Journal of the Optical Society of America A
OPTICS, IMAGE SCIENCE, AND VISION

How to make a #theDress

Christoph Witzel[1,*] **and Matteo Toscani**[2]

[1]Department of Psychology, University of Southampton, University Road, SO17 1BJ Southampton, UK
[2]Experimental Psychology, Justus-Liebig-Universität Gießen, Otto-Behaghel-Str. 10F, 35392 Gießen, Germany
*Corresponding author: cwitzel@daad-alumni.de

Received 4 November 2019; revised 8 February 2020; accepted 15 February 2020; posted 18 February 2020 (Doc. ID 381311); published 19 March 2020

If we completely understand how a phenomenon works, we should be able to produce it ourselves. However, the individual differences in color appearance observed with #theDress seem to be a peculiarity of that photo, and it remains unclear how the proposed mechanisms underlying #theDress can be generalized to other images. Here, we developed a simple algorithm that transforms any image with bicolored objects into an image with the properties of #theDress. We measured the colors perceived in such images and compared them to those perceived in #theDress. Color adjustments confirmed that observers strongly differ in how they perceive the colors of the new images in a similar way as for #theDress. Most importantly, these differences were not unsystematic, but correlated with how observers perceive #theDress. These results imply that the color distribution is sufficient to produce the striking individual differences in color perception originally observed with #theDress—at least as long as the image appears realistic and hence compels the viewer to make assumptions about illuminations and surfaces. The algorithm can be used for stimulus production beyond this study.

https://doi.org/10.1364/JOSAA.381311

1. INTRODUCTION

Individual differences may originate from a small number of common perceptual factors; thus, investigating them may help to understand fundamental determinants of perception [1]. The malleability of color appearance by implicit assumptions has been brought to the spotlight by the striking individual differences in the perception of #theDress. Some observers see the dress in that photo as white and gold, while others perceive it as blue and black. We will call this phenomenon #theDress effect.

Independent laboratories across the world provided evidence that the perceived colors depend on what lighting conditions observers assume in the scene in the photo: Observers who assume the dress is in the shadow tend to see its colors as white–gold; those who assume that it is in bright, direct light see it as blue–black ([2–12]; for review, see [13]). However, many of those observations are merely correlational: When the dress in the photo is seen in a certain color (e.g., white–gold), the illumination is judged correspondingly (e.g., dark and bluish). It seems plausible that the assumption about the illumination causes the perception of the dress; yet, the inverse causal relationship cannot be excluded with certainty, namely, that observers infer the illumination based on the color that they see on the dress. It could also be that a yet unknown other cause determines both the perceived color of the dress and the illumination; see, e.g., [14–17].

Additional evidence suggests that the differences in interpretation are possible because the sensory color signal (chromaticities) of the dress is distributed along the daylight locus, the curve along which natural daylight varies [6–12,18–20]. Rotating chromaticities away from the daylight locus (while keeping luminance information) reduce the individual differences. Individual differences completely disappear when rotated 180 deg [19,20]. However, these observations merely show that the distribution along the daylight locus is necessary for the individual differences. It does not prove that the color distribution is sufficient to elicit those striking differences in color perception, or whether there are other properties of that photo that are necessary to produce #theDress effects. To date, the effects on color appearance observed with #theDress remain a peculiarity of that photo, and it is unclear whether the proposed mechanisms underlying #theDress are general principles that affect the color appearance of other images. Yet, if we completely understand how a phenomenon works, we should be able to produce it ourselves.

This is the purpose of the present study. We developed an algorithm to transform photos into images with properties like #theDress. We then measured color appearance and color naming for those new variants of #theDress and tested whether they produce similar individual differences as #theDress.

2. METHOD

In previous experiments, we observed that repeated viewing of #theDress may contaminate measurements of color appearance [20]. We wanted to make sure that similarities between the new images and #theDress do not emerge due to observers' tendency to provide consistent answers across trials within a session. For this reason, we first measured #theDress in a preliminary online study, in which #theDress was seen before the new images were seen. Then, we did the inverse in the laboratory: we first measured the new images, and then #theDress. Cross validation between the online and laboratory measurements allowed us to assess the role of the sequencing for our results. The average time between the two measurements was 12 days (SD 29 days).

