraction of the total number of trials were excluded for data analysis.

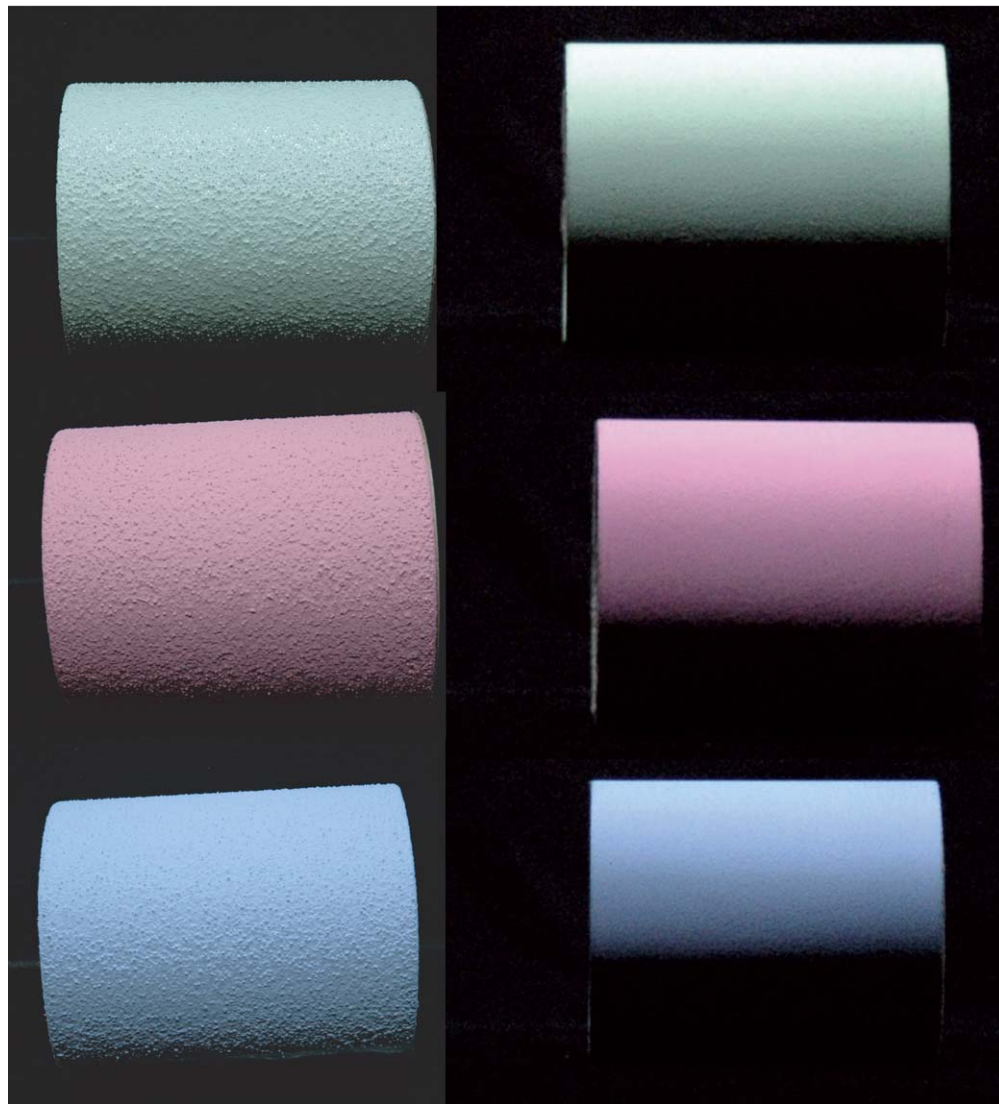


Figure 12. The rough stimuli (left column) and the smooth stimuli (right column) used in Experiment 2 illuminated by D65.

