Journal of Vision (2017) 17(1):40, 1–14

Differences in illumination estimation in #thedress

Matteo Toscani
Department of Psychology, Giessen University,
Giessen, Germany

Karl R. Gegenfurtner
Department of Psychology, Giessen University,
Giessen, Germany

Katja Doerschner
Department of Psychology, Giessen University,
Giessen, Germany
Department of Psychology & National Magnetic
Resonance Research Center, Bilkent University,
Ankara, Turkey

We investigated whether people who report different colors for #thedress do so because they have different assumptions about the illumination in #thedress scene. We introduced a spherical illumination probe (Koenderink, Pont, van Doorn, Kappers, & Todd, 2007) into the original photograph, placed in fore-, or background of the scene and—for each location—let observers manipulate the probe's chromaticity, intensity and the direction of the illumination. Their task was to adjust the probe such that it would appear as a white sphere in the scene. When the probe was located in the foreground, observers who reported the dress to be white (white perceivers) tended to produce bluer adjustments than observers who reported it as blue (blue perceivers). Blue perceivers tended to perceive the illumination as less chromatic. There were no differences in chromaticity settings between perceiver types for the probe placed in the background. Perceiver types also did not differ in their illumination intensity and direction estimates across probe locations. These results provide direct support for the idea that the ambiguity in the perceived color of the dress can be explained by the different assumptions that people have about the illumination chromaticity in the foreground of the scene. In a second experiment we explore the possibility that blue perceivers might overall be less sensitive to contextual cues, and measure white and blue perceivers' dress color matches and labels for manipulated versions of the original photo. Results indeed confirm that contextual cues predominantly affect white perceivers.

Introduction

The phenomenon started with a photograph of a dress posted on the social networking service Tumblr (http://swiked.tumblr.com/post/112073818575/guys-please-help-me-is-this-dress-white-and). A heated debate arose about the colors of the materials of which the depicted dress was made of (Figure 1): Looking at the same photograph on the same monitor some people reported to see a blue dress with black lace ("blue perceivers"), while the person next to them might insist on seeing a white dress with a golden lace ("white perceivers"). This sparked a public interest in visual perception and was a powerful illustration of the fact that what we perceive is subjective and does not necessarily match physical reality (Brainard & Hurlbert, 2015). A scientific report following the informal debate found that 57% of 1400 respondents to an internet survey reported the dress to be blue and black whereas 30% reported it to be white and gold and about 10% saw it as blue and brown. Approximately 10% of the respondents reported that they could switch between any of the color combinations (Lafer-Sousa, Hermann, & Conway, 2015).

Melgosa and colleagues (Melgosa, Gómez-Robledo, Isabel Suero, & Fairchild, 2015) compared spectroradiometric measurements of the dress photograph presented on a typical LCD monitor (HP w255hc) with those taken from an original model of this dress (from the retailer Roman Originals), assuming a D65 illuminant and a CIE 1931 colorimetric standard observer. They found that measurements of the original dress were substantially lower in the CIE L* dimension than those of the one in the photography. Moreover, while the color of the dress body was bluish in hue in both, the photography and the original dress, the lace in the photograph was brownish whereas in the original dress it was black. The latter difference would be compatible with the effect of white balance as many

Citation: Toscani, M., Gegenfurtner, K. R., & Doerschner, K. (2017). Differences in illumination estimation in #thedress. *Journal of Vision*, *17*(1):22, 1–14, doi:10.1167/17.1.22.

doi: 10.1167/17.1.22 Received July 8, 2016; published January 31, 2017 ISSN 1534-7362

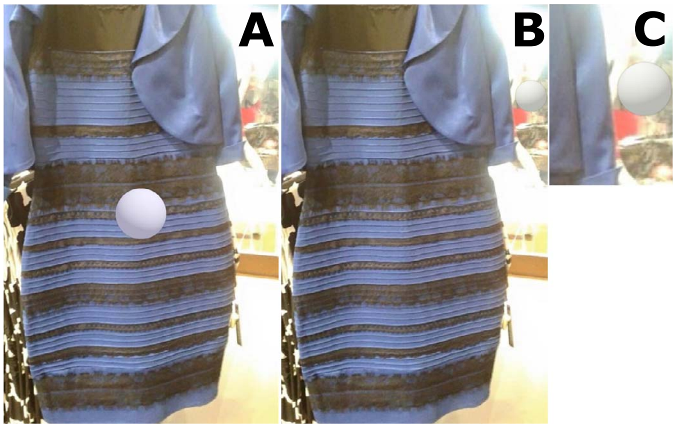

Figure 1. Stimuli Experiment 1. The probe placed in foreground (A) and background positions (B) of the original dress photograph. The right inlay (C) highlights the location of the background probe. Both probe locations were clearly visible to the observer. Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

smartphone cameras perform automatically. Whereas these physical differences between original and photograph might explain why some people might have called the blue dress "white," it does not explain why some people stuck with the "blue" interpretation.

In order to go beyond the informal reports of people, researchers took this question to the lab and compared asymmetric color matches on the dress photograph between white and blue perceivers, with some mixed results. Whereas Gegenfurtner, Bloj, & Toscani (2015) showed that differences between matches of the two groups were significant only in luminance—not in chromaticity—Lafer-Sousa et al. (2015) found that asymmetric color matches differed in both domains for their much larger group of internet observers. Despite these differences between Gegenfurtner et al. (2015) and Lafer-Sousa et al. (2015), their combined results show that white and blue perceivers did perceive different colors and did not simply use different color labels for identical percepts. Yet, this still leaves the question of why these perceptual differences existed.

It is well known that the perceived color of a surface patch does not correspond to a photometric measurement of the spectrum of light reflected from that surface but that, instead, it depends on an active processing of the whole scene by the visual system (Foster, 2011; Hurlbert, 1999). Humans (and many other animals) possess the ability to assign a constant color to an object despite a change in retinal stimulation as it occurs when the spectrum of the illumination changes (as during the course of a day). The implication for the dress is that when the visual system assigns a color to the cloth of the garment, it must also take into account the illumination of the scene in which the photograph was taken. It seems, however, that #thedress photograph provides information about the scene structure and/or illumination that could be consistent with several interpretations. What exactly is the ambiguity in the photograph that could give rise to the observed differences in perceived dress color?

Gegenfurtner et al. (2015) hypothesized that the ambiguity arises because the distribution of colors *within the dress* closely matches the distribution of natural daylights (Granzier & Valsecchi, 2014; Wyszecki & Stiles, 1982). Previous results (Beer, Dinca, & Macleod, 2006; Bosten, Beer, & MacLeod, 2015; Pearce, Crichton, Mackiewicz, Finlayson, & Hurlbert, 2014; Witzel, Valkova, Hansen, & Gegenfurtner, 2011) have lent support to the idea that people are less certain in color estimation along the daylight locus (a region of the color space along whose axis' daylight varies). More specifically, when people are asked to produce neutral settings of objects (e.g., of a uniform disk), the settings tended to vary along the daylight locus (Witzel et al., 2011). Pearce et al. (2014) found that discrimination of illumination changes is particularly poor along the daylight locus. Gegenfurtner et al. (2015) rotated the color distribution of #thedress photograph, forcing it to lie away from the daylight locus and showed that the ambiguity in perceived dress color disappeared with this manipulation. Thus, the distribution of colors in the photograph along daylight locus creates a degree of uncertainty that might force observers resort to their priors in interpreting the image – and it is those differences in priors that distinguishes a "blue" from a "white" perceiver. What then, are these priors about?