A. Participants

Seventy participants (56 women, 14 men; average age 23.3 ± 5.1 years) took part in the laboratory experiment. Seventy-two observers (57 women, 15 men; 23.3 ± 4.9 years) participated in the preliminary online survey. Of those, 69 participants took part in both the laboratory and the online measurement. All observers were students of the Justus-Liebig-University Gießen and were compensated by eight Euros per hour or course credit. None of the observers had any color deficiencies as tested by self-report (online survey) and the HRR polychromatic plates [21].

B. Apparatus

In the lab, stimuli were presented on a computer monitor with a spatial resolution of 1920×1200 pixels, a refresh rate of 59 Hz, and a color resolution of 10 bits per channel. CIE xyY1931 specifications of the channels were R = [0.6851, 0.3110, 27.1]; G = [0.2133, 0.7267, 69.4]; B = [0.1521, 0.0450, 4.7]. RGB values in the lab and in the online experiment were characterized relative to that monitor, and the monitor white point (xyY1931 = [0.3337, 0.3515, 101.2]) was assumed for CIELUV transformations.

C. Stimuli

From previous investigations, we know that the background has little importance for the individual differences in the perception of #theDress [3,5,12,20,22]. For this reason, we could simplify #theDress by cutting the dress from its original background and showing it on a uniform black background.

Figure 1 illustrates our algorithm for producing #theDress-like images. A precondition of the algorithm is that the image to be processed includes two parts with a lighter and a darker