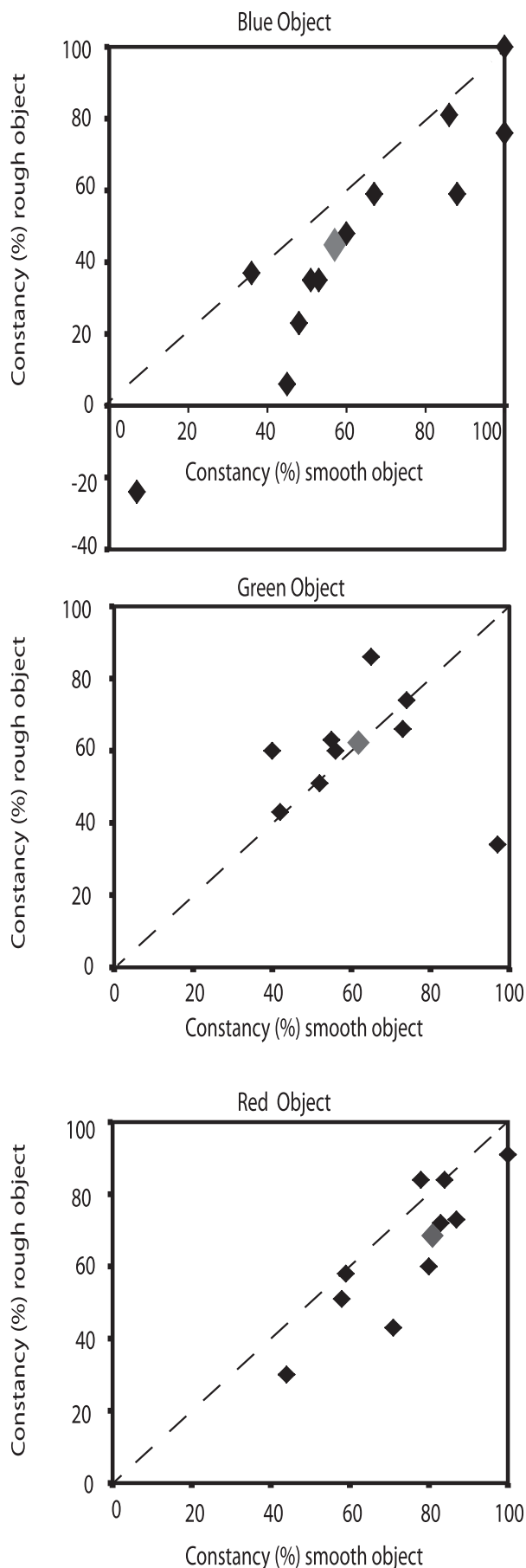
Figure 13 shows the results for Experiment 2. The figure is similar to Figure 7. The color constancy percentages for each subject (indicated by different symbols) for the rough objects (y-axis) as a function of the constancy percentage obtained for the smooth objects (x-axis) are shown. If there was no difference in

	Good trials (%)
Blue rough object	92%
Blue smooth object	96%
Green rough object	80%
Green smooth object	88%
Red rough object	88%
Red smooth object	92%

Table 3. Percentages of good trials for Experiment 2, shown separately for each cylinder/object.

constancy for the rough and smooth objects, all symbols would lie on the unity line. The data are shown separately for each color category in a separate panel. From this figure we can conclude the following:

Of primary interest is that most symbols lie below the unity line indicating superior constancy performance for the smooth objects compared to the rough objects. However, this effect seems once again to be dependent on the color group, as for the green cylinder this effect is absent. What is surprising to see is that the positive effects of smooth objects on constancy performance are quite large considering that we did not manipulate surface roughness in extreme ways. For example, looking at the data for the blue object, one subject obtained a constancy performance of only 5% for the rough object, while his performance for the smooth object was about 45%, which is obviously a huge difference. It seems to be generally the case that



the improvement for smooth objects was higher for observers with a low overall constancy.