In order to perceive the color of a surface, the visual system must take into account the spectral and intensity information of the illumination (Boyaci, Doerschner, & Maloney, 2006; Boyaci, Doerschner, Snyder, & Maloney, 2006; Boyaci & Maloney, 2004; Doerschner, Boyaci, & Maloney, 2004; Foster, 2011). Authors that have investigated the dress phenomenon so far (Brainard & Hurlbert, 2015; Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Winkler, Spillmann, Werner, & Webster, 2015) have all proposed an explanation based on the estimation of the illumination of the scene, suggesting that differences in illumination estimation would lead to different color percepts. For example, light blue pixel values (within the boundaries of the dress) could be due to a bluish illumination on a white dress, or be due to a neutral light on a blue dress. To demonstrate this dependence on estimated scene illumination, Lafer-Sousa et al. (2015) cut the dress from the original photograph and embedded it in two new scenes that differed dramatically in background luminance and color. They found that—regardless of whether participants were white or blue perceivers—the color of the dress depended predominantly on the context in which it was presented, and they interpreted their results as evidence that the difference in perceived dress color must be due to differences in perceived illumination. Taken together, this suggests that the visual system of white and blue perceivers might discount different illuminations in order to arrive at the respective perceived dress color in the original photograph. However, this idea has not been tested explicitly thus far.

How is the illumination estimated in the first place? A scene provides information (cues) about the spatial and spectral content of the illumination (Foster, 2011; Maloney, 1999). Such cues can be, for example, the shading patterns on an undulated object, cast shadows, or specular highlights, to name a few. From these cues, the visual system is able to form a so-called "equivalent illumination" model of the scene (Brainard, Brunt, & Speigle, 1997) and its interpretation of the retinal image structure is consistent with this model. For example, Brainard et al. (1997) showed that errors in surface color matching were consistent with an observer who incorrectly estimates the illumination color and discounts it by using the incorrect estimate (i.e., "the equivalent illuminant"). Others (e.g., Boyaci & Maloney, 2004; Boyaci et al., 2006; Doerschner, Boyaci, & Maloney, 2004) found that observers' achromatic matches are affected by the estimated spatial and spectral distribution of the light sources in a 3D scene. Thus, there is evidence—albeit inferred—that the visual system can indeed estimate illumination properties from images when judging surface color.

Other researchers have probed illumination estimation explicitly: Koenderink, Pont, van Doorn, and Kappers (2007) measured the perceived illumination at multiple locations in a set of different realistic scenes. They used a probe consisting of a white sphere at a given location and asked the observers to adjust the appearance of the probe to appear "as if it is a white sphere embedded in the scene." In order to achieve that appearance, observers could adjust the direction, diffuseness, and intensity of the illumination. Observers' adjustments were close to veridical and rather consistent among participants. This demonstrates that humans not only take the illumination into account when judging surface color but are also able to explicitly estimate the properties of the illumination of a given scene. The method by Koenderink and colleagues (2007) has also been successfully used to probe the representation of the illumination in real complex scenes (Kartashova, Sekulovski, de Ridder, te Pas, & Pont, 2016; Xia, Pont, & Heynderickx, 2014) and drawings (Kartashova, de Ridder, te Pas, Schoemaker, & Pont, 2015).

Here we adopt the method introduced by Koenderink et al. (2007) to explicitly measure observers' estimated illumination in #thedress photograph by asking them to adjust a sphere-probe that we placed either into the foreground or the background in the image (Figure 1) such that it would appear like a white sphere in the scene. In order to achieve their setting, observers adjusted the color of the probe, as well as the direction and intensity of the direct and the intensity of the diffuse illumination components. If differences in the perceived color of the dress were due to different estimates of the illumination in the photograph, then we would expect to find differences in the "white settings" of the probe between white and blue perceivers (Experiment 1).

The work by Winkler et al. (2015) suggests that blue things tend to be mistaken more frequently as white, due to the attribution of the bluish tint to an indirect illumination (shadow), but this tends to happen to white perceivers more frequently; and Winkler et al. attribute this to a pronounced asymmetry in the blue-yellow color labeling (see Winkler et al., 2015, for more details). Blue observers do not show this asymmetry and thus show some "resiliency" to this phenomenon. This could suggest that blue perceivers might be less sensitive to contextual information (such as illumination chromaticity priors) than white perceivers. To test this idea, we introduced subtle contextual cues in the form of incremental or decremental luminance contrast patterns on the dress (Experiment 2). If our idea was correct, we would expect blue perceivers to be less variable in their color judgments of the dress across contrast conditions than white observers

Taken together these experiments assess for the first time the perceived illumination difference hypothesis directly, and test whether the phenomenon might in part be explained by a difference in sensitivity to contextual cues.

Experiment 1

Overview

Here we measure observers' estimated illumination in #thedress photograph by asking them to adjust a sphere-probe that was placed either on the foreground or the background in the image (Figure 1) such that it would appear like a white sphere in the scene. In order to achieve their setting, observers adjusted the color of the probe, as well as the direction and intensity of the direct and the intensity of the diffuse illumination components. If differences in the perceived color of the dress were due to different estimates of the illumination in the photograph, then we would expect to find differences in the "white settings" of the probe between "white" and "blue" perceivers. We also collected color matches of the dress as in Gegenfurtner et al. 2015.

Methods

Observers

Thirty-eight observers participated in Experiment 1. All participants had seen #thedress picture before. Observers had normal or corrected-to-normal visual acuity and normal color vision. They gave written, informed consent in agreement with the local ethics committee of Giessen University and in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

Stimuli

The dress picture was presented on an OLED screen (SONY, Trimaster EL OLED). The image larger dimension (height) was equivalent to the vertical dimension of the screen, and the image was resized with a locked aspect ratio. Participants sat 62 cm from the center of the computer screen with their head stabilized by a chinrest. Under these viewing conditions, the image size was $28^\circ \times 19^\circ$ of visual angle. In Experiment 1 we used the original photograph of the dress.

The sphere probe could occur in one of two positions within the photograph: in the "foreground," right in front of the dress (approximately at the center of the image), or in the "background," approximately 8° visual angle to the right and up from the center (see Figure 1A, B). The size of the probe was 3.45° visual angle when in the foreground and was reduced to 2.12° when placed in the background, in order to achieve a more compelling impression of greater pictorial distance (Figure 1B). The reason for two using two probe locations was to verify that probe settings varied indeed with the perceived illumination in different parts of the photograph: We expected that the brightly lit area behind the dress would elicit different probe settings (at least in perceived intensity) than would the foreground location.

The probe was rendered as a matte (Lambertian) sphere on which observers could adjust (a) the overall hue and chroma, (b) the direction and the intensity of a collimated light beam directed onto the sphere, and (c) the intensity of the ambient light, resulting in six adjustable parameters.

The experiment was written by us in MATLAB using the Psychtoolbox routines (Brainard, 1997; Kleiner & Pelli, 2007).