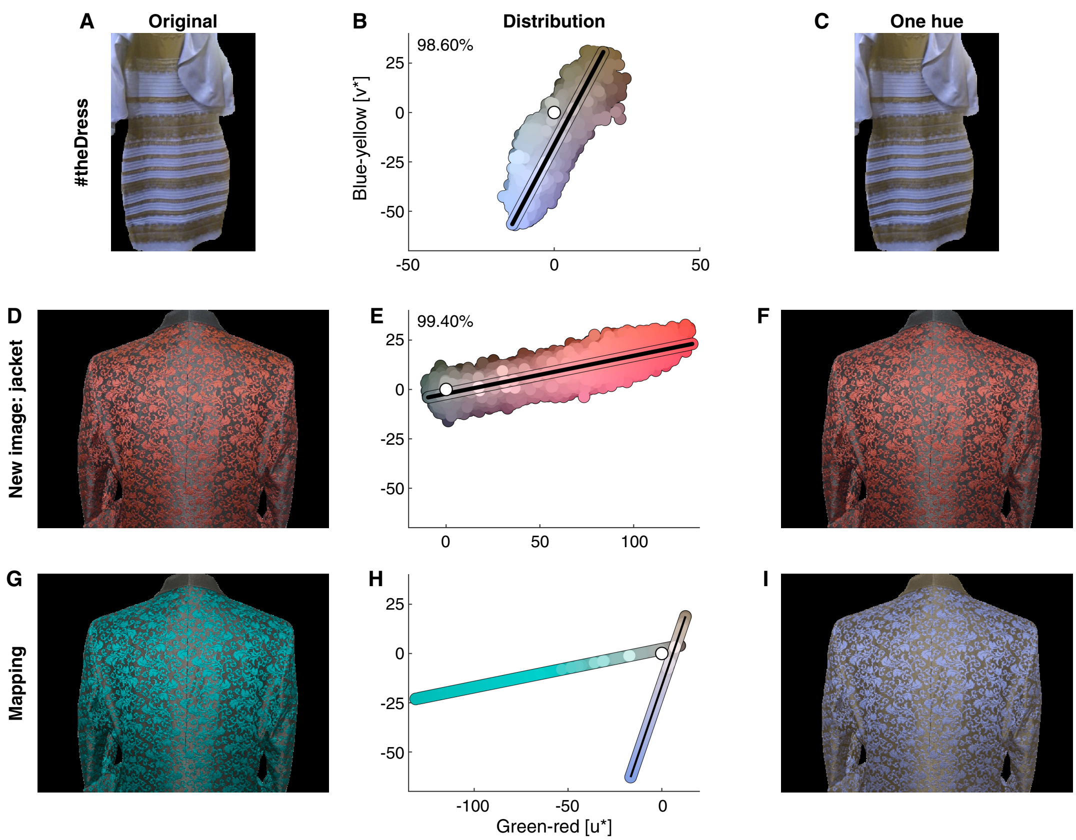

Fig. 1. Illustration of algorithm. The first step is the projection of the chromaticities of #theDress onto the first principal component. Panel (A) reproduces the original photo of #theDress with a black background. Panel (B) represents the distribution of chromaticities of #theDress in CIELUV color space. The black line corresponds to the first principal component of those chromaticities. Panel (C) shows the image that corresponds to the chromaticities projected onto the first principal component [black line in (B)]. In the second step, the same is done with a new image, such as the photo of a jacket [panels (D)–(F)]. To project the bluish chromaticities onto the lighter and the brownish chromaticities onto the darker patterns, the algorithm mirrors chromaticities when necessary. This is the reason the jacket is red in panel (F) and green in panel (G) after mirroring chromaticities [cf. reddish and green line in panels (E) and (H)]. Then the chromaticities [greenish line in panel (H)] are projected onto the principal component of #theDress, and mean and standard deviation are set to those of #theDress [bluish line in panel (H)]. Panel (I) illustrates the resulting image. In [23], a MATLAB algorithm is available to try out the algorithm illustrated here.

color of arbitrary chromaticities. In the example, this was a red–black jacket [Fig. 1(D)]. The algorithm starts with projecting the chromatic distribution of #theDress [Fig. 1(B)] onto the first principal component of that distribution, implying that chromaticities vary only along one hue direction [black line in Fig. 1(B)]. Figure 1(C) shows the resulting image of the dress. The projection onto the principal component barely changes the appearance of #theDress, as can be seen by comparing Fig. 1(C) with the original #theDress in Fig. 1(A).

We apply the same procedure to the new images, as illustrated with the example of the jacket in Figs. 1(D)–1(F). Note again that the transformation of the image is barely visible [Fig. 1(D) versus Fig. 1(F)].

Then, the chromaticities along the principal component of the new image are projected onto the principal component of #theDress. Our algorithm ascertains that the chromaticities of the lighter part of the new image are projected onto the lighter bluish part of #theDress, and the chromaticities from the darker brownish part of the new image are projected onto the darker part of #theDress. For this, our algorithm calculates the correlation between L* and the principal component in #theDress and the new image. If the signs of the correlations differ, the chromaticities of the lighter part of the new image would project onto the darker part of #theDress. In this case, our algorithm inverts the chromaticities of the new image, which is equivalent to mirroring the distribution along the hue direction. Due to this inversion, the jacket in Fig. 1(F) becomes green in Fig. 1(G), and the greenish line of the jacket in Fig. 1(H) is oriented to the opposite direction of the line in Fig. 1(E).

After projecting chromaticities of the new image onto those of #theDress, we also set the mean and the standard deviation of the distribution of the new image to the ones of #theDress [Fig. 1(H)]. Figure 1(I) shows the resulting image of the jacket. To try out the algorithm, a MATLAB program and example images are available in [23].

In addition to the jacket in Fig. 1, we processed three images with this algorithm, showing a tie, an egg, and a fish [Figs. 2(A)–2(C)]. We added a fifth image to this set [Fig. 2(D)]. That image is based on a photo of sandals that appeared on BuzzFeed on November 20, 2016 [24]. That photo has a similar chromatic distribution as #theDress and produces very similar individual differences as #theDress [20,24]. Here, we cut out a "peephole" of that photo to remove shape information about the identity of the sandals in that photo. Then, we processed the image with our algorithm [23] so that chromaticities align with one hue direction. The resulting image allowed us to test whether objects in the photo need to be recognizable to produce individual differences, or whether the chromatic distribution is sufficient. As can be seen from Figs. 1 and 2, each of these images has a light, more bluish, and a dark, more brownish, part with colors close to those of the body and lace of #theDress.

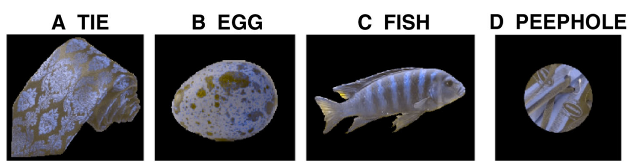

Fig. 2. Stimuli. Images in panels (A)–(D) were created with the algorithm illustrated in Fig. 1 and implemented by the code file [23]. For #theDress and jacket, see Figs. 1(C) and 1(I).