Repeated measures showed a main effect of the factor “Reflectance,”  $F(1, 35) = 8.906$ ,  $p < 0.01$ . Paired sample  $t$  test revealed a significant superior color constancy performance for the smooth objects compared to the rough ones for the blue cylinders,  $t(11) = -4.647$ ,  $p < 0.01$ , and for the red cylinders,  $t(11) = -2.615$ ,  $p < 0.05$ , but not for the green cylinders,  $t(11) = -1.069$ ,  $p = \text{ns}$ . A Fisher’s least significant difference (LSD) post hoc test showed significant differences in color constancy between the color groups when the reflectance (either glossy or matte) was the same. Color constancy was significantly higher for the red rough cylinders compared to the blue rough cylinders,  $F(5, 11) = 14.985$ ,  $p < 0.01$ , and also significantly higher for the green rough cylinders compared to the blue rough cylinders  $F(5, 11) = 14.985$ ,  $p < 0.01$ . No significant difference could be detected between the other pairwise comparisons.

Figure 14 shows the results for the categorical color constancy index. The reasoning of this figure is identical to that of Figures 8 and 11. The results show that, with exception to the red color category, no difference in the categorical color constancy index can be observed between the rough and the smooth objects. The  $\chi^2$  test showed overall no significant difference between the smooth and the rough objects ( $p = 0.45$ ). No significant differences could be detected between the different color groups.

## Results for inclusion of poor color matches

### Effects of background

A main effect of background on color constancy performance could be detected,  $F(1, 10) = 6.043$ ,  $p = 0.032$ . Color constancy increased from 59% to 69% when a colored background was presented compared to when a black background was presented. However, no

Figure 13. Results of Experiment 2 (exclusion of below-average quality color matches). The color constancy indices (%) are shown for the smooth objects (x-axis) plotted against the color constancy indices for the rough objects (y-axis). Each data point represents the average color constancy performance for a single observer. The data are shown separately for each color category in a separate panel. The gray diamond represents the average color constancy index across observers. The diagonal line represents the unity line. If there was no difference between color constancy performance for the smooth versus the rough objects, all data points would lie on the unity line. The panels show that for most observers, the symbols are located on the right side of the unity line indicating superior constancy performance for the smooth objects.



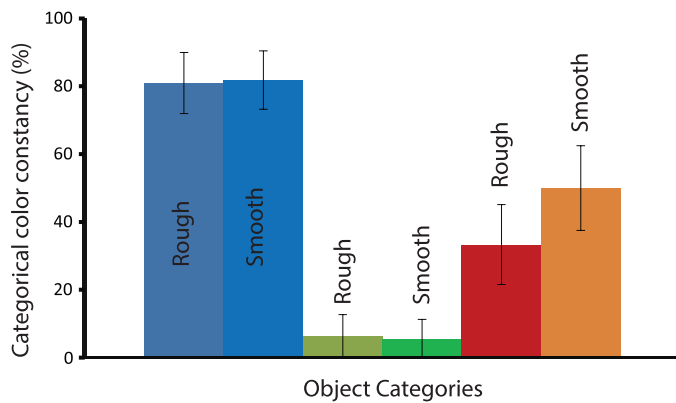


Figure 14. Results for Experiment 2 (exclusion of below-average quality color matches). The average categorical color constancy index (with their standard error) is shown on the y-axis as a function of the object categories (rough vs. smooth). The results are shown separately for each of the three color groups as shown in the representative color.

interaction between the factors “surface” and “background” could be revealed,  $F(1, 10) = 0.012$ ,  $p = ns$ .