Procedure

Prior to the adjustment task observers where asked in what color they saw the cloth and the lace of the dress (they had to choose between a blue/black and a white/gold percept). Depending on their response we labelled them as blue ($N = 20$) or white perceivers ($N = 18$) for our records and the analysis. The experiment consisted of two trials: Observers first completed the foreground and then the background condition. At the beginning of each trial, all adjustable parameters were set to randomly chosen values. The observers' task was to adjust the appearance of the sphere such that it looked like a matte white sphere embedded in the scene. We choose "white" as our target color since it has been shown that the other achromatic option "gray" varies substantially in chromaticity between observers to begin with (Witzel et al., 2011). Moreover, people may aim for different target luminance values when thinking about gray, e.g., midgray, dark gray, or light gray. By choosing a 'white' target we were aiming to narrow down this range without dictating a particular luminance value explicitly (we verified this in a separate control experiment, not reported here).

A USB numeric pad allowed the observers to adjust the color of the light illuminating the sphere. Chromatic adjustments were done in the spherical representation of DKL color space (Derrington, Krauskopf, & Lennie, 1984); thus, observers could independently adjust the hue and the chroma of the illumination on the sphere, by pressing "4" and "6," and "8" and "2" keys, respectively. In order to help participants discriminate tactilely between the two pairs of buttons, we covered them with velvety and rough surfaces, respectively. The chroma was coded so that the extremes of the adjustment scale were two opposite colors in DKL space and the neutral point was in the center. DKL coordinates were calculated as $\text{Luminance} = 0.29\ \text{R} + 0.64\ \text{G} + 0.07\ \text{B}$, $\text{red-green} = 0.75\ \text{R} - 0.57\ \text{G} - 0.17\ \text{B}$, and $\text{blue-yellow} = 0.31\ \text{R} + 0.63\ \text{G} - 0.93\ \text{B}$. The primaries of our monitor had C.I.E. xyY values of $\text{R} = (0.6751, 0.3226, 46.15)$, $\text{G} = (0.1912, 0.7263, 103.2)$, $\text{B} = (0.1413, 0.0506, 11.3)$. The direction (elevation and azimuth) of the collimated light was adjusted by moving the mouse. Elevation settings could vary between $-\pi/2^\circ$ and $\pi/2^\circ$ azimuth settings between 0° and $2\pi^\circ$ (see Figure 2). The intensity of the collimated light source was increased by pressing the left mouse button and decreased by pressing the right one. The ambient light intensity was increased and decreased by pressing the + and − keys, respectively. After completing the adjustment, observers had to press the spacebar to proceed to the next trial. Participants thought the task to be intuitive and easy.

Matching

We used the identical procedure as in Gegenfurtner et al. (2015). Observers adjusted a circular patch to match the color of the dress in the original photograph. Patch and dress photograph were displayed on the same monitor.

Analysis

We compared parameter settings between white and blue perceivers. We conducted separate analyses for (a)

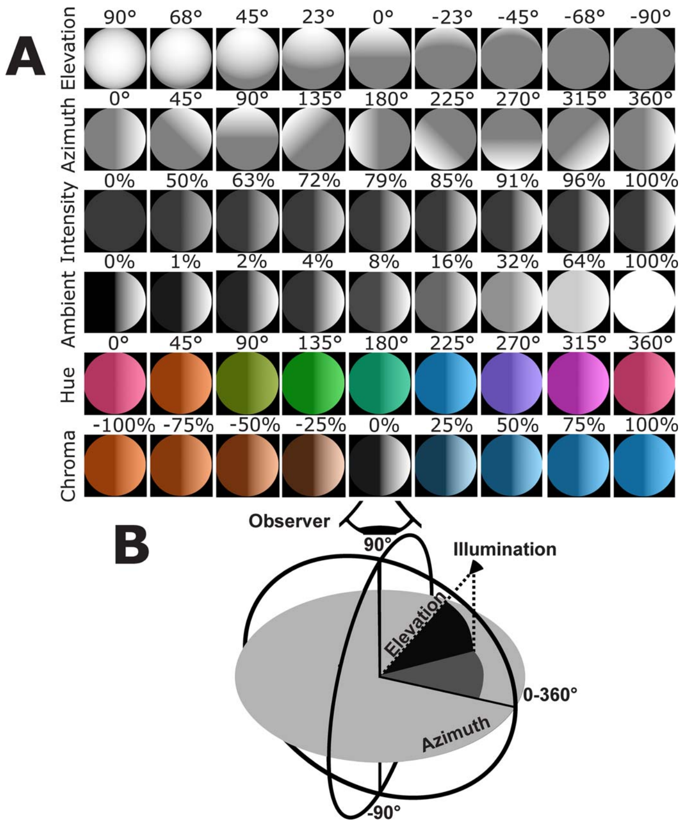

Figure 2. Effects of illumination parameter values on probe appearance. (A) Shown are the six degrees of freedom for the illumination adjustment. Participants could adjust four spatial parameters, i.e., elevation, azimuth and intensities of ambient, and collimated components and two chromatic parameters, i.e., hue and chroma. Different ratios of collimated and ambient components can result in various diffuseness levels. In our analysis we assess this diffuseness level across probe locations by computing a "punctate- diffuse ratio". (B) Representation of illumination direction in terms of azimuth and elevation related to the position of the observer.

adjusted probe chromaticity, (b) perceived collimated light source direction (elevation and azimuth), and (c) perceived collimated and ambient light intensities, described below. We also analyzed color matching data of white and blue perceivers.

Results

Overview

Figure 3 shows the average probe setting of blue and white perceivers for foreground and background conditions. It illustrates that for both types of perceivers, average settings differ between foreground and background conditions—suggesting that our experimental manipulation of probe location in the pictorial space of the photograph was indeed successful. If observers saw the sphere as a foreign body and not part of the scene, we should have seen no difference in settings between foreground and background conditions. More importantly, the figure also illustrates that white perceivers' average probe chromaticity setting appears to be more bluish than that of blue perceivers, suggesting that, indeed, the perceived illumination differed between the two groups. A detailed analysis of all measured parameters follows next.

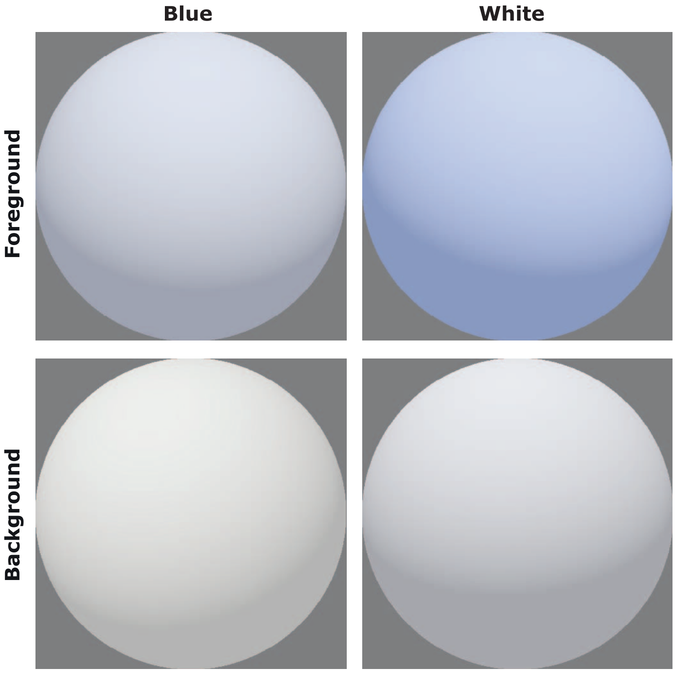

Figure 3. Average probe settings. Settings for the two perceptual groups for the sphere placed in the two positions, in front of the dress (top row) and in the back of the dress (bottom row).

