Seeing "the Dress" in the Right Light: Perceived Colors and Inferred Light Sources

Andrey Chetverikov
Faculty of Psychology, University of Iceland, Reykjavik, Iceland;
Department of Psychology, Saint Petersburg State University, Saint
Petersburg, Russia; Cognitive Research Lab, Russian Academy of
National Economy and Public Administration, Moscow, Russia

Ivan Ivanchei
Department of Psychology, Saint Petersburg State University, Saint
Petersburg, Russia

Abstract

In the well-known "dress" photograph, people either see the dress as blue with black stripes or as white with golden stripes. We suggest that the perception of colors is guided by the scene interpretation and the inferred positions of light sources. We tested this hypothesis in two online studies using color matching to estimate the colors observers see, while controlling for individual differences in gray point bias and color discrimination. Study 1 demonstrates that the interpretation of the dress corresponds to differences in perceived colors. Moreover, people who perceive the dress as blue-and-black are two times more likely to consider the light source as frontal, than those who see the white-and-gold dress. The inferred light sources, in turn, depend on the circadian changes in ambient light. The interpretation of the scene background as a wall or a mirror is consistent with the perceived colors as well. Study 2 shows that matching provides reliable results on differing devices and replicates the findings on scene interpretation and light sources. Additionally, we show that participants' environmental lighting conditions are an important cue for perceiving the dress colors. The exact mechanisms of how environmental lighting and circadian changes influence the perceived colors of the dress deserve further investigation.

Corresponding author:
Andrey Chetverikov, Department of Psychology, University of Iceland, Saemundargata 2, 101 Reykjavík, Iceland.
Email: andrey@hi.is

Introduction

The perception of color does not depend only on the physical characteristics of stimuli. An interesting example of this is an ambiguous photo of a harshly illuminated dress ("The dress (viral phenomenon)," 2015). While some viewers perceive the dress as white with gold stripes, others see it as blue-and-black. The perception of colors is quite stable although anecdotal reports suggest that it may switch when observers see the image again after some time. If these reports are correct, "the dress" poses a considerable challenge for vision science. While some color illusions create strikingly discrepant perceptions of physically equal stimuli (Lotto & Purves, 1999), the dress is perceived differently by different (or even the same) observers without additional manipulations of context or other factors. And unlike already known bistable phenomena (e.g., the Necker cube, Schwartz, Grimault, Hupe, Moore, & Pressnitzer, 2012), the changes in perceived interpretation of the dress are rare and limited to a part of the population. The uniqueness of the dress image makes it worth studying as it may provide insights in the mechanisms of color vision and perceptual stability.

We suggest that the crucial element of this illusion is the inferred position of the light sources in the image. When estimating colors, the visual system relies not only on a particular patch of color but also on the image as a whole (Shevell & Kingdom, 2008). Sources of illumination are particularly important because the same object could look different under different lighting. Observers are capable of discounting the effects of illumination, a phenomenon known as color constancy (Foster, 2011). Although estimation of the illuminant may not always help or even be necessary (Granzier, Brenner, & Smeets, 2009; Granzier & Valsecchi, 2014), it might provide a cue for the resolution of ambiguity and help to interpret natural scenes (Boyaci, Doerschner, Snyder, & Maloney, 2006; Yang & Shevell, 2003).

The hypothesis that the ambiguity of the dress perception is related to the discounting the effects of illumination was suggested by Conway in an interview for Wired magazine (Rogers, 2015). The results of a study reported later by Lafer-Sousa, Hermann, and Conway (2015) confirmed that intuition. Introducing an obvious cue for the interpretation of the lighting conditions shifted observers' perception in the predicted way: Under a cool illuminant observers perceived the dress as white-and-gold, while with a warm illuminant the dress was more likely to be interpreted as blue-and-black. They also measure perceived colors of the dress using a color-matching procedure and found that the distribution of the colors indeed differed between the observers who interpreted the dress differently with notable peaks for different labels. Gegenfurtner, Bloj, and Toscani (2015) also discuss the importance of luminance assumptions for the dress perception. However, they suggested that the distribution of colors is unimodal rather than multimodal with different labels corresponding mainly to differences in lightness.

In the present study, we were mainly interested in the assumptions of observers about the direction of the illumination. If the dress is illuminated from the front, for example, by a camera flash, then it may be overexposed. If overexposure is discounted, the perception of colors might be biased toward dark colors leading to the perception of the blue-and-black dress. If there is, on the other hand, a strong light source behind the dress, such as sunlight through a window, then the dress might be underexposed. Observers discounting underexposure might perceive colors lighter than the real ones, resulting in a white-and-gold impression. This hypothesis is consistent with the explanation proposed by Lafer-Sousa et al. (2015). However, while Lafer-Sousa et al. (2015) assessed the effects of the illuminant color, we were mostly interested in the direction of illumination. Moreover, we believe that observers' assumptions by themselves provide an important way to test the hypotheses regarding the sources of individual differences in perception.

After a pilot study in which we quizzed observers about the sources of the illumination in the image (Appendix), we ran an online study that combined color-matching procedure with tests of color discrimination, tests for biases in gray point, and questions about the interpretation of light sources and the scene. In addition, we analyzed circadian changes in perceived color of the dress. We assumed that both the spectrum of light and amount of light varies during the day and may influence the colors of objects (Judd et al., 1964; Romero, Hernández-Andrés, Nieves, & García, 2003). Accordingly, it could affect the estimated light sources as well as the perceived colors of the dress.

Study 1

Methods

Participants. One thousand one hundred and eighteen participants completed the quiz (participants who did not complete the quiz fully or provided incorrect data were excluded beforehand) during the period from March 2, 2015 to March 30, 2015. Their reported age was from 10 to 75 years (eight participants did not report their age), 454 men and 657 women (seven selected "Other" as their gender). The study was available in English ($N = 337$) and in Russian ($N = 781$). The data on gray point perception biases were collected from 449 participants.

Procedure and materials. The study was conducted with JavaScript following a pilot study that used Google Forms for the data collection (see Appendix). All colors and their transformations were based on the assumption that sRGB color space is used, which is the default option for all modern web browsers (Çelik, Lilley, & Baron, 2011). The transformations to Lab colors were made with the *convertColor* function from *grDevices* package in R (R Core Team, 2015) assuming the standard illuminant D65.

The study was presented on a white background. The originally published photo of the dress was resized to 300×455 pixels. The image was positioned in the left part of the screen, the questions and instructions were presented later. Next to a photo of the dress, the following instructions were presented:

> People perceive colors of this dress differently. We aim to gather some data on this phenomenon to understand it better. We are interested in your perception and not in the 'true' colors, and there are no right and wrong answers.

Then, participants selected the color of the two sets of stripes by picking a color with a standard hue, saturation, and lightness (HSL) color picker. A separate scale for hue, and a 2D space for saturation and lightness were presented for each set of stripes. To avoid color priming, we defined two sets of stripes as "the stripes with grating (the ones that are similar to the sleeves)" (further described in text as white/blue stripes) and "the other stripes" (further described in text as gold/black stripes). We asked participants to state "what does the color of... look like" to emphasize that we were interested in the perception of colors and not in the "true" colors of the dress.

We used a simple color discrimination test that mimicked the standard Farnsworth-Munsell 100 hue test and other computer color deficiency tests based on hue discrimination (Melamud, Simpson, & Traboulsi, 2006; Shin, Park, Hwang, Wee, & Lee, 2007); 72 circles with equidistant hues in an HSL color space (saturation and lightness were set to 0.5, the average value) were presented in three rows of 24. In addition, we included a separate row of 24 circles ranging from black to white with equidistant lightness. The first circles in each row were the same for each participant; all other circles were shuffled. For each

row, participants sorted the circles from left to right by dragging them so that the next circle was, to them, most similar to the previous one (e.g., Figure S4 and additional details).

We then tested bias in the perception of a gray point. Participants had to make the presented square (40×40 pixels) "as neutral (gray) as possible" using two sliders. The sliders controlled the *a* and *b* channels of Lab color space with lightness preset to average value. We added a random constant shift on each channel ranging from −20 to 20 so that simply setting both sliders to a middle point would not work.