D. Online Survey

The procedure for the online survey followed the one described in detail by Witzel *et al.* [5]. It consisted of four parts. In the first, observers entered personal information (gender, age, glasses, color deficiencies).

In the second part, observers were asked to choose a color term to describe the color of the lace and the body of #theDress by selecting one of 14 color terms, respectively. The color terms were the 11 basic color terms (pink, red, orange, yellow, green, blue, purple, brown, black, gray, and white), plus gold, bronze, and silver. After that, participants answered three questions about the light that illuminates the dress: (1) "Is the dress in the shadow?" with the response options: "Yes," "No," or "I don't know"; (2) "From which direction is the dress illuminated?" From "the front" (the direction of the observer), "the back" (same light as in the background), "Both," or "I don't know"; (3) "Is the dress illuminated by the flash of the camera?" "Yes," "No," or "I don't know. Then, we asked observers to rate the brightness and color of the light that illuminates the dress. They could choose a number between 0 and 10. For the brightness judgments, 0 meant dark and 10 light. The color judgments concerned the blue–yellow hue direction, with a rating of 0 meaning blue, 5 colorless and 10 yellow. Two questions about the illumination of the whole scene followed. (1) "The dress is illuminated by the same light as the background": "Yes," "No," "I don't know"; (2) "The photo is overexposed": "Yes," "No," "I don't know."

In the third part, observers chose color terms to describe each part of the jacket, tie, egg, fish, and peephole. Finally, a fourth part asked observers whether they have seen #theDress before, whether they can switch the way they see the dress, and what color the dress has in reality.

E. Laboratory Experiment

In the laboratory experiment, observers completed the color naming and then color adjustment tasks described previously [12] for the two parts of the object in each image (Fig. 3). For each image, the lighter part was judged first. The image was presented in the center of the screen, and observers named the color of that part using 13 color terms (the same as the survey, but without silver). Then the image was displayed on the left side of the screen, and on the right side a disk was shown in a random color. The observer adjusted the color of the disk so that

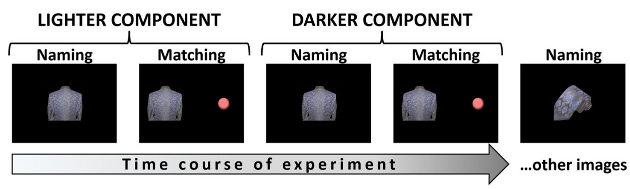

Fig. 3. Procedure of laboratory experiment. Sequence of naming and adjustments for the jacket. In this example, the sequence continues with the tie after the jacket. In the experiment, the sequence of images was randomized except that #theDress was always the last sequence of the experimental session.

it matched the color they perceived in the image. After finishing this sequence for the light part, they did the same for the dark part of the respective object.

The observer did this first for all new images (i.e., without #theDress). The order of the blocks with each image was random. For all observers, the very last block consisted of color naming and color matching for #theDress. While participants saw #theDress before seeing any other image in the online survey, they saw the other images before seeing #theDress in the laboratory experiment.

3. RESULTS

A few observers reported that they mixed up adjustments and naming for the respective two parts of some of the objects. For some observers who did not mention this during the experiment, it seemed obvious from the raw data. For the main results below, we swapped the adjustments and color terms of the two parts if it was obvious (e.g., if the brownish part was adjusted or named blue and the bluish part brown or black). We also deleted responses in which observers gave the same response to the dark and light parts (e.g., body and lace). We did this to exclude the possibility that observers who consistently swapped or repeated answers produced spurious correlations across observers. In any case, the main results also held for the uncorrected raw data—just correlations were slightly lower; we will provide additional information on this in each section below. The raw data are available on Zenodo [25].

A. Adjustments

There was one observer who made almost the same adjustments (Euclidean distance <10) for the two parts of the dress, the jacket, the egg, and the peephole, and there were two such observers for the fish. These data were excluded from the analyses below.

Figure 4 illustrates the color adjustments of the disk to match the lighter (blue circles) and darker (yellow circles) of the images. As for the matches of #theDress, the adjustments of each part of tie, jacket, fish, and peephole were distributed along the blue–yellow direction. Adjustments of the egg were more scattered across color space [Fig. 4(D)].