Adjusted probe chromaticity

Figure 4 shows probe settings in DKL color space of blue and white perceivers at foreground (Figure 4A) and background (Figure 4B) locations. The settings for the chromatic DKL axes (blue-yellow and red-green) tended to vary along the daylight locus (continuous, colored line in Figure 4A and B). The red dashed line shows the first principal component of the combined data, illustrating its close alignment with the daylight locus. Settings along the red-green and the blue-yellow axis were highly correlated ($r = 0.81$, $p < 0.001$; $r = 0.51$, $p < 0.01$); therefore, we performed a PCA to reduce the data to a single dimension. The first principal component explained 97% and 77% of the total variance for foreground and background conditions, respectively. The obtained PCA scores along the maximum dimension of variability were used as dependent variable in a 2 (probe location) × 2 (types of

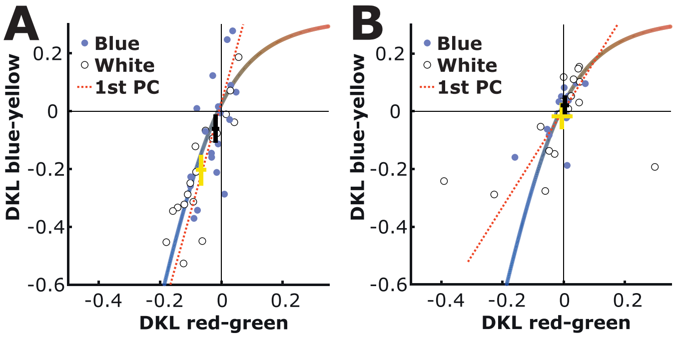

Figure 4. DKL color settings for the sphere probe. Blue circles represent the individual settings of the observers who reported the body of the dress as blue and black, and white circles represent the settings of the observers who reported it as white and gold. The colored continuous line represents the daylight locus. The red dashed line shows the first principal component of the combined data, illustrating its close alignment with the daylight locus. The horizontal and vertical lines represent the average in the setting in their interception and its standard errors along the two DKL axes: Black lines represent average settings of the observers who reported the body of the dress as blue, and gold lines represent average settings of observers who reported the body of the dress as white. (A) Settings along the red-green and the blue-yellow axis for the sphere placed in the foreground. (B) Settings along the red-green and the blue-yellow axis for the sphere placed in the background.

perceiver) mixed-design ANOVA. Overall, when the sphere was placed in the foreground, white perceivers tended to set the sphere probe to a more bluish color than blue perceivers, yielding a statistically significant main effect of both perceiver type, $F(1, 72) = 4.54$, $p < 0.04$. Moreover, for perceiver types, probe settings tended to be more bluish in the foreground than in the background, yielding a statistically significant main effect of probe location, $F(1, 72) = 12.58$, $p < 0.001$. Although the difference in chromaticity between foreground and background locations tended to be more pronounced for white perceivers, the interaction of probe location and perceiver type did not reach statistical significance ($p = 0.163$). Figure 5 shows the mean values for LM and S cone contrast and PCA scores for each condition.

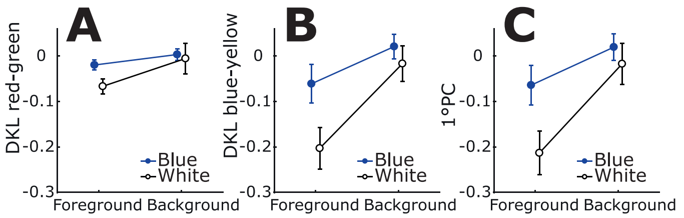

Figure 5. Mean settings along DKL chromatic axes and mean PCA values. Shown are average settings (*y* axis) of blue and white perceivers (blue and white symbols respectively) for DKL red-green axis (A) the DKL blue-yellow axis (B), and PCA scores (C). Errorbars are 1 *SE*. Although the difference in chromaticity between foreground and background locations tended to be more pronounced for white perceivers, the interaction of probe location and type of perceiver did not reach statistical significance.

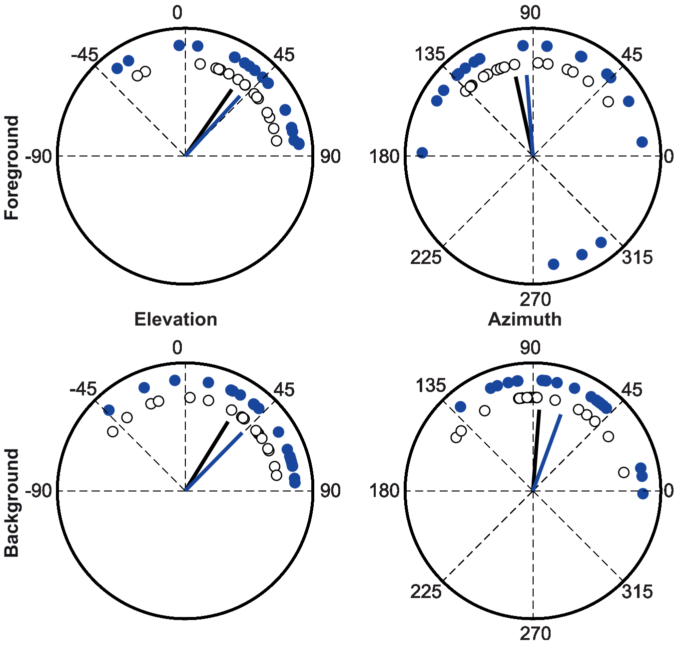

Figure 6. Individual settings and averages for the elevation and azimuth of the collimated light. The top and bottom rows show probe settings for the foreground and background conditions, respectively. Blue data points denote blue perceivers, and white data points, white perceivers (printed with slight offset for visualization purposes). Lines show vector averages of settings in each condition (blue, for blue perceivers, and black, for white perceivers).

Perceived collimated light source direction

Koenderink et al. (2007) parameterized the measured visual light field by slant and tilt (as well as intensity and directedness). Analogously, we analyze azimuth and elevation settings separately. Mean values for perceived elevation and azimuth are indicated by solid lines in Figure 6 (also see Table 1). We next tested whether observers' light direction settings were entirely random or whether they clustered around a mean

	Foreground		Background	
Blue perceivers *(azimuth)*	94.425	*13.318*	69.945	*7.536*
White perceivers *(azimuth)*	102.285	*6.654*	85.58	*7.482*
Blue perceivers *(elevation)*	42.306	*7.4286*	44.478	*7.574*
White perceivers *(elevation)*	35.671	*6.671*	31.804	*8.586*

Table 1. Mean settings for collimated light source elevation and azimuth. *Notes*: Values are shown along with the corresponding standard error.