Then, we asked participants the following questions about the light sources in the photo with the possible answers "Yes"—"Not sure"—"No":

(1) Is there a light source in front of the dress (from the photographer's side)?
(2) Is there a light source behind the dress?
(3) Is there a light source to the right of the dress?
(4) Is there a light source to the left of the dress?

We subsequently asked, "Is it a window or a mirror behind the dress?" with the possible answers being "window," "mirror," and "other." Then, participants were asked whether "this dress is more of a . . ." with two options, white-and-gold and blue-and-black, shown in randomized order to avoid "default option" bias. The remaining questions were:

(1) Have you seen this image earlier (before taking this questionnaire)? (Yes or No)
(2) Does your perception of the dress change from time to time? (Yes or No)
(3) If it does change, then how often? (Open Question)
(4) If you are able to see different versions of the image, can you voluntarily switch from one perception to another? (Yes or No)
(5) Your age
(6) Your gender (M, F, or Other)
(7) Did you fill this questionnaire before? (Yes or No)

In addition, the information about the participant's city was collected on the basis of the information provided by the web browser (IP address). After the study, participants received feedback on their accuracy in the color discrimination test and were shown the colors chosen by others.

Results

Perceived colors. First, we verified whether there are measurable differences in color perception when people do not need to name the color. The palette of matched colors (Figures S1, S2) shows that people do indeed perceive the dress differently (see Table 1 for descriptive statistics). Analysis of colors in CIELAB color space shows that the "white/blue" stripes of the white-and-gold dress are lighter, $t(1062.7) = 36.23$, $p < .001$, $d = 2.18$[1] less bluish, $t(1111.6) = 34.04$, $p < .001$, $d = 2.02$, and less greenish, $t(901.5) = 17.40$, $p < .001$, $d = 1.01$, than the same stripes of the blue-and-black dress. The "gold/black" stripes are also lighter, $t(1098.1) = 36.52$, $p < .001$, $d = 2.19$ and less bluish, $t(1023.2) = 34.39$, $p < .001$, $d = 2.08$, in white-and-gold perception than in blue-and-black (Figure 1).

The perceived colors of the dress change more often among people who see it as blue-and-black, $\chi^2(1, N = 1116) = 5.18$, $p = .023$, Cramer's $V = .070$ (Table 2). Analysis of the distances in color space demonstrated that people whose perception of the dress changes from time to time see the dress less similarly to other people in their group ($t(367.4) = -4.05$, $p < .001$,

Table 1. Descriptive Statistics on Perceived Dress Colors in Study 1.

		Blue-and-black						White-and-gold					
		R	G	B	L	A	b	R	G	B	L	a	b
"White/blue" stripes	M	98	127	203	54	13	−42	193	203	223	81	1	−11
	SD	39	36	34	12	14	17	41	35	29	13	7	13
"Gold/black" stripes	M	73	63	48	27	3	10	168	137	61	59	4	42
	SD	40	35	33	15	8	15	39	37	39	14	10	16
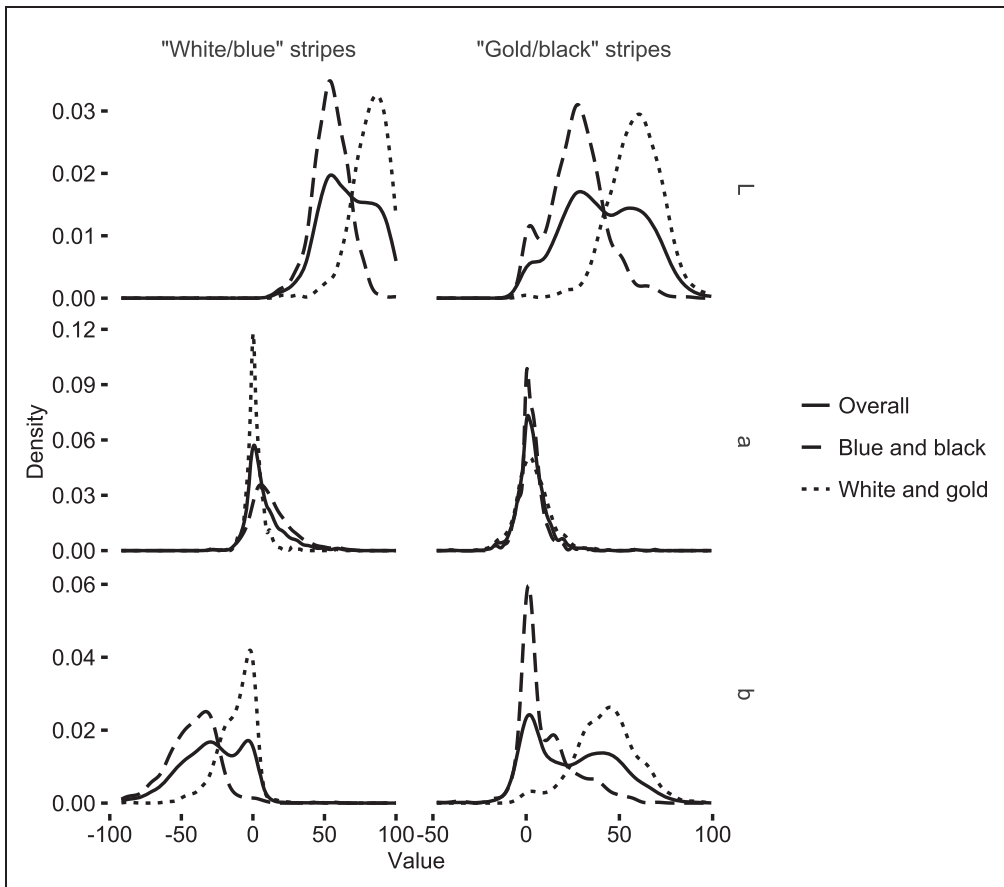
Figure 1. Distribution densities for perceived colors of the dress in Study 1, CIELAB color space.

Table 2. Age, Gender, and Stability of Perception in Study 1.

	Women		Men		Age		
	N	%	N	%	M	SD	% perception changes
Blue-and-black	360	55	252	56	27.4	8.7	27
White-and-gold	297	45	202	44	27.7	9.0	21

$d = 0.30$ and $t(367.7) = -2.90$, $p = .004$, $d = 0.22$ for "white/blue" stripes and "gold/black" stripes, accordingly) and more similarly to the people in the other group ($t(400.8) = 2.12$, $p = .034$, $d = 0.15$ and $t(410.0) = 3.15$, $p = .002$, $d = 0.23$, accordingly).

The distribution of perceived colors in case of the b channel is not unimodal (Hartigans' dip test for unimodality (Hartigan & Hartigan, 1985; Maechler, 2014) for the white/blue stripes, $D = 0.02$, $p = .007$; for the gold/black stripes, $D = 0.02$, $p < .001$, Figure 1). Although the distribution of perceived colors on L channel also looks bimodal, the results of the dip test did not support this conclusion, $D = 0.01$, $p = .951$ for white/blue stripes and $D = 0.01$, $p = .228$ for gold/black stripes. An alternative approach, based on likelihood ratio of unimodality against bimodality (Holzmann & Vollmer, 2008; Schwaiger, Holzmann, & Vollmer, 2013), indicated that bimodal distribution provides a better fit both for the L channel, $LR = 12.08$, $p < .001$ and $LR = 11.04$, $p < .001$, and the b channel, $LR = 69.72$, $p < .001$ and $LR = 408.71$, $p < .001$ for white/gold and blue/black stripes, respectively. Neither of the tests indicated that distribution on a channel is unlikely to be unimodal.

Age and gender. We did not find any significant relation between dress categorization and age, $t(1049.4) = -0.56$, $p = .577$, $d = 0.03$, or gender, $\chi^2(1, N = 1111) = 0.03$, $p = .862$, $V = .010$ (Table 2). However, there were only 88 participants older than 39 years, and the lack of any findings on age-related differences may be due to sample limits.