Figure 15 shows the results for Experiment 2 when all data are included for analysis. The color constancy percentages for each subject (indicated by different symbols) are shown for the rough objects (y-axis) as a function of the constancy percentage obtained for the smooth objects (x-axis). The data are averaged across color groups. If there was no difference in constancy for the rough and smooth objects, all symbols would lie on the unity line. From this figure we can conclude the following:

Of primary interest is that most symbols lie below the unity line indicating superior constancy performance for the smooth objects compared to the rough objects. However, this effect seems once again to be dependent on the color group. Repeated ANOVA showed a trend for the factor “Reflectance,”  $F(1, 11) = 4.529$ ,  $p = 0.057$ . Paired sample  $t$  tests revealed only a significant superior color constancy performance for the smooth objects compared to the rough ones for the blue cylinders,  $t(10) = -4.647$ ,  $p = 0.02$ , but not for the red cylinders,  $t(10) = -2.615$ ,  $p = ns$ , and the green cylinders,  $t(10) = -1.069$ ,  $p = ns$ . A Fisher’s least significant difference (LSD) post hoc test showed significant differences in color constancy between the color groups when the reflectance (either glossy or matte) was the same. No significant interaction could be detected between the other factors.

The results of Experiment 2 for the categorical color constancy indices showed only superior categorical color constancy performance for the red smooth object compared to the red rough object (54% vs. 33%). No apparent differences in categorical color constancy could be detected between the green smooth object and the green rough object (both indices of 8%) and blue

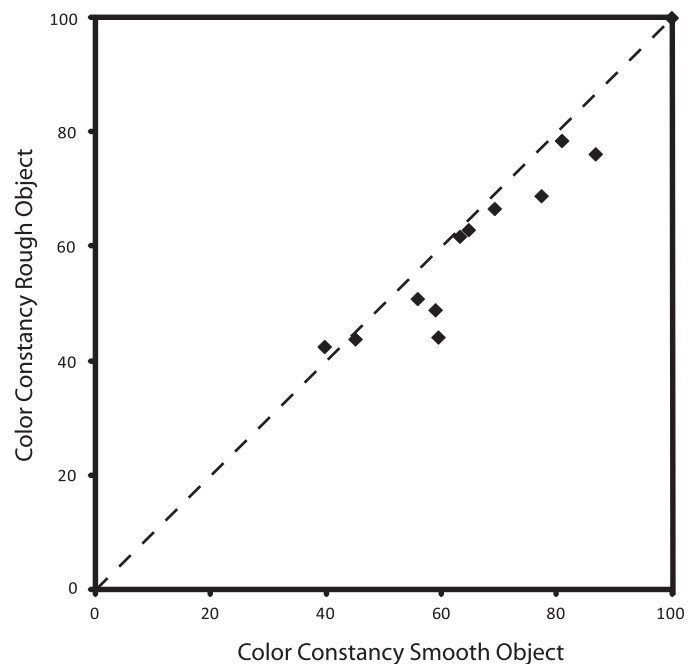


Figure 15. Results of Experiment 2 (inclusion of all trials). The color constancy indices (%) are shown for the smooth objects (x-axis) plotted against the color constancy indices for the rough objects (y-axis). Each data point represents the average color constancy performance for a single observer. The results are averaged across the three color groups. If there was no difference between color constancy performance for the smooth versus the rough objects, all data points would lie on the unity line. The panel shows that for most observers, the symbols are located on the right side of the unity line indicating superior constancy performance for the smooth objects.

color category (both indices of 79%). A  $\chi^2$  test showed overall no significant difference between the smooth and the rough objects ( $p = 0.4$ ) and no significant difference between the separate color groups.

## General discussion

Our results show that material properties like gloss and roughness can potentially have significant effects on color constancy.

Our results are in line with Xiao et al. (2012), who concluded that results found for flat matte tests may not be easily generalized to three-dimensional objects. We also found that material properties (i.e., gloss) can influence color constancy. Our results are most consistent with those of their reduced-cue condition, in which they silenced local contrast as a cue to the illuminant change, and by doing so they could separate the effect of local contrast from other processes that mediate constancy. These authors found that color

constancy was systematically worse for a matte disk (two-dimensional object) that they used compared to the spheres that these authors used (simulated three-dimensional object). These same authors also tested whether there were differences between matte versus glossy spheres, but they could not detect any effect. This is different from our results. Quite possibly, the fact that we used more observers and used real objects instead of simulated stimuli could perhaps explain the differences between our study and the results of Xiao et al. (2012).