	Foreground	Background
Blue perceivers	0.8503 *0.0183*	0.8156 *0.0263*
White perceivers	0.8580 *0.0263*	0.8601 *0.0227*

Table 2. Collimated-ambient ratio averages. *Notes*: Shown are averages and standard errors.

value. For this we conducted a Rayleigh test for nonuniformity of circular data for azimuth settings for all experimental conditions. In all cases we rejected the null hypothesis that settings were uniformly distributed along a circle (all $p < 0.0001$ and $z > 12.89$; $p < 0.013$ and $z = 4.23$ for blue perceivers in the foreground condition). In order to conduct a similar analysis for elevation settings, we first multiplied them by 2 in order to bring them into a circular range. Also for elevation data we rejected the null hypothesis that settings were uniformly distributed along a circle (all p values $<$ 0.008 and $z > 4.64$; $p < 0.043$ and $z = 3.12$ for white perceivers in the background condition). This suggests that observers perceived the direction of the illumination across probe locations in a similar manner, namely from the front and above (also see Figure 3). This provides further evidence that the probe appeared as a plausible object immersed in the scene.[1] The Harrison-Kanji test circular equivalent of a 2-way ANOVA (Harrison and Kanji, 1988) on azimuthal settings yielded a significant main effect for perceiver type, $F(1, 72) = 5.33$, $p < 0.024$ and location $F(1, 72) = 4.36$, $p < 0.04$. There were no significant differences in elevation settings between groups and conditions.

Collimated-ambient ratio

Observers in Experiment 1 also adjusted the intensities of collimated and ambient light sources. In the analysis we converted these two settings to a combined score: the so called "diffuse-punctate ratio" (Boyaci, Maloney, & Hersh, 2003). This is simply computed by dividing the intensity of the collimated component by the sum of collimated and ambient intensity (see Table 2 for averages). The 2 (perceiver type) × 2 (probe location) mixed ANOVA on the diffuse-punctate ratio yielded no statistically significant main effects and interaction. The 2 × 2 mixed ANOVA on over-all probe DKL luminance settings yielded a significant main effect for probe location, $F(1, 36) = 15.41$, $p < 0.0001$, but not for perceiver type and no significant interaction (see Table 3).

	Foreground	Background
Blue perceivers	0.8135 *0.0498*	0.9415 *0.0221*
White perceivers	0.7311 *0.0716*	0.9739 *0.0239*

Table 3. DKL luminance settings for the probe. *Notes*: Shown are averages and standard errors.

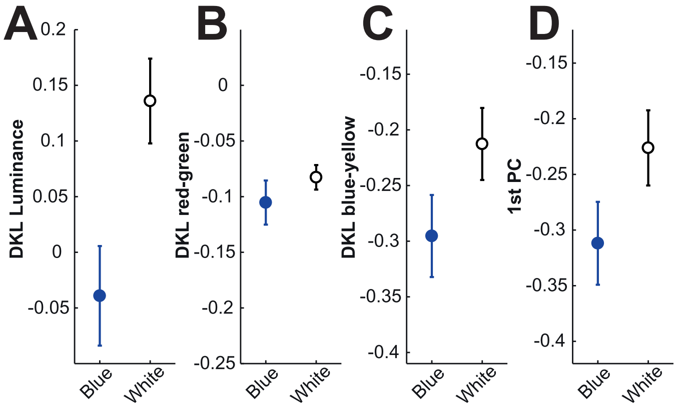

Figure 7. Color matches Experiment 1. Shown are average settings of blue and white perceivers (blue and white symbols, respectively) for DKL Luminance (A), the DKL red-green axis (B), DKL blue-yellow axis (C), and PCA scores (D). As in Gegenfurtner et al. 2015, we find only significant differences along the Luminance dimension.

Color matches

Consistent with the results of Gegenfurtner et al. (2015), we found that blue and white perceivers' color matches separate well in luminance (Figure 7A, $t = -2.9587$, $p < 0.01$) but not in chromaticity: Figure 7B–D, all $|t \text{ values}(27)| < 1.7$ all p values >0.1 in DKL color space.

Summary Experiment 1

The main finding with respect to our starting hypothesis is that white perceivers tended to perceive the illumination in the dress photograph to be more bluish than did blue perceivers. Furthermore, inspecting Figure 5, it appears that white perceivers estimated the difference in chromaticity of the illumination in the front and back of the dress scene to be more pronounced than blue perceivers—though this interaction did not reach statistical significance due to the substantial spread of the white perceiver settings. These two observations point to the possibility that blue perceivers tend to be less affected by contextual cues. We tested this idea next.

Experiment 2

Overview

In Experiment 2 we introduced contextual cues in the form of incremental or decremental luminance patterns on the original dress photo. Our aim was to find out whether white and blue perceivers differ in their

Figure 8. Stimuli Experiment 2. Six versions of the original dress photograph. Each features color texture elements that were either lower (left to right: 1–3) or higher in luminance (4–6) with respect to the dress. These textures were intended to serve as subtle contextual cues; e.g., an actual white texture might make it more difficult to perceive the dress as 'white'. Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

sensitivity to these cues, and whether this would manifest itself in unchanging color estimates of the dress for blue perceivers—across conditions. Informally, we noticed that the dress seemed to look more white with lower luminance textures and more bluish with the high luminance textures (Figure 8). Observers participated in two tasks: (a) a matching task as in Gegenfurtner et al. (2015), and (b) a naming task. We expected white—more than blue—perceivers to be sensitive to our manipulations of contextual cues.

Methods

Observers

Twelve observers from Experiment 1 participated also in Experiment 2. All participants had seen #thedress picture before. Observers had normal or corrected-to-normal visual acuity and normal color vision. They gave written informed consent in agreement with the local ethics committee of Giessen University and in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

Stimuli

Stimuli were modified versions of the original dress photograph and included colored texture elements that were either higher or lower in luminance with respect to the dress (Figure 8). These textures were intended to serve as subtle contextual cues. For example, a texture with positive contrast may make it less likely that the dress is perceived as white, since white is usually assigned to the highest luminance in a framework (Gilchrist et al., 1999); conversely, a texture with negative contrast may make the dress appear lighter. However, we would like to point out that this manipulation was rather ad-hoc and not parametric. Thus, it was an empirical question whether our manipulations had an effect at all.

Procedure

We labelled participants as white (seven) and blue perceivers (five) according to their self-reports of how they saw the original dress. There were six conditions in each of the three subexperiments—each condition corresponding to the color of the superimposed flower texture (Figure 8). The order of conditions was randomized for each observer. Participants were first asked to match the color of each dress by adjusting the color and brightness of a square patch (as in Gegenfurtner et al., 2015). After that they were asked to provide a color label for each dress condition.

Analysis

We first noted the number of times participants changed their dress color label with respect to the original dress, and compared the number of label changes across white and blue perceivers. We then aggregated data into two groups: a *negative contrast group* in which the flower textures (black, blue, purple) were darker than the original dress color and should tend to induce a white dress percept and a *positive contrast group*, in which the flower textures (white, yellow, and pink) were brighter than the original dress color and should tend to induce a blue dress percept. We then analyzed the differences of the matching data between Experiment 2 and Experiment 1, for each contrast group of Experiment 2. We expected different patterns for white and blue perceivers as detailed below. Overall, we expected that flower texture contrast manipulations would modulate white perceivers' matches more than those of blue perceivers.