Color discrimination. No significant differences were found for accuracy of color discrimination test, $p > .3$. Error scores for participants in the white-and-gold group were 16.98 ($SD = 46.90$) and 19.39 ($SD = 46.43$) for white-black scale and for other colors, respectively. For participants in the blue-and-black group, they were 16.91 ($SD = 46.12$) and 21.77 ($SD = 49.34$), respectively. No correlations between dress colors values on different channels and error scores were found.

Observers' gray point. No significant differences were found for the observers' gray point. Color settings for neutral gray square did not differ between the groups (a channel: $t(390.9) = -0.59$, $p = .556$, $d = 0.06$; b channel: $t(401.0) = 0.28$, $p = .777$, $d = 0.03$). No correlation was found between the amount of bias on either of channels and perceived colors of the dress stripes (Pearson's r, all $ps > .130$).

Inferred light sources. People who saw the blue-and-black dress were more likely to agree that there is a light to the left, $\chi^2(2) = 19.75$, $p < .001$, $V = .130$, but not to the right of the dress, $\chi^2(2) = 0.56$, $p = .755$, $V = .020$, and less likely to agree that a light source was behind the dress, $\chi^2(2) = 25.24$, $p < .001$, $V = .150$ (Figure 2). But the most striking differences were related to the light in front of the dress; 19% of white-and-gold group agreed and 65% disagreed that there is a light in front, as compared to 47% and 33% in the blue-and-black group, $\chi^2(2) = 125.43$, $p < .001$, $V = .330$. That is, people in the blue-and-black group were twice more likely to see the dress as illuminated from the front.

Background interpretation. A total of 65% of observers in the blue-and-black group interpreted the background of the photo as a mirror, while 29% described it as a window as opposed to 54% and 36% in the white-and-gold group, $\chi^2(2) = 14.37$, $p < .001$, $V = .110$. This result was consistent with the interpretation of light sources: Those who believed that there was a light in front of the dress was less likely to interpret the background as a window (binomial regression, $B = -0.22$, $Z = -2.30$, $p = .021$).
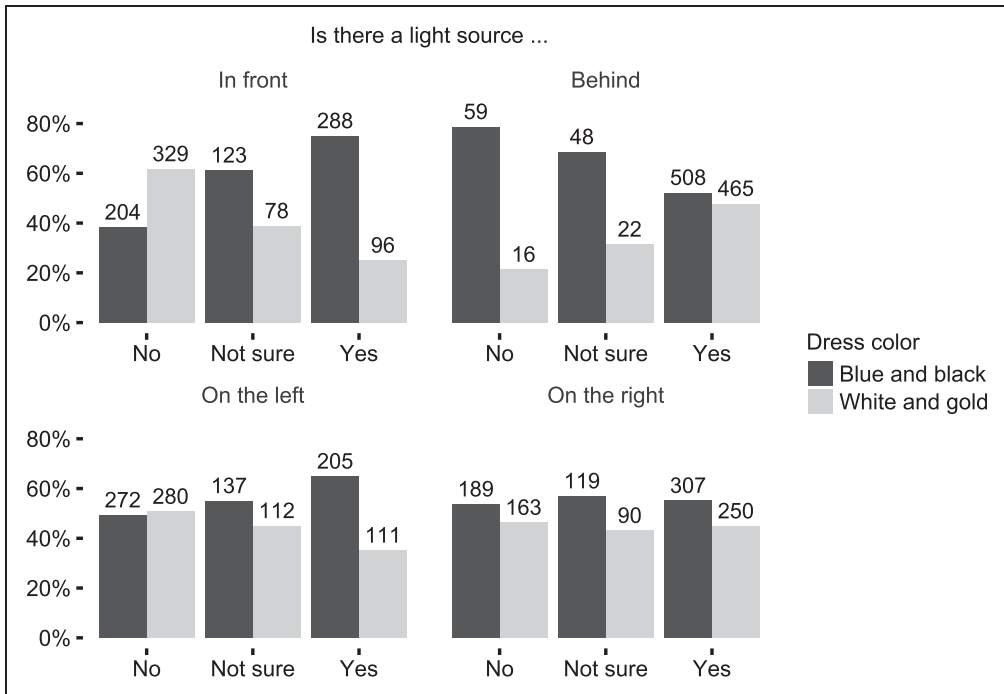
Figure 2. Inferred light sources in "the dress" photo in Study 1. The numbers on top of the bars show responses frequencies.

Circadian changes. Because the spectrum of light may vary with geographical latitude, amount of light pollution, and weather conditions, and so on (Hernández-Andrés, Romero, Nieves, & Lee, 2001), we analyzed the data from Moscow, the city most represented in our sample ($N = 318$). The dress was more often perceived as blue-and-black during the evening and night (6 p.m.–6 a.m.) and as white-and-gold during the morning and day (6 a.m.–6 p.m.), as indicated by a quadratic effect obtained from binomial regression with time of the day as a predictor, $B = -6.10$, $SE = 2.09$, $Z = 2.92$, $p = .003$ (Figure 3). The time of day also affected the inference regarding the frontal light source. The proportion of "yes" responses to the presence of frontal lighting in each period was higher in the evening and night periods compared to morning and day periods (binomial regression $B = 5.16$, $SE = 2.06$, $Z = 2.51$, $p = .012$). Analysis of the separate channels of the matched colors with linear regression indicated a quadratic effect of time of day both for lightness, $B = -72.49$, $SE = 23.02$, $t = -3.15$, $p = .002$, $R^2_{part} = .02$, and for the b channel, $B = -80.09$, $SE = 34.38$, $t = -2.33$, $p = .020$, $R^2_{part} = .01$, with lighter and more yellowish perception of the dress during the day (Figure 4). Mediation tests indicated that answers about the frontal light source mediated 21% of the effect of time of day for both channels. Time of the day was not related to the observers' gray point. The data from the rest of the sample analyzed in the same way as the data from Moscow sample did not show any significant effects.

Discussion

Our results show that the interpretation of the dress colors is not simply a naming difference. Rather, observers actually do perceive it differently and those who are in between the
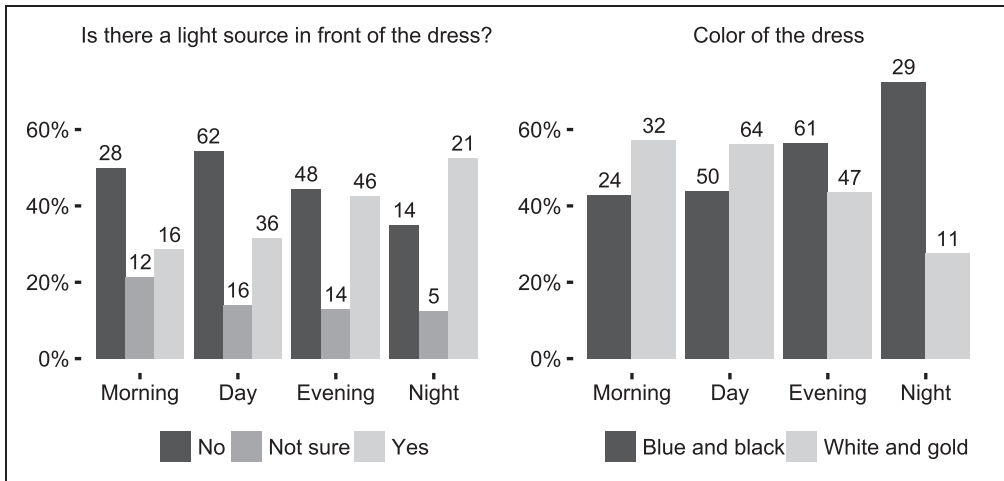
Figure 3. Changes in inferred illumination and perceived dress color as function of time of the day.
Note. Morning—6 a.m. to noon, day—noon to 6 p.m., evening—6 p.m. to midnight, night—midnight to 6 a.m.
The numbers on top of the bars show responses frequencies.
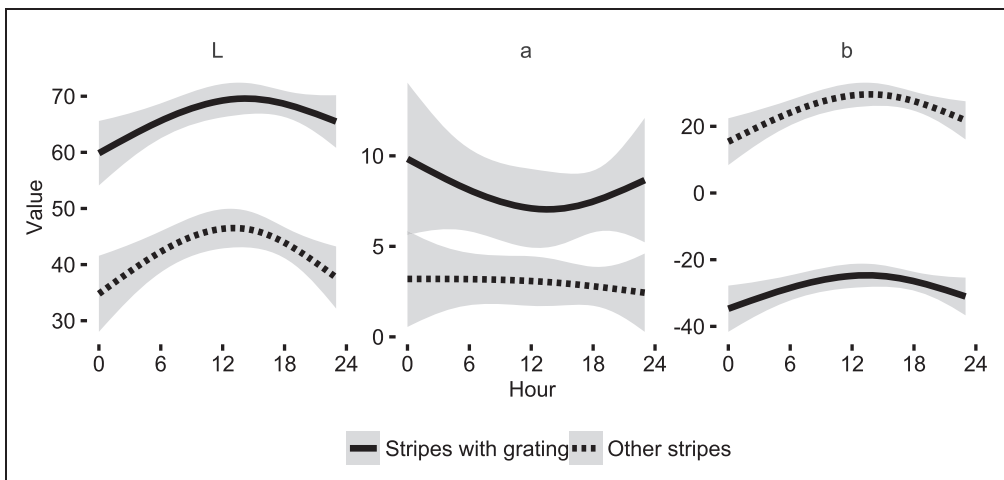
Figure 4. Color values in CIELAB color space as function of the time of day. The smoothed estimates are based on the regression model with linear and quadratic effects of the time of the day as predictors.