Following an earlier approach [12], we projected each observer's adjustments onto the first principal component of the six dimensions, i.e., lightness, u^*, and v^* for the two parts of each image (cf. red lines in Fig. 4). Table 1 reports the explained variance and weights of the first principal component for each of the six dimensions. The first principal component explained 72% of the variance of #theDress, 47% of the tie, 58% of the jacket, 38% of the egg, 47% of the fish, and 58% of the peephole [blue bars in Fig. 8(A)]. The scores of the principal component provide a single point for each observer in the six-dimensional space. Positive values of the scores corresponded to data points towards the light, yellowish direction of color space. For all images, the blue–yellow dimension (v^*) of either the lighter or the darker part (tie) of the object yielded the highest weight on the principal component (bold numbers in Table 1).

Correlations across observers between the scores of the first principal component are a way to determine whether individual

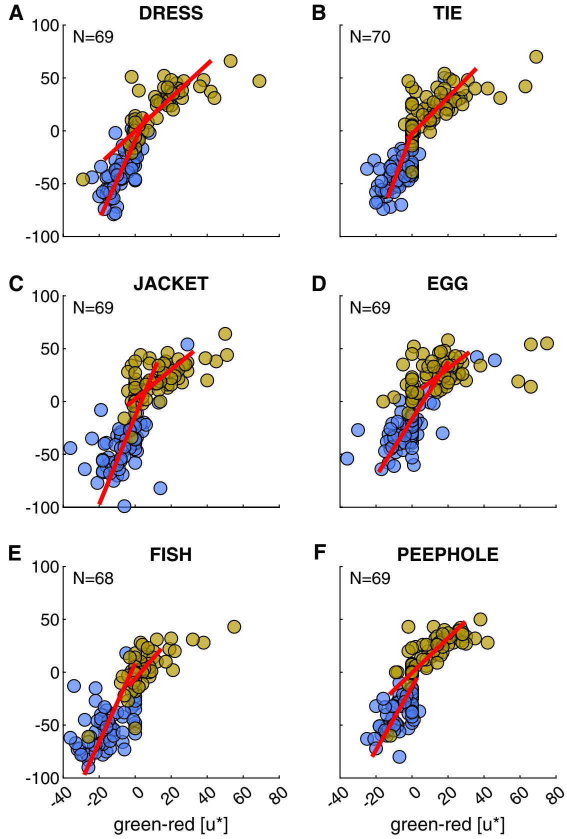

Fig. 4. Adjustments. Green–red (u^*) color dimension on the x axis, blue–yellow (v^*) on the y axis. Each data point represents the match of one participant, blue for the light part of the image, and yellow for the dark one. Red lines represent the first principal component of the color adjustments projected onto the u^* v^* plane. Note that the two red lines are two parts of the same principal component in six-dimensional space. This is possible because four dimensions, i.e., u^* and v^* for the dark and the light part, respectively, are represented in the same u^* v^* plane for illustration purposes (for details, see [12]).

Table 1. First Principal Component of Adjustments

	Expl Var[a]	Light Part[b]			Dark Part[c]		
		L^*	u^*	v^*	L^*	u^*	v^*
Dress	72.0	0.24	0.16	**0.60**	0.27	0.37	0.59
Tie	47.5	0.30	0.13	0.57	0.29	0.35	**0.60**
Jacket	58.2	0.17	0.20	**0.85**	0.20	0.23	0.33
Egg	37.9	−0.06	0.32	**0.86**	0.12	0.22	0.29
Fish	46.5	0.19	0.23	**0.85**	0.28	0.15	0.29
Peep	58.2	0.25	0.20	**0.61**	0.30	0.34	0.56

[a]Explained variance in percent.
[b]Principal component weights of the light part.
[c]Principal component weights of the dark part.

differences of the responses are related across images [5,20,26]. So, we calculated correlations across observers between scores for #theDress and those for each new image. Figure 5 illustrates the correlations through scatter plots. The scores for all images were positively correlated with those for #theDress. The tie ($r(67) = 0.59$, $P < 0.001$), the jacket ($r(66) = 0.54$,