Results

Color label changes

Table 4 shows the proportion of dress color label changes for each perceiver group and each *contrast*

	Negative contrast group			Positive contrast group		
	Black	Blue	Purple	White	Yellow	Pink
Blue perceivers	0.2	0.2	0.2	0	0.2	0
White perceivers	0.29	0.29	0.29	0.43	0.43	0.57

Table 4. Color label changes. *Notes*: This is computed as the number of participants who changed their color label (from white to blue or vice versa) in a given contrast group divided by the total number of participants. Our prediction was that white perceivers would tend to change their color label more frequently when flower texture cues implied a blue dress. Conversely, we expected blue perceivers to be less sensitive to our experimental manipulations, and thus change their color label less frequently. This is exactly what we found.

group. A hierarchical logistic regression analysis showed that white perceivers changed their color label more frequently with respect to their original perceived dress color than did blue observers, $t(86) = -2.2271$, $p < 0.0263$. This was consistent with our idea that blue perceivers would be less affected by the experimental manipulation of contextual cues. Moreover, the number of color label changes depended on the flower texture condition, i.e., observers tended to change color labels more frequently in those flower texture conditions that would not be consistent with their original label, $t(86) = 2.5388$ $p < 0.011$. This effect was primarily driven by the white perceivers who tended to change their dress color label to 'blue' in the white, yellow and pink conditions, i.e., in the positive contrast group.

Color matches across Experiments 1 and 2

Figure 9 plots the differences between color matches in Experiment 2 and Experiment 1 in terms of DKL luminance and chromaticity values, aggregated for each texture contrast group (positive, negative). Consider Figure 9A: Values around 0 indicate no differences between matches of Experiment 1 and 2; a positive value can be interpreted that the dress was matched higher in luminance in Experiment 2, and a negative value that it was matched as darker in luminance. White perceivers matched the dress darker in the positive contrast condition, causing a big negative difference in luminance matches. This resulted in a significant main effect of contrast group, $F(1, 6) = 25.054$, $p < 0.002$, and a significant interaction of perceiver type and contrast group, $F(1, 6) = 12.831$, $p < 0.012$. This was consistent with the fact that white perceivers changed their dress label to "blue" in the positive contrast condition, whereas blue perceivers did not change their dress color label in any condition. Interestingly, there were no differences in changes of chromaticity settings (S or L + M) across texture conditions for either perceiver group (see Figure 9B and C). There was no significant main effect of perceiver type.

Summary Experiment 2

Results of Experiment 2 show that white perceivers were more affected by contextual cues, causing a change in color percept for the dress from white to blue as indicated by changes in dress labels and luminance matches. Blue perceivers' dress labels and color matches were not affected by the introduction of contextual cues.

Overall summary

In Experiment 1 we tested the hypothesis that people who report different colors of the dress, i.e., white and blue perceivers also perceive the dress to be illuminated by different light sources. We measured observers' perception of the illumination in the photograph by having them adjust a spherical light probe—placed in the foreground and background with respect to the

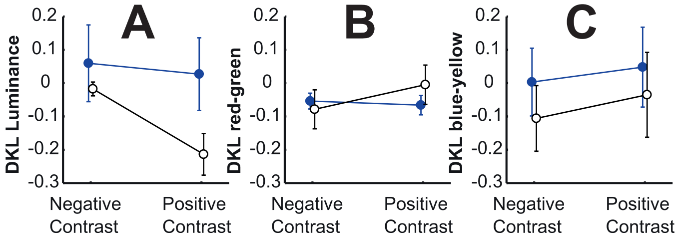

Figure 9. Color matching Experiment 2. Shown are the differences between color matches in Experiment 2 and Experiment 1 for DKL Luminance (A), DKL red-green axis (B), and DKL blue-yellow axis (C). Blue lines and symbols denote blue perceivers, and black lines and symbols denote white perceivers. Consider panel A: Values around 0 would indicate no differences between matches in Experiment 1 and 2, positive values mean that the dress was matched higher in luminance in Experiment 2, and a negative value that it was matched as lower in luminance. White perceivers matched the dress darker in the light texture condition, consistent with their dress label change to "blue". There were no differences in chromaticity settings changes across texture conditions for either perceiver group (panels B and C).

dress—such that it would appear like a white sphere embedded in the scene. Results showed that white perceivers tended to perceive the illumination in the foreground more bluish than blue perceivers. There was no difference in perceived illumination chromaticity between perceiver groups when the light probe was in the background position. We also found no differences in estimated illumination direction and intensities between groups.

In a second experiment, we tested the idea that blue perceivers might be less sensitive to contextual cues than white perceivers. We superimposed positive and negative contrast flower patterns on the dress in the original photograph, with the goal of steering the perceived color of the dress towards blue or white, respectively. For both contrast conditions, observers were asked to match the dress color and to provide color labels for the cloth. Results indicate that the positive contrast condition caused white perceivers to switch color label category from white to blue more often than did blue perceivers. Consistent with this labelling, white perceivers tended to match the dress in the positive contrast condition with a lower luminance than in the negative contrast condition. There was no change in labeling or matching across conditions for blue perceivers, suggesting that they were unaffected by the contextual cues.

Discussion

"The color-changing dress" has been an internet phenomenon in which the colors of a dress were reported to be different by different people. The first published studies aiming to explain the phenomenon (e.g., Winkler et al., 2015; Lafer-Sousa et al., 2015; Brainard & Hurlbert, 2015; Gegenfurtner et al., 2015) interpreted this striking ambiguity within the color constancy framework, which proposes that the visual system discounts illumination properties when estimating surface reflectance (Boyaci, Doerschner, & Maloney, 2006; Boyaci, Doerschner, Snyder et al., 2006; Brainard et al., 1997; Brainard & Maloney, 2004, 2011; Doerschner et al., 2004; Foster, 2011).

Here we show that blue and white perceivers differ in their assumption about the color of the illumination in the original dress photograph (Experiment 1). Our results lend direct support to the idea that blue and white perceivers see the dress in a different color because they discount different illumination colors. Taken together, our findings not only suggest that blue and white perceivers discount illumination that differ in chromaticity, but that on average, blue perceivers rely less on contextual cues when estimating surface color (Experiment 2). Several of these results stand in contrast to previous works—we will review and discuss these next.

Results in context

Our results differ from those of other studies that have investigated the perception of the dress or the perceived illumination of the scene in three ways: (a) Whereas some studies suggest that perceptual differences in perceived dress color are primarily chromaticity-based, we find differences between perceiver types mostly in terms of luminance matches; (b) conversely, whereas other studies report differences in the perceived intensity of the illumination in the dress photograph, we find primarily that perceiver types differ in their estimates of illumination chromaticity; and (c) lastly, whereas previous studies suggest that white and blue perceivers differ in how they perceive direction and position properties of the illumination in the photograph, we could not find such differences. We subsequently discuss implications of our results for two other factors that have been proposed to account for the differences in perceived dress color, namely the idea that blue surfaces tend to be seen as white and the chronotype of the observer.