The improvement of color constancy with the addition of specular highlights is consistent with previous reports (Yang & Maloney, 2001; Yang & Shevell, 2003). Yang and Maloney (2001) used simulated scenes containing a uniform background plane perpendicular to the observer's line of sight and a small number of specular, colored spheres resting on the uniform background. Observers adjusted the color of a small, simulated test patch to appear achromatic. They found that the specular highlight cue had a significant influence on the achromatic settings. However, an important difference between our results and those of Yang and Maloney (2001) is that the latter found no effect of a single specular highlight (one sphere) on achromatic setting, but only found an effect of specular highlights when there were many identical specular highlights (eleven spheres). These authors concluded from their study that the visual system requires multiple identical highlights before giving weight to any one of them. We show here, by using real objects, that the presence of one object with specular reflections already suffices to considerably enhance color constancy performance. However, it is hard to state the exact number of highlights that are present on a given surface. For example, looking at Figure 2 one cannot state whether the glossy cylinder contains only one quite large highlight or whether the object contains multiple smaller highlights that the visual system might use for estimating the illumination. An additional complication in making a direct comparison between our current results and those of Yang and Maloney (2001) is that the latter used virtual stimuli while we used real stimuli. It is still uncertain whether one can generalize between using virtual scenes and when using real scenes (e.g., Hedrich & Bloj, 2010; Granzier, Brenner, & Smeets, 2009c). Also, the amount of color constancy seems to be dependent on the type of task (Brenner, Granzier, & Smeets, 2011) and the size of the visual field (Hansen, Walter, & Gegenfurtner, 2007). All these factors were unequal between our current study and the one of Yang and Maloney (2001).

From Experiment 2 we can conclude that subjects are slightly better able to estimate surface reflectance for smooth objects than for rough objects. However, this effect seems to depend on the object's color. Why

there is a positive effect of smooth surfaces on constancy performance for some colors but not for others is unclear. Interestingly, there was no advantage for the green and the red objects, for which constancy was already higher than for the blue color group. It seems that the advantage for smooth objects, or equivalently, the disadvantage for rough objects, mainly occurs at low overall levels of constancy. The harder it is for subjects to estimate surface reflectance under a given illumination, the more they are affected by the rough object surfaces, which have more interreflections, which in turn could make surface estimation more difficult. This hypothesis is, however, hard to test as it is hard to define from the outset what makes a difficult combination of object's color and illumination's chromaticity. It was again the blue color group that showed the largest effect in Experiment 1.

It is also noteworthy to state that the Munsell book that observers used to make their color match in Experiment 2 contained only smooth chips. It would have been interesting to test whether similar results would have been found when observers could have used a Munsell book containing only rough chips. However, it was not our main purpose to systematically investigate the effects that the different versions of the Munsell book have on color constancy performance.

It is noteworthy that the superior color constancy performance found for the glossy objects (and the smooth objects) was dependent on the way in which color constancy was computed, either categorically or by way of color matching. This shows that the effects of gloss and roughness on color constancy performance might depend on the task at hand, as previous studies have suggested that there might be subtle differences between color naming (categorical color constancy) and color matching in a color constancy paradigm (Brenner, Granzier, & Smeets, 2011). Also noteworthy is the fact that we were able to find an effect of gloss and, to a lesser extent, surface roughness on color constancy, even when not using extreme examples of both dimensions (very rough or very shiny objects). Therefore, the results of surface gloss and roughness might be larger when both dimensions are more exaggerated.