blue-and-black and white-and-gold camps have corresponding intermediate perception of colors. The perceived colors are consistent with inferred light sources: The blue-and-black group is twice as likely to agree that there is a light source in front of the dress supporting the discounting of overexposure explanation. Not only do the inferred light sources differ—the interpretation of the background as either a window or a mirror is also consistent with an effect on inferred light sources on perceived colors. Putting both factors together, only 10% of the blue-and-black group said that there was a window and no front light, while only 11% of white-and-gold group agreed that there is a front light and that the background is a mirror.

Study 1 also demonstrated that the perceived colors of the dress depend on the time of the day. The dress is most often perceived as blue-and-black during the evening and night hours (from 6 p.m. to 6 a.m.). The changes in the reported interpretation also correspond to the circadian changes in the matched colors indicating that observers see the dress differently and not simply interpret its colors differently during different periods of the day. The observed pattern of changes can be interpreted in several ways. First, the spectrum of light varies during the day with "warmer" light when the sun is near the horizon. The peak of the changes in perceived colors of the dress is close to the noon and may stem from the discounting of the "cold" daylight. It is also possible that the perceived color is influenced by artificial sources of light. After sunset, which occurs in Moscow at 6 p.m. in the beginning and at 7 p.m. in the end of March, participants' perception of the dress may depend on the spectrum of the lamps used. If most of lamps used provide "warmer" light than daylight, participants may discount this "warm" illuminant and hence perceive the dress as "black and blue." Finally, the amount and direction of light rather than its spectrum may influence the interpretation of the dress. At night, if the computer display is the major source of illumination, participants might interpret the dress as overexposed, because its surrounds are brightly lit compared to the rest of the environment. In contrast, during the day there is a higher chance of illumination from behind the observer due to the sunlight coming through the windows. Moreover, the ambient light is more diffuse reducing the importance of display as a source of the illumination. Although we favor the latter explanation, all three of them can contribute to the resulting pattern of changes.

Study 2

While the findings of Study 1 are promising, the extent of their validity is limited by unknown colorimetric characteristics of the displays used by participants. Unlike laboratory studies, where one can use a colorimeter to ensure the quality of color representation, large-scale web studies are limited in the control over participants' displays. Most of the displays are not calibrated, their physical dimensions and distance from participant to display are unknown. To assess the accuracy of our color-matching procedure, in Study 2 we asked observers to participate at least twice using different devices. This allowed us to estimate the reliability of color-matching procedure we used. If display characteristics play an important role in the matched colors, then we would expect to observe little or no correlation between responses on different devices. If, on the other hand, they are not important, then the colors matched by the same observer on different devices should be similar. Such design also allows us to estimate the importance of display size for the perception of the dress in online studies. Lafer-Sousa et al. (2015) found that increasing the image size increases the probability of white-and-gold perception. However, it is unclear if this factor actually influences the colors perceived by non-naïve observers when they are able to choose the most convenient distance to the display as in Study 1.

The second goal of this study was to assess the role of environmental lighting conditions and the ability of observers to make inferences regarding the lighting conditions in the dress photo. Different light sources have different spectrums and the dominant wavelengths in the light spectrum may bias the perception of the dress providing a cue for its interpretation. The differences in light sources might also explain the circadian variations observed in Study 1.

Methods

Materials and procedure. The study followed the same procedure as Study 1 except that the color discrimination test was removed and several new questions were introduced. First, two

questions about the environmental lighting conditions and the lighting conditions in the photo were added. The participants were asked "What are the lighting conditions around you now?" and "What are the lighting conditions in the photo?" with seven possible answers: "daylight" (natural lighting), "luminescent lamp," "incandescent lamp," "halogen lamp," "LED lamp," "display light only" (only for the question about environmental lighting), and "other." The order of answers was randomized for each participant. Second, we asked participants to select the type of device they used ("desktop," "laptop," "tablet," "smartphone," and "other"). Then they were asked to provide the name of the device (or, in case of desktop, the name of the display), its diagonal size (in inches), width and heights (in centimeters, these parameters were marked as optional), and screen resolution (detected automatically, but observers were able to correct it if the automatic detection did not work). Finally, we asked participants, which city they are from to avoid relying on geolocation.

Participants. Three hundred ninety one participants completed the quiz during the period from June 15, 2015 to June 25, 2015. Their reported age was from 17 to 53 years (36 participants did not report their age), 135 men, 246 women; 113 participants participated in the study at least twice using different devices, but only 11 of them participated more than two times. Thus, we did not analyze the data on third and later attempts separately. The distribution of responses by device type was as follows: Desktop or TV display—126, Laptop—161, Smartphone—158, Tablet—58, Other—3. Most of the participants were from the Moscow region ($N = 191$ responses) or from the St. Petersburg region ($N = 238$ responses). The rest of the sample ($N = 86$) were from a number different cities with less than 15 responses from each city.

Results

Age and gender. We analyzed the data from those participating in the study for the first time. Contrary to the results of the first study, we did find a significant relation between dress categorization and age, $t(338.6) = -2.60$, $p = .010$, $d = 0.27$. People who saw the dress as white-and-gold were on average older than those who saw it as blue-and-black (Table 3). Gender was related to the interpretation of the dress colors at a tendency level, $\chi^2(1) = 3.20$, $p = .074$, $V = .090$.

Perceived colors. The results on perceived colors are similar to the findings of Study 1. Analysis of colors in CIELAB color space shows that the white/blue stripes of the white-and-gold dress are lighter, $t(510.6) = 30.18$, $p < .001$, $d = 2.64$, less bluish, $t(503.5) = 25.92$, $p < .001$, $d = 2.26$, and less greenish, $t(375.6) = 12.23$, $p < .001$, $d = 1.04$, than the same stripes of the blue-and-black dress. The gold/black stripes are also lighter, $t(512.0) = 26.14$, $p < .001$, $d = 2.31$, and less bluish, $t(469.8) = 23.01$, $p < .001$, $d = 2.05$, for white-and-gold perception

Table 3. Age, Gender, and Stability of Perception in Study 2.

|	Women		Men		Age		
	N	%	N	%	M	SD	% perception changes
Blue-and-black	119	48	79	59	25.0	4.7	24
White-and-gold	127	52	56	41	26.4	5.6	20

than blue-and-black (Figure 7). A likelihood ratio test of unimodality against bimodality (Holzmann & Vollmer, 2008; Schwaiger et al., 2013) indicated that bimodal distribution provides a better fit both for L, $LR = 14.39$, $p < .001$ and $LR = 4.05$, $p = .022$ for white/blue and gold/black stripes, and b channels, $LR = 17.34$, $p < .001$ and $LR = 162.78$, $p < .001$, respectively. No significant differences in the a channel of perceived colors were found.