Differences in chromaticity versus intensity matches for dress color

White and blue perceivers not only use different color terms to describe the dress, but in fact their perception differs. Lafer-Sousa et al. (2015) found that asymmetric color matches of blue and white perceivers differ in both, chromaticity and luminance value. On the other hand—and consistent with Gegenfurtner et al. (2015) we find only differences in luminance between perceiver types to reach statistical significance. This difference might be in part accounted for by methodological differences: Whereas Lafer-Sousa et al. (2015) used a color picker tool, we directly let participants navigate through DKL space. However, since we could not directly test their tool, it is difficult explain precisely what aspect of the task led to this discrepancy. Another difference to their study is the mere number of participants. Using an online survey, the authors tested hundreds of observers so that even small differences in color settings would reach statistical significance.

Differences in intensity versus chromaticity settings for illumination

Though it is generally believed that differences in perceived dress color depend on different assumptions that white and blue perceivers make about the illumination, there have not been any studies—except this one—directly assessing the perceived illumination in the original dress photograph. We found that while both perceiver types, equally, judged intensities of the illumination in the background to be higher and more neutral than in the foreground, only in the foreground white perceivers judged the illumination to be significantly more bluish (but not darker) than did blue perceivers. In contrast to this, other studies that have indirectly assessed the perceived illumination in the dress photograph concluded that perceived light chromaticity did not predict differences in perceived dress color; instead, that light position did: i.e., a light source perceived in the front would correlate with a blue dress color (Chetverikov & Ivanchei, 2016; Witzel, Racey, & O'Regan, in press; also see Discussion below).

In order to demonstrate that perceived illumination dictates perceived dress color, Lafer-Sousa et al. (2015) cropped the dress from the original context and placed it into a new image that was intended to provide unambiguous cues to the illumination. Their manipulation affected the color naming reports and reduced the ambiguity; however, it is not clear whether the dress was truly perceived as part of the scene and in the same illumination framework. The observed reduction in ambiguity does not necessarily mean it was caused by a difference in the assumptions about the illumination. Since both, color and intensity, of the immediate dress surround were manipulated, low-level contrast cues might in part explain their finding.

Witzel et al. (in press) controlled for such low level cues and created stimuli in which the immediate luminance background would predict the opposite of what contrast cues would predict; i.e., a dark background made the dress appear blue, or conversely a bright background made the dress appear white. Using an adjustment paradigm similar to ours—though with a difference that their probe was not embedded in the actual scene and color of the illumination components could not be adjusted—they showed that white perceivers tended to perceive a lower intensity illumination in the foreground of the dress, compared to blue perceivers. Whereas this difference in intensity setting between perceiver types is conflicting with our findings, and might be attributable to the difference in method (using a context-embedded light probe vs. a separate light probe), we do not know whether observers in their study also perceived different illumination chromaticities, since this was not assessed explicitly by Witzel et al., in press.

Illumination spatial distribution

In an online survey, Chetverikov and Ivanchei (2016) asked participants to indicate whether they thought that a light source was present in the dress photograph either in front of or behind, or to the left or right side of the dress. They found that blue perceivers were twice more likely to perceive a light source illuminating the dress from the front. In contrast in our experiment, we find no differences between perceiver types in terms of how they perceived the spatial properties of the illumination in the scene. White and blue perceivers' estimations of light direction and intensity were similar, and both perceiver types were equally likely to see the back of #the dress photograph as illuminated more than the front. There was no difference between observers as to how brightly illuminated they perceived the front of the dress. Again, these conflicting findings might in part be explained by the difference in method: Forced-choice answers on a survey might influence and restrict observers' judgments in a different way than direct estimation tasks, as the ones we used in this experiment.

One additional possibility that might help to explain differences might have to do with the perceived complexity of the illumination in the dress photograph: Light in real scenes is far more complex than a single direct component or multiple direct components. Light bounces around and gets reflected from surfaces in a scene—creating complex patterns at a given point in space (i.e., the light-field, Gershun, 1939). Using a Lambertian probe in our experiment, we can only assess the perceived low pass components of this potentially complex light field (Doerschner, Boyaci, & Maloney, 2007), but of course, that does not rule out that observers perceived a more complex lighting setup in the photograph. Thus, every task and experiment might capture a different aspect of that percept, and hence result in discrepancies between studies.

Is blue the new white?

It has been reported that surfaces whose color varies along bluish directions are more likely to be perceived as neutral as compared to surfaces that comparably vary along yellowish, reddish, or greenish directions (Winkler et al., 2015). This finding was interpreted as a tendency to attribute bluish tints to the illuminant rather than to the surface, which could in part explain why the dress image is particularly ambiguous.

Winkler et al. (2015) found large individual differences in the estimated achromatic boundaries along the blue-yellow axis. However, these boundaries did not predict observers' percepts of the dress. Witzel et al. (in press) also finds that whereas there is an overall "blue" bias in generic subjective white-point settings, the variability in generic subjective white-point settings is insufficient to explain the variability in reported dress colors.

Yet, white perceivers tended to have larger blue-yellow asymmetries in their achromatic boundaries than did blue perceivers, and Winkler et al. (2015) suggest that these asymmetries might be one factor, which contributes to the individual differences in perceiving the color of the dress.

The smaller asymmetry in blue perceivers might suggest a weaker reliance on priors about illumination properties (e.g., its chromaticity) when estimating surface reflectance. Priors are simply a different kind of contextual cue (information accumulated over time and space), and we have shown here that blue perceivers rely less on contextual cues than do white perceivers.

Thus, the idea that blue things tend to appear less colorful, because part of their hue is attributed to the chromaticity of the illumination, is potentially consistent with the results of Experiment 2—however, only for white perceivers.

Chronotypes

If illumination chromaticity priors depend on exposure, people with different circadian rhythms might have different priors (Lafer-Sousa et al., 2015). Though we cannot rule out that chronotype might play a role in whether people perceive the dress as white or blue, our results suggest that there exists a difference in observers as to what degree contextual information and prior enters their surface reflectance estimate.

Concerns

Experiment 2

We altered the original photograph such that the textile of the dress appeared to be printed with flower patterns that were either physically brighter or darker than the dress color. Assuming that the photograph can be split into several illumination frameworks (Cataliotti & Gilchrist, 1995), e.g., the foreground being one and the background being another, we reasoned that a very bright pattern (e.g., white) should make it very difficult to perceive the dress color as white, since the highest luminance in a framework—in this case our flower texture—is generally perceived as white, and there cannot be two "whites" within the same illumination framework (Cataliotti & Gilchrist, 1995; Gilchrist et al., 1999). Conversely, we thought that the addition of a dark pattern should make it more difficult to perceive the dress as dark blue, analogous to a reverse Staircase Gelb effect (Cataliotti & Gilchrist, 1995; Gilchrist et al., 1999). In addition to our luminance manipulation, we added color cues. The bright, white flower texture was much closer to the achromatic point than the perceived white of the dress in the original photograph—confronting the 'white' perceiver with conflicting information. Analogously, the dark blue flower texture would confront the 'blue' perceiver with an actual dark blue color, making it possibly harder to maintain the blue percept of the dress. Note, that for the respective other type of perceiver, there would be no conflicting information. We found that this manipulation affected white perceivers' color matches, and one might be tempted to propose that the introduction of a bright white texture onto the dress might affect white perceivers' illumination estimate, in that the illumination should appear less blue, when compared to the original photograph. We should remember, however, that our manipulation was neither systematic nor parametric, and that all the original cues to the illumination were still present in the photograph. Thus, it is difficult to predict how our manipulation should affect the illumination estimates. Certainly more investigations are needed to explore which and to what extent contextual cues are used by 'white' observers (remember, blue perceivers are less sensitive to such manipulations) when estimating surface color.