In sum, our results show that there can be significant effects of material properties (roughness and gloss studied here) on color constancy performance. Our experiments show that it is feasible to build real-life three-dimensional objects that have controlled surface properties. It should be the goal of vision science to use stimuli that approach the complex visual world that we encounter in daily life.

*Keywords:* color perception, color constancy, illumination, material perception, shape, texture

## Acknowledgments

This research was supported by grant number DFG grant GE 879/9 to KRG.

Commercial relationships: none.

Corresponding author: Jeroen J. M. Granzier.

Email: j\_granzier@hotmail.com.

Address: Department of Psychology, Justus Liebig University, Giessen, Germany.

## References

- Arend, L. E., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America. A. Optics, Image Science, and Vision*, *A3*, 1743–1751.
- Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology*, *20*(1), 38–64.
- Blackwell, K. T., & Buchsbaum, G. (1988). Quantitative studies of color constancy. *Journal of the Optical Society of America. A. Optics, Image Science, and Vision*, *5*, 1772–1780.
- Bloj, M. G., & Hurlbert, A. C. (2002). An empirical study of the traditional Mach card effect. *Perception*, *31*(2), 233–246.
- Bramão, I., Faísca, L., Petersson, K. M., & Reis, A. (2010). The influence of surface color information and color knowledge information in object recognition. *The American Journal of Psychology*, *123*(4), 437–446.
- Bramão, I., Reis, A., Petersson, K. M., & Faísca, L. (2011). The role of color information on object recognition: A review and meta-analysis. *Acta Psychologica*, *138*(1), 244–253.
- Brenner, E., Granzier, J. J. M., & Smeets, J. B. J. (2011). Color naming reveals our ability to distinguish between a colored background and colored light. *Journal of Vision*, *11*(7):8, 1–16, <http://www.journalofvision.org/content/11/7/8>, doi:10.1167/11.7.8. [PubMed] [Article]
- de Almeida, V. M., Fiadeiro, P. T., & Nascimento, S. M. (2004). Color constancy by asymmetric color matching with real objects in three-dimensional scenes. *Vision Neuroscience*, *21*(3), 341–345.
- Foster, D. H. (2011). Color constancy. *Vision Research*, *51*(7), 674–700.
- Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, *10*(13), 805–808.
- Giesel, M., & Gegenfurtner, K. R. (2010). Color appearance of real objects varying in material, hue, and shape. *Journal of Vision*, *10*(9):10, 1–21, <http://www.journalofvision.org/content/10/9/10>, doi:10.1167/10.9.10. [PubMed] [Article]
- Granzier, J. J. M., Brenner, E., & Smeets, J. B. J. (2009a). Reliable identification by color under natural conditions. *Journal of Vision*, *9*(1):39, 1–8, <http://www.journalofvision.org/content/9/1/39>, doi:10.1167/9.1.39. [PubMed] [Article]
- Granzier, J. J. M., Brenner, E., & Smeets, J. B. J. (2009b). Can illumination estimates provide the basis for color constancy? *Journal of Vision*, *9*(3):18, 1–11, <http://www.journalofvision.org/content/9/3/18>, doi:10.1167/9.3.18. [PubMed] [Article]
- Granzier, J. J. M., Brenner, E., & Smeets, J. B. J. (2009c). Do people match surface reflectance fundamentally differently than they match emitted light? *Vision Research*, *49*, 702–707.
- Granzier, J., & Gegenfurtner, K. R. (2012). Effects of memory color on color constancy for unknown colored objects. *i-Perception*, *3*(3), 190–215.
- Granzier, J., Toscani, M., & Gegenfurtner, K. R. (2012). The role of eye movements in chromatic induction. *Journal of the Optical Society of America A*, *29*, A356–A368.
- Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision*, *7*(4):2, 1–15, <http://www.journalofvision.org/content/7/4/2>, doi:10.1167/7.4.2. [PubMed] [Article]
- Hedrich, M., & Bloj, M. (2010). Basic color names for 2-D samples: Effects of presentation media and illuminants. *Ophthalmic and Physiological Optics*, *30*(5), 638–645.
- Hedrich, M., Bloj, M., & Ruppertsberg, A. I. (2009). Color constancy improves for real 3D objects. *Journal of Vision*, *9*(4):16, 1–16, <http://www.journalofvision.org/content/9/4/16>, doi:10.1167/9.4.16. [PubMed] [Article]
- Ishihara, S. (1969). *Tests for color blindness: 38 plate edition*. Tokyo, Japan: Kanehara Shuppan Co. Ltd.
- Jameson, D. (1985). Opponent-color theory in light of physiological finding. In D. Ottoson & S. Zeki (Eds.), *Central and peripheral mechanisms of color vision*, 8–102. New York: Macmillan.
- Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences, USA*, *96*, 307–312.
- Maloney, L. T. (2002). Illuminant estimation as cue combination. *Journal of Vision*, *2*(6):6, 493–504,