Perception of the dress on different devices. Only three participants changed their opinion regarding the dress color when they participated in the study the second time. Figure 5 shows that the matched colors were also quite consistent. The perceived colors on the first and the second pass through the test were correlated: $r(111) = .78$, $p < .001$ and $r(111) = .73$, $p < .001$ for the L channel, $r(111) = .40$, $p < .001$ and $r(111) = .50$, $p < .001$ for the a channel, and $r(111) = .70$, $p < .001$ and $r(111) = .85$, $p < .001$ for the b channel of the first and the second sets of stripes, respectively. That is, the a channel which was almost the same for the two variants of perceived colors was the most device-dependent. On the other hand, L and b channels that separated the observers in Study 1 were reliable according to traditional guidelines on test–retest reliability (Kaplan & Saccuzzo, 2009)

We then estimated the impact of image size on the perception of the dress. Answers with sizes below 3.1° or above 16.1° of v.a. corresponding to the top and bottom 2.5% of the data were removed as outliers. Figure 6 shows a slight trend toward decreasing chances of white-and-gold perception with increasing image size (in degrees of visual angle). However, the binomial regression indicated that this effect was not significant, $B = -0.03$, $SE = 0.04$,
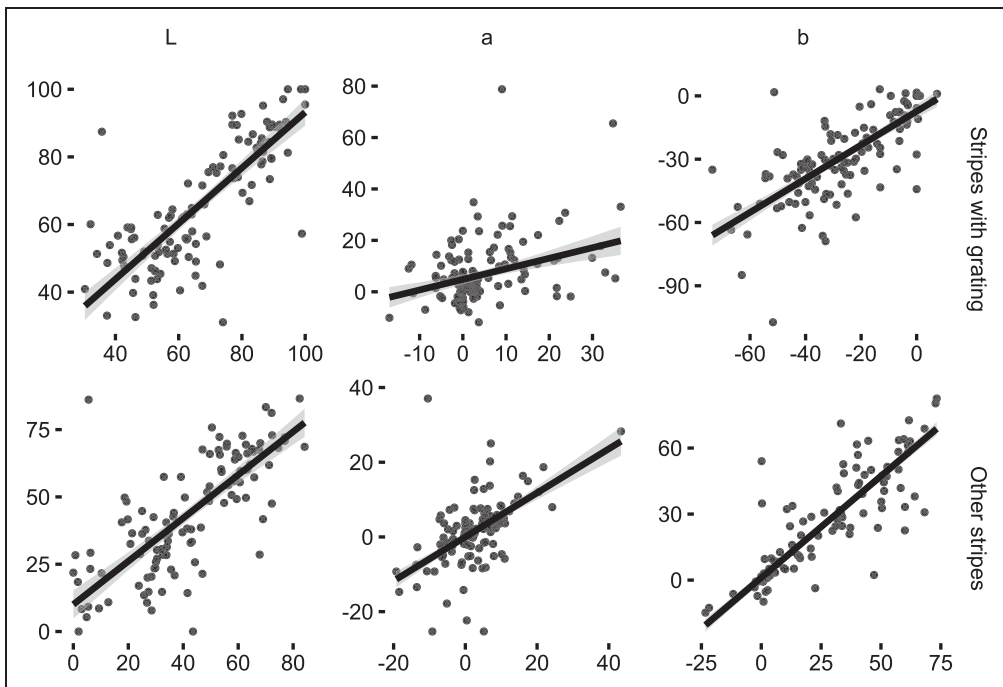
Figure 5. Correlations between perceived colors on the first (x-axis) and the second (y-axis) measure in the Study 2 for the same observers using different devices. Each dot shows the data from one participant.
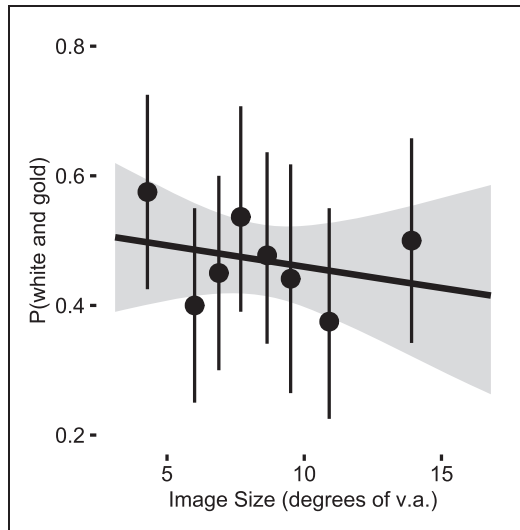
Figure 6. Probability of white-and gold-perception as function of the dress image angular size. Line shows fitted binomial regression prediction (with shaded area showing 95% confidence interval), dots show binned means with 95% confidence intervals as bars.

$Z = -0.82$, $p = .411$. Similar results were obtained when we used stimulus size in centimeters as predictor, $B = -0.03$, $SE = 0.04$, $Z = -0.74$, $p = .460$. A comparison of responses by device type also did not indicate any association between dress color and device, $\chi^2(3, N = 503) = 1.53$, $p = .675$, $V = .060$.

Inferred light sources. Replicating the results of the first study, observers who saw the dress as blue-and-black were more likely to agree that there was a light source in front, $\chi^2(2) = 35.84$, $p < .001$, $V = .270$, and on the right, $\chi^2(2) = 6.09$, $p = .048$, $V = .110$, and less likely to agree that there is a light behind the dress, $\chi^2(2) = 7.22$, $p = .027$, $V = .120$, than those who saw it as white-and-gold. Answers about the light source on the left, $\chi^2(2) = 0.68$, $p = .711$, $V = .040$, were distributed similarly in the two groups (Figure 7).

The distribution of answers regarding the interpretation of the background in the image was similar to the previous study; 62% of observers in "blue and black" group interpreted the background of the photo as a mirror while 29% described it as a window as opposed to 55% and 34% in the "white and gold" group. However, this difference failed to reach significance, $\chi^2(1) = 1.34$, $p = .248$, $V = .060$.