Conclusion

There is evidence that in perceiving the color of objects the visual system takes into account the context in which the surface is located (Boyaci & Maloney, 2004; Boyaci, Doerschner, & Maloney, 2006; Boyaci, Doerschner, Snyder et al., 2006; Brainard et al., 1997; Brainard & Maloney, 2004, 2011; Doerschner et al., 2004). Here we directly probe the perceived illumination in the dress photograph and relate it to perceived dress color. Our findings suggest that people who perceive the dress as blue might rely less on contextual cues when estimating surface color.

Keywords: #thedress, color vision, illumination estimation, scene interpretation, interindividual differences

Acknowledgments

This research was supported by a Sofja Kovalevskaja Award from the Alexander von Humboldt Foundation, sponsored by the German Federal Ministry for Education and Research. The work was further supported by the Deutsche Forschungsgemeinschaft SFB TRR 135 and by BMBF-NSF grant 01GQ1304.

Commercial relationships: none.

Corresponding author: Matteo Toscani.
Email: matteo.toscani@psychol.uni-giessen.de.
Address: Department of Psychology, Giessen University, Giessen, Germany.

References

Beer, R. D., Dinca, A., & MacLeod, D. I., (2006). Ideal white can be yellowish or bluish, but not reddish or greenish. *Journal of Vision*, *6*(6): 417, doi:10.1167/6.6.417. [Abstract]

Bosten, J. M., Beer, R. D., & MacLeod, D. I. A. (2015). What is white? *Journal of Vision*, *15*(16):5, 1–19, doi:10.1167/15.16.5. [PubMed] [Article]

Boyaci, H., Doerschner, K., & Maloney, L. T. (2006). Cues to an equivalent lighting model. *Journal of Vision*, *6*(2):2, 106–118, doi:10.1167/6.2.2. [PubMed] [Article]

Boyaci, H., Doerschner, K., Snyder, J. L., & Maloney, L. T. (2006). Surface color perception in three-dimensional scenes. *Visual Neuroscience*, *23*(3–4), 311–321.

Boyaci, H., & Maloney, L. T. (2004). The effect of an illuminant direction cue based on cast shadows on lightness perception in three-dimensional scenes. *Journal of Vision*, *4*(8): 121, doi:10.1167/4.8.121. [Abstract]

Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision*, *3*(8):2, 541–553, doi:10.1167/3.8.2. [PubMed] [Article]

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433–436.

Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image. 1. Asymmetric matches. *JOSA A*, *14*(9), 2091–2110.

Brainard, D. H., & Hurlbert, A. C. (2015). Colour vision: Understanding #TheDress. *Current Biology*, *25*(13), R551–R554.

Brainard, D. H., & Maloney, L. T. (2004). Perception of color and material properties in complex scenes. *Journal of Vision*, *4*(9): i, ii–iv, doi:10.1167/4.9.i. [PubMed] [Article]

Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illumination models. *Journal of Vision*, *11*(5):1, 1–18, doi:10.1167/11.5.1. [PubMed] [Article]

Cataliotti, J., & Gilchrist, A. (1995). Local and global processes in surface lightness perception. *Perception & Psychophysics*, *57*(2), 125–135.

Chetverikov, A., & Ivanchei, I. (2016). Seeing "the Dress" in the right light: Perceived colors and inferred light sources. *Perception*, *45*(8), 910–930.

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

Doerschner, K., Boyaci, H., & Maloney, L. T. (2004). Human observers compensate for secondary illumination originating in nearby chromatic surfaces. *Journal of Vision*, *4*(2):3, 92–105, doi:10.1167/4.2.3. [PubMed] [Article]

Doerschner, K., Boyaci, H., & Maloney, L. T. (2007). Testing limits on matte surface color perception in three-dimensional scenes with complex light fields. *Vision Research*, *47*(28), 3409–3423.

Foster, D. H. (2011). Color constancy. *Vision Research*, *51*(7), 674–700.

Gegenfurtner, K. R., Bloj, M., & Toscani, M. (2015). The many colours of "the dress." *Current Biology*, *25*(13), R543–R544.

Gershun, A. (1939). The light field. *Journal of Mathematics and Physics*, *18*(1), 51–151.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., & Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, *106*(4), 795.

Granzier, J. J., & Valsecchi, M. (2014). Variations in daylight as a contextual cue for estimating season, time of day, and weather conditions. *Journal of Vision*, *14*(1):22, 1–23, doi:10.1167/14.1.22. [PubMed] [Article]

Harrison, D., & Kanji, G. K. (1988). The development of analysis of variance for circular data. *Journal of Applied Statistics*, *15*(2), 197–223.

Hurlbert, A. (1999). Colour vision: Is colour constancy real? *Current Biology*, *9*(15), R558–R561.

Kartashova, T., de Ridder, H., te Pas, S. F., Schoemaker, M., & Pont, S. C. (2015). The visual light field in paintings of Museum Prinsenhof: Comparing settings in empty space and on objects (pp. 93941M–93941M). Presented at the SPIE/IS&T Electronic Imaging, International Society for Optics and Photonics, March 2015, San Francisco, CA.

Kartashova, T., Sekulovski, D., de Ridder, H., te Pas, S. F., & Pont, S. C. (2016). The global structure of the visual light field and its relation to the physical light field. *Journal of Vision*, *16*(10):9, 1–16, doi:10.1167/16.10.9. [PubMed] [Article]

Kleiner, M. B., & Pelli, D. D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(1), 1–16.

Koenderink, J. J., Pont, S. C., van Doorn, A. J., Kappers, A. M., & Todd, J. T. (2007). The visual light field. *Perception*, *36*(11), 1595–1610.

Lafer-Sousa, R., Hermann, K. L., & Conway, B. R. (2015). Striking individual differences in color perception uncovered by "the dress" photograph. *Current Biology*, *25*(13), R545–R546.

Maloney, L. T. (1999). Physics-based approaches to modeling surface color perception. In *Color vision: From genes to perception* (pp. 387–416). Cambridge, UK: Cambridge University Press.

Melgosa, M., Gómez-Robledo, L., Isabel Suero, M., & Fairchild, M. D. (2015). What can we learn from a dress with ambiguous colors? *Color Research & Application*, *40*(5), 525–529.

Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., & Hurlbert, A. (2014). Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. *PloS One*, *9*(2), e87989.

Winkler, A. D., Spillmann, L., Werner, J. S., & Webster, M. A. (2015). Asymmetries in blue–yellow color perception and in the color of "the dress." *Current Biology*, *25*(13), R547–R548.

Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner, K. R. (2011). Object knowledge modulates colour appearance. *I-Perception*, *2*(1), 13–49.

Witzel, C., Racey, C., & O'Regan, J. K. (in press). The most reasonable explanation of "the dress": Implicit assumptions about illumination. *Journal of Vision*, in press.

Wyszecki, G., & Stiles, W. S. (1982). *Color science* (Vol. 8). New York, NY: John Wiley.

Xia, L., Pont, S. C., & Heynderickx, I. (2014). The visual light field in real scenes. *i-Perception*, *5*(7), 613–629.