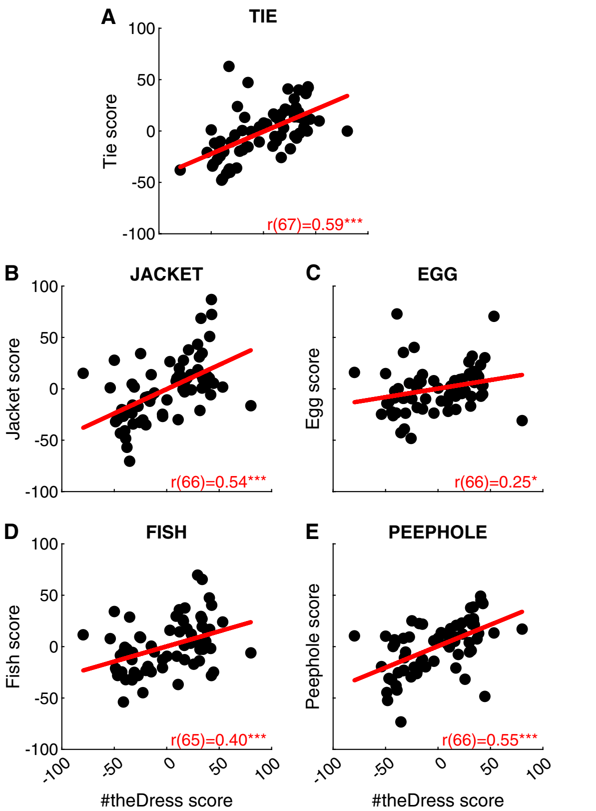

Fig. 5. Correlations of adjustments. Scores of the first principal component of #theDress color adjustments are shown along the x axis, those for the new images along the y axis. Each panel corresponds with one of the new images. Each data point is the principal component score computed on one participant. Pearson's correlation coefficients (r) are reported in the lower right corner in red, with its degree of freedom and statistical significance: $^{*}P < 0.05$, $^{***}P < 0.001$.

$P < 0.001$), and the peephole ($r(66) = 0.55$, $P < 0.001$) yielded the highest, the fish ($r(65) = 0.40$, $P < 0.001$), and the egg ($r(66) = 0.25$, $P = 0.04$) the lowest correlations [see blue bars in Fig. 8(C)]. All correlations were also significant with the uncorrected data except for the egg.

B. Naming (Laboratory)

Figure 6 illustrates naming answers for #theDress and the new images in the laboratory experiment. The light part of all images was mostly called blue or white, and in a fewer cases also purple (upper charts in Fig. 6). The dark part of all images was mostly described as gold, black, brown, or bronze (lower charts in Fig. 6). The fish [Fig. 6(E)] yielded different responses than the other images: There were fewer individual differences for the light part because most observers called it "blue," and most observers described the dark part with the achromatic color terms black or gray instead of gold or brown. For #theDress, the jacket, and the fish, five observers gave the same answer to both parts; this was the case for four observers with the tie and the peephole, and for seven observers with the egg. These answers were excluded from the analyses below.

To assess the systematic variation of color naming across observers, we adopted an earlier approach [5] to calculate naming scores: We coded each of the 14 color terms as a dummy variable ($1 =$ chosen, $0 =$ not chosen), separately for the light and the dark part (e.g., body and lace). Then we performed principal component analyses with those 28 binary variables. The first principal component represents the correlations in naming between the two parts. For example, a combination used by many observers, e.g., blue for the body and black for the lace of #theDress, yields a score with a high absolute value. Whether the minima and maxima of the scores coincide with -1 and 1 depends on the relative frequencies of responses. To calculate the naming scores, we shifted the scores so that the minimum and maximum naming score is -1 and 1. The x axis in Fig. 7 illustrates the naming scores with the example of #theDress.

For all stimuli, the naming score yielded a minimum (-1) for blue and black, and a maximum ($+1$) for white and gold. Scores in between were color-term combinations that were less often combined. For example, the naming scores around -0.5 for #theDress in Fig. 7 corresponded to blue–brown and blue–bronze, scores around 0 to blue–gold, purple–blue, and purple–bronze, and values around 0.5 to purple–gold, and combinations of white with brown, yellow, and bronze.

We calculated correlations between those naming scores and the adjustment scores, as illustrated for #theDress in Fig. 7. The naming scores were positively correlated with the adjustments