The interpretation of the character of light sources in the photograph (Figure 8) was not associated with the perceived colors of the dress, $\chi^2(5, N = 208) = 8.84$, $p = .116$, $V = .210$. Notably, white-and-gold perception was dominant for answers "Daylight" and "LED," but there were not enough responses in these categories to draw any meaningful conclusions.
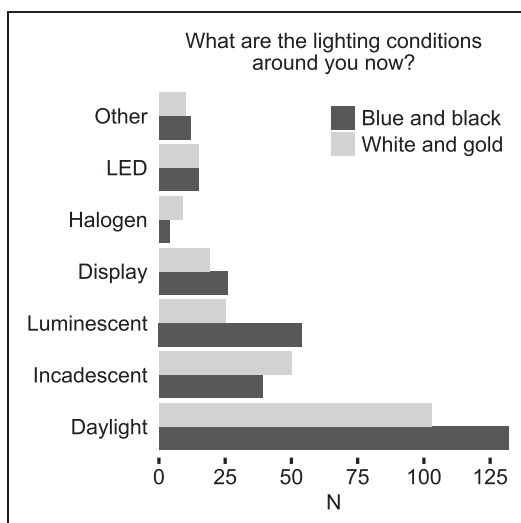
Environmental lighting conditions. Figure 9 shows the distribution of environmental lighting for the two groups of observers. Perception of the dress color was different among different lighting conditions, $\chi^2(6) = 13.84$, $p = .031$, $V = .160$. Luminescent lamps and LCD display that typically provide more "blue" light compared to incandescent, LED, and halogen lamps, were more likely to be associated with blue-and-black perception, $\chi^2(1) = 10.08$,
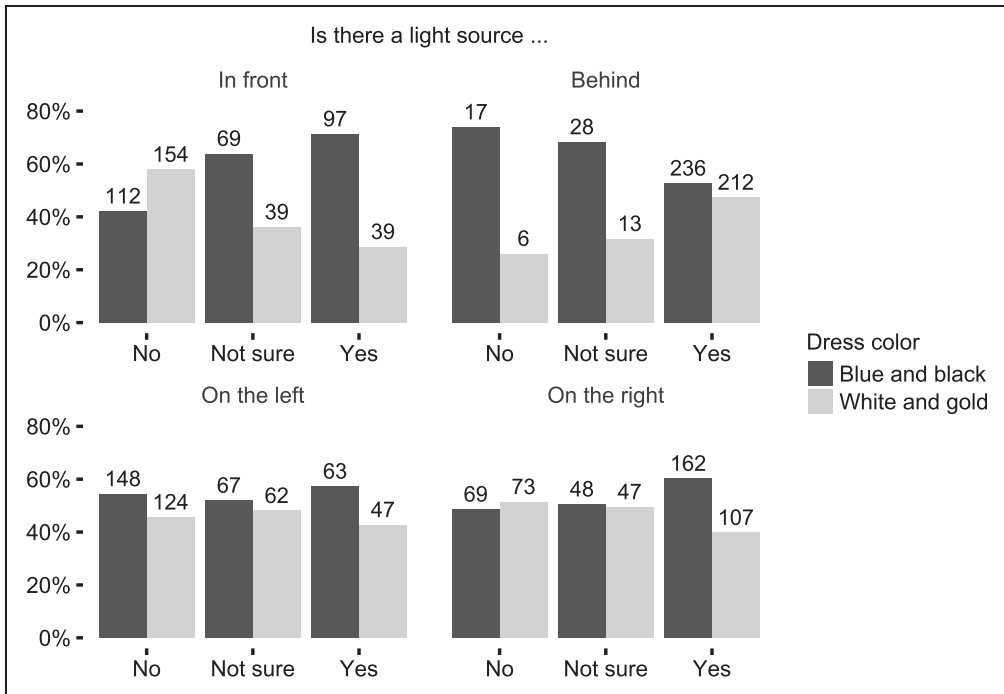
Figure 7. Inferred light sources in "the dress" photo in Study 2. The numbers on top of the bars show responses frequencies.
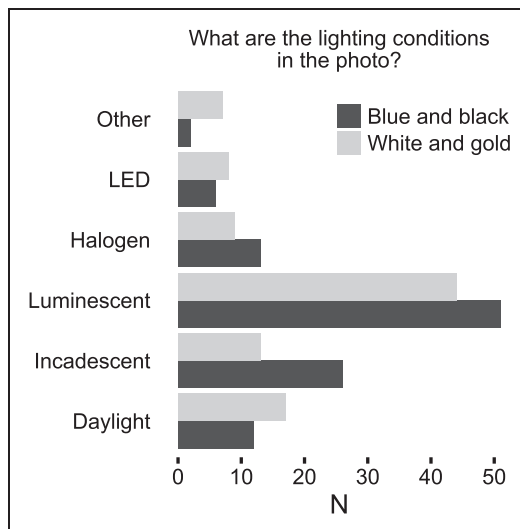
Figure 8. Estimated lighting conditions in the dress photo in Study 2.

Figure 9. Environmental lighting conditions, Study 2.

$p = .001$, $V = .200$. Daylight was also associated with a higher probability of blue-and-black perception, but as the daylight spectrum varies during the day, we analyzed it separately.

Circadian variations. Figure 10 shows the distribution of dress colors during the day for observers who answered "daylight" when asked about environmental lighting conditions. The pattern of results in the two most represented cities in our sample, Moscow ($N = 89$) and St. Petersburg ($N = 98$), was similar. In contrast to what was observed in Study 1, the probability of the blue-and-black perception was higher during the day (noon–6 p.m., 64% and 59% for Moscow and St. Petersburg, respectively) than during the evening (6 p.m.– midnight, 60% and 52%) or during the morning (6 a.m.–noon, 43% and 50%). Given that the results from the two cities were similar and the overall amount of responses with no artificial lighting was limited, we did the statistical tests on the combined data. Binomial regression indicated a marginally significant quadratic trend in the effect of time of the day on the probability of interpretation of the dress as white-and-gold, $B = 3.99$, $SE = 2.12$, $Z = 1.88$, $p = .060$. There were no significant effects of the time of the day on the inferences regarding the frontal light sources. However, the pattern of results for this variable was also opposite of the one observed in Study 1. Fewer participants agreed that there was a light source in front of the dress in the evening than during the day or morning hours (Figure 10).

Discussion

Study 2 demonstrates that the color-matching procedure we used in Study 1 provides reliable results. When observers participate in the study the second time using different device, the colors they match are similar even though the actual colors of the dress photo would depend on the display properties. We believe that two factors provide such reliability. First, color matching and presentation of an image are similarly biased by the display. If, for example, the dress colors would be darker in one display than in another display, the colors shown by color picker would be similarly darker, and the colors matched by observers would be similar.
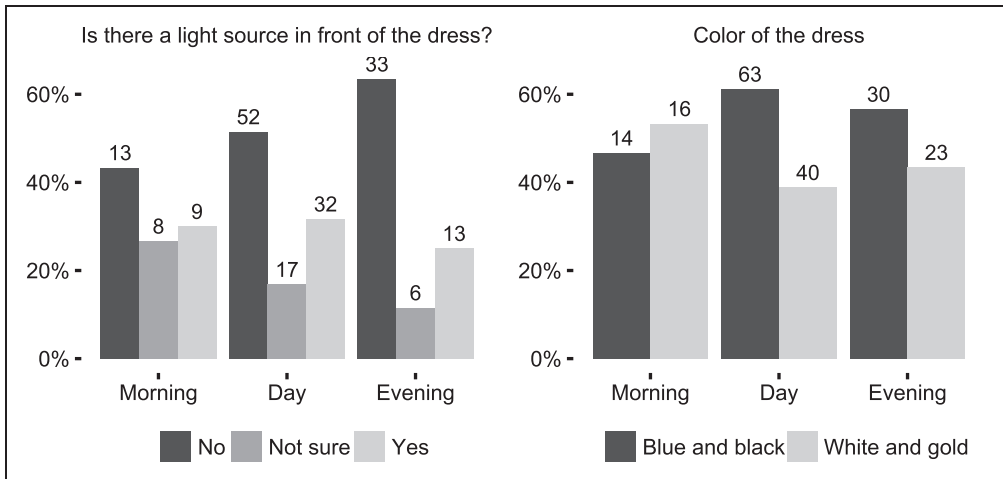
Figure 10. Circadian changes in the perceived colors of the dress for observers in the daylight environment, Study 2. The numbers on top of the bars show responses frequencies. There was only one answer during the "night" period (not shown).

Second, the matched colors are highly dependent on perceptual processing. This is what makes the dress image so interesting in the first place. The L and b channel values of matched colors are unlikely to be described by a unimodal distribution. The bimodality (or multimodality) restricts the variation of colors thus adding to the reliability of the color-matching procedure.

The second major finding of Study 2 is the influence of environmental lighting conditions on perceived colors. When luminescent lamps or LCD displays were the dominant sources of light, participants perceived the dress as blue-and-black rather than as white-and-gold. But with incandescent, LED, or halogen lamp illumination, white-and-gold perception was dominant. This again demonstrates that ambient light may serve as a cue for the interpretation of the dress colors. Note that this result goes against the discounting explanation, because observers were biased by more "warm" illuminants to perceive the dress as "white and gold." Moreover, we did not find any evidence of the circadian changes observed in Study 1. We will discuss this discrepancy in more details in the "General Discussion" section.

In addition, in contrast to Study 1, we did find an effect of age on the perceived color of the dress in Study 2. In agreement with previous findings (Lafer-Sousa et al., 2015), older participants perceive the dress as white-and-gold more often. However, the effect of gender did not reach significance.

General Discussion

Two studies reported in this article demonstrate that inferred positions of light sources were different among observers perceiving the dress as white-and-gold and those perceiving it as blue-and-black. The most striking differences are related to the frontal light source. Supporting under- or overexposure explanation, the participants in blue-and-black group more often perceived the dress as illuminated from the front than those in white-and-gold group. Importantly, the estimated character of lighting conditions (daylight, luminescent lamp, etc.) within the image provided by participants in Study 2

did not predict the perceived color of the dress. That is, the inferences regarding the positions of light sources but not regarding their spectrum were associated with observers' perception of the dress color.