- <http://www.journalofvision.org/content/2/6/6>, doi:10.1167/2.6.6. [PubMed] [Article]
- Maloney, L. T., & Brainard, D. H. (2010). Color and material perception: Achievements and challenges. *Journal of Vision*, 10(9):19, 1–6, <http://www.journalofvision.org/content/10/9/19>, doi:10.1167/10.9.19.
- Olkkonen, M., & Brainard, D. H. (2010). Perceived glossiness and lightness under real-world illumination. *Journal of Vision*, 10(9):5, 1–19, <http://www.journalofvision.org/content/10/9/5>, doi:10.1167/10.9.5. [PubMed] [Article]
- Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of Vision*, 10(9):16, 1–22, <http://www.journalofvision.org/content/10/9/16>, doi:10.1167/10.9.16. [PubMed] [Article]
- Smithson, H. E. (2005). Sensory, computational and cognitive components of human color constancy. *Philosophical Transactions of the Royal Society Biological Sciences*, 360, 1329–1346.
- Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Percept Psychophys*, 61(6), 1140–1153.
- Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Science*, 5(5), 211–215.
- Troost, J. M., & de Weert, C. M. (1991). Naming versus matching in color constancy. *Perception & Psychophysics*, 50(6), 591–602.
- Uttl, B., Graf, P., & Santacruz, P. (2006). Object color affects identification and repetition priming. *Scandinavian Journal of Psychology*, 47(5), 313–325.
- Valberg, A., & Lange-Malecki, B. (1990). Color constancy in Mondrian patterns: A partial cancellation of physical chromaticity shifts by simultaneous contrast. *Vision Research*, 30, 371–380.
- World Medical Association. (2004). *Ethical principles for medical research involving human subjects*. Retrieved from <http://www.wma.net/en/30publications/10policies/b3/>.
- Wurm, L. H., Legge, G. E., Isenberg, L. M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance*, 19(4), 899–911.
- Xiao, B., & Brainard, D. H. (2008). Surface gloss and color perception of 3D objects. *Visual Neuroscience*, 25(3), 371–385.
- Xiao, B., Hurst, B., MacIntyre, L., & Brainard, D. H. (2012). The color constancy of three-dimensional objects. *Journal of Vision*, 12(4):6, 1–15, <http://www.journalofvision.org/content/12/4/6>, doi:10.1167/12.4.6. [PubMed] [Article]
- Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vision Research*, 41, 2581–2600.
- Yang, J. N., & Shevell, S. K. (2003). Surface color perception under two illuminants: The second illuminant reduces color constancy. *Journal of Vision*, 3(5):4, 369–379, <http://www.journalofvision.org/content/3/5/4>, doi:10.1167/3.5.4. [PubMed] [Article]
- Zaidi, Q., & Bostic, M. (2008). Color strategies for object identification. *Vision Research*, 48(26), 2673–2681.