Study 1 demonstrated that perception of the dress changes during the day, with lighter and more yellowish colors around noon. This variation is consistent with discounting explanation: observers discard the illumination from the "cold" daylight and perceive the dress as more "warm." The inferred position of light sources mediates this effect considerably. During the day observers are less likely agree that there is a light source in front of the dress associated with blue-and-black perception. This demonstrates the dependency of a high-level scene interpretation on the context of the perceived image that helps to disambiguate the scene. The lack of correlation between the accuracy of color discrimination or observers' gray point and perceived dress colors further supports the idea.

However, in Study 2 we were not able to replicate the previously observed circadian changes in perceived colors of the dress. Moreover, while in Study 1 white-and-gold perception was dominant during the morning and the day period but not at night and during the evening, in Study 2 during the day blue-and-black perception was more dominant that during the morning or during the evening. Study 2 differs from Study 1 in several aspects. First, while Study 1 was done in March, Study 2 was done in June during a period close to the summer solstice. Thus, the "day" and the "evening" periods in the two studies were quite different. While in March "evening" was after sunset (6 p.m. in Moscow in the beginning or March, 7 p.m. in the end), in June almost all answers in the "evening" were before sunset (around 9 p.m. in Moscow and 10 p.m. in St. Petersburg). A seasonal variations in the environmental color statistics could also play part in the discrepant findings (Webster, Mizokami, & Webster, 2007). Second, the participants may have seen the dress image multiple times since its first appearance. Given that for most of them the perceived colors do not change, the conditions under which they saw the dress the first time may play overwhelming role in the subsequent perception. Note that the results of Study 2 are in agreement with the pilot study (Appendix) that have been conducted in the end of February. Thus, we believe that either the seasonal variations or the previous exposure to the image may lead to the absence of circadian variation in Study 2.

Finally, the first study did not account for environmental lighting conditions. In Study 2, we found that different ambient lighting may result in different perceived colors with "warmer" illumination resulting in "white and gold" perception. This can explain why Study 1 demonstrated the dominance of blue-and-black perception during the evening and at night when the illumination from luminescent lamps or devices screens might have been dominant. However, it might be difficult to explain from this point of view why white-and-gold perception was dominant during the day in Study 1, and in general the findings of Study 2 regarding the environmental lighting are against the discounting explanation. Speculatively, other factors such as the brightness of ambient lighting or its direction might be more important than its spectrum.

The studies described in this article also allow to resolve the discrepancy regarding the distribution of perceived colors between groups of observers in the previously published works. While Lafer-Sousa et al. (2015) described the distribution as bimodal (or weakly trimodal), Gegenfurtner et al. were in favor of unimodal distribution (Gegenfurtner et al., 2015). The data from both studies presented here speak in favor of bimodality. Both lightness and "blueness" of matched colors had at least two peaks in the distribution corresponding to the differences in the interpretation of the dress. We made our best effort to avoid biases in observers' answers by using labels that did not include color names when we referred to the stripes of the dress. However, interpretation of the dress as "blue and

black'' or ''white and gold'' in the media still could play a role in the results leading to categorical perception.

One caveat related to the interpretation of our findings is that there is a possibility that the devices observers use are somehow depended on their individual differences. For example, observers could adjust color settings on each of their devices. This is very improbable, because it is unlikely that many observers would adjust their display settings and for many devices (e.g., most of the smartphones) such settings are unavailable. However, the possibility cannot be discarded.

It is also possible that observers' answers regarding the positions of light sources are based on the perceived colors of the dress. That is, observers can infer that there is a light in front of the dress because they perceive it as blue-and-black or, vice versa, there is no light in front of the dress because it looks white-and-gold. As in case of the suggested inference of the dress color basing on the light sources, the reverse inference is unlikely to be a conscious inference. To make it, observers need to be aware of how the dress ''looks like'' and how it ''really is.'' The majority of the sample reported that their perception of the dress is stable, that is, they have conscious access to only one of the perceptual interpretations of the image (the data on the position of the light sources reanalyzed excluding observers with changing perception yielded the same results with slightly larger effect sizes). The reverse inference explanation, in our opinion, leaves open the question of why observers make such assumptions regarding the position of light sources. That is, if the dress color is already perceived in a specific way, it is unclear what is additionally explained by assuming that the light sources are in specific positions. Nevertheless, the design of our experiment does not allow to reject the possibility of the reverse inference. The interpretation of the scene background, in contrast, can be based on the perceived light sources and be more dependent on a posthoc rationalization. The presence of a strong light behind the dress is rather evident from the image. Speculatively, if observers agree that there is front light source, then the light sources in the background can be interpreted as a reflection.

The dress image poses a serious challenge for vision science, because it is quite different both from other phenomena usually classified as ''color illusions'' and from other bistable images. We advocate a high-level explanation of the perceived colors of the dress: utilizing cues from various sources, including ambient light, observers create the interpretation of a scene that drives color perception. Explicit assumptions of observers about the light sources and the image background agree with that explanation. Due to the design of our study, it is possible that these assumptions are based on the perception of dress colors or merely correlated with it. However, the observed differences are consistent with under- or overexposure explanations and may be difficult to explain otherwise.

Declaration of Conflicting Interests

Funding

Acknowledgements

Note

1. We used Student's (Welch's) *t*-test for unequal variances. The exact number of degrees of freedom varies because of the Welch's approximation. In addition, not all subjects answered all questions (e.g., data on age were missing for some subjects), which may lead to the variation in degrees of freedom between the tests.

References

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

Çelik, T., Lilley, C., & Baron, L. D. (2011). *CSS color module level 3: W3C recommendation 07 June 2011*. Retrieved from http://www.w3.org/TR/2011/REC-css3-color-20110607/

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

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

Granzier, J. J. M., Brenner, E., & Smeets, J. B. J. (2009). Can illumination estimates provide the basis for color constancy? *Journal of Vision*, *9*, 18.1–18.11.

Granzier, J. J. M., & Valsecchi, M. (2014). Variations in daylight as a contextual cue for estimating season, time of day, and weather conditions. *Journal of Vision*, *14*, 1–23.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, *13*, 70–84.

Hernández-Andrés, J., Romero, J., Nieves, J. L., & Lee, R. L. (2001). Color and spectral analysis of daylight in southern Europe. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *18*, 1325–1335.

Holzmann, H., & Vollmer, S. (2008). A likelihood ratio test for bimodality in two-component mixtures with application to regional income distribution in the EU. *AStA Advances in Statistical Analysis*, *92*, 57–69.

Judd, D. B., Macadam, D. L., Wyszecki, G., Budde, H. W., Condit, H. R., Henderson, S. T., . . . Simonds, J. L. (1964). Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America*, *54*, 1031.

Kaplan, R. M., & Saccuzzo, D. P. (2009). *Psychological testing: Principles, applications, and issues* (7th ed). Belmont, CA: Wadsworth.

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

Lotto, R. B., & Purves, D. (1999). The effects of color on brightness. *Nature Neuroscience*, *2*, 1010–1014.

Maechler, M. (2014). *Diptest: Hartigan's dip test statistic for unimodality - Corrected code*. Retrieved from http://cran.r-project.org/package=diptest

Melamud, A., Simpson, E., & Traboulsi, E. I. (2006). Introducing a new computer-based test for the clinical evaluation of color discrimination. *American Journal of Ophthalmology*, *142*, 953–960.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from https://www.R-project.org/

Rogers, A. (2015, February 26). The science of why no one agrees on the color of this dress. *Wired*. Retrieved from http://www.wired.com/2015/02/science-one-agrees-color-dress/

Romero, J., Hernández-Andrés, J., Nieves, J. L., & García, J. A. (2003). Color coordinates of objects with daylight changes. *Color Research and Application*, *28*, 25–35.

Schwaiger, F., Holzmann, H., & Vollmer, S. (2013). *Bimodalitytest: Testing for bimodality in a normal mixture*. Retrieved from https://www.uni-marburg.de/fb12/stoch/forschung/rpackages/bimodalitytestwin.zip

Schwartz, J.-L., Grimault, N., Hupe, J.-M., Moore, B. C. J., & Pressnitzer, D. (2012). Multistability in perception: binding sensory modalities, an overview. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 896–905.

Shevell, S. K., & Kingdom, F. A. A. (2008). Color in complex scenes. *Annual Review of Psychology*, *59*, 143–166.

Shin, Y. J., Park, K. H., Hwang, J. M., Wee, W. R., & Lee, J. H. (2007). A new color vision test to differentiate congenital and acquired color vision defects. *Ophthalmology*, *114*, 1341–1347.

The dress (viral phenomenon). (2015). Retrieved from https://en.wikipedia.org/w/index.php?title= The_dress_%28viral_phenomenon%29&oldid=692077870

Webster, M. A., Mizokami, Y., & Webster, S. M. (2007). Seasonal variations in the color statistics of natural images. *Network*, *18*, 213–233.

Yang, J. N., & Shevell, S. K. (2003). Surface color perception under two illuminants: The second illuminant reduces color constancy. *Journal of Vision*, *3*, 369–379.

Appendix. The Pilot Study

Method

Procedure. The pilot study was run early after the appearance of "the dress" (from February 27, 2015 until March 3, 2015). The study was done using the Google Forms interface and initially consisted of the image of the dress along with the four questions related to the positions of the light sources, and the questions about the interpretation of the dress color and the question "Have you seen this image earlier (before taking this questionnaire)?" The questions regarding the gender and age were added few hours after starting the study. Google Forms also automatically records the time of the participation (UTC) but not the time zone of participants. Giving that the information about the study was first published in Saint Petersburg and Moscow-based online communities, we assume that most of the participants were from those cities and analyze the time data accordingly. The wording of the questions was the same as in Study 1.

Participants. In total, 883 participants completed the quiz during the period from February 27, 2015 to March 3, 2015. The majority of participants (93%) filled the questionnaire during the first two days of the study. For part of them ($N = 246$), age and gender were not recorded, the rest reported their age as 12 to 63 years old, 174 men, 466 women (Table 4).

Results

As in Study 1 and Study 2, observers were sharply divided in their perception of the dress. Dress categorization was related to age, $t(612.9) = 2.60$, $p = .009$, $d = 0.21$ (participants with blue-and-black perception were younger), but not gender, $\chi^2(1, N = 640) = 1.05$, $p = .305$, $V = .04$.

Table 4. Age and Gender Distribution in the Pilot Study.

	Women		Men		Unknown		Age ($N = 637$)	
	N	%	N	%	N	%	M	SD
White-and-gold	240	52	81	47	126	52	23.9	6.4
Blue-and-black	226	48	93	53	117	48	22.7	5.2
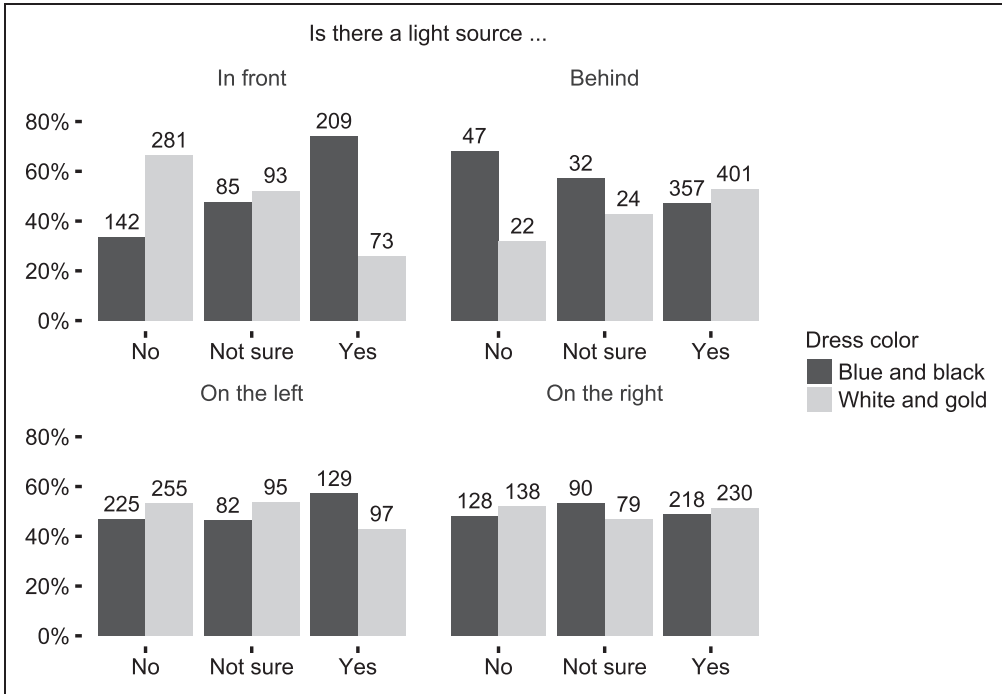
Figure 11. Inferred light sources in "the dress" photo in the pilot study. The numbers on top of the bars show responses frequencies.
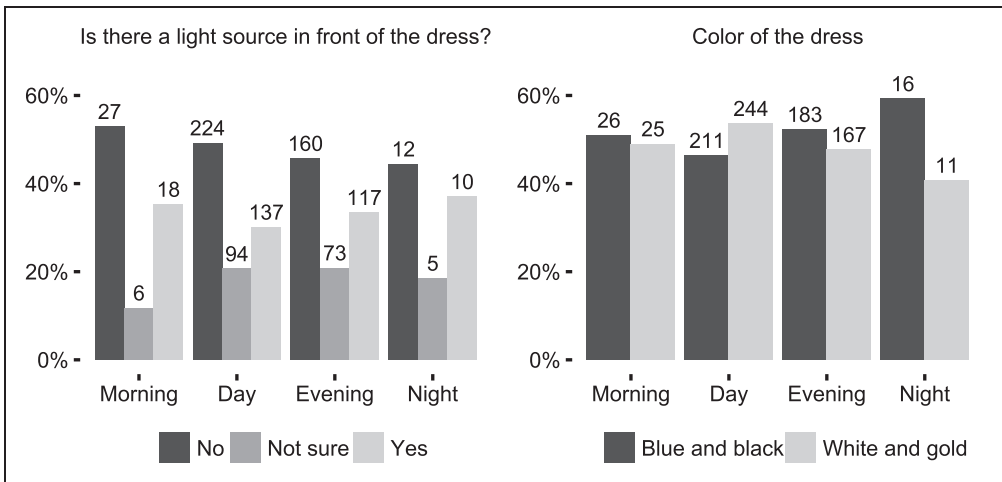
Figure 12. Changes in inferred illumination and perceived dress color as function of time of the day during. *Note.* Morning—6 a.m. to noon, day—noon to 6 p.m., evening—6 p.m. to midnight, night—midnight to 6 a.m. The numbers on top of the bars show responses frequencies.

Participants in white-and-gold group and blue-and-black group agreed about the presence of the light source on the right from the dress, $\chi^2(2, N = 883) = 1.28$, $p = .528$, $V = .04$, but not on the left, $\chi^2(2, N = 883) = 7.22$, $p = .027$, $V = .09$, behind the dress, $\chi^2(2, N = 883) = 12.62$, $p = .002$, $V = .12$, or in front of the dress, $\chi^2(2, N = 883) = 111.50$, $p < .001$, $V = .36$ (Figure 11).

Binomial regression indicated a significant quadratic effect of the time of the day on the perceived color of the dress, $B = -4.16$, $SE = 2.04$, $Z = 2.04$, $p = .042$ (Figure 12). Blue-and-black perception was dominant during the evening (52%) and night (59%) hours while white-and-gold perception was more probable during the day (54%). During the morning hours, the probabilities of white-and-gold perception and blue-and-black perception were almost equal (51% vs. 49%). The data on inferred light sources followed the same pattern as in the Study 1 with higher probability of "Yes" answer regarding the frontal light source during the evening, but the quadratic effect of the time of the day did not reach significance, $B = 1.87$, $SE = 2.09$, $Z = 0.89$, $p = .372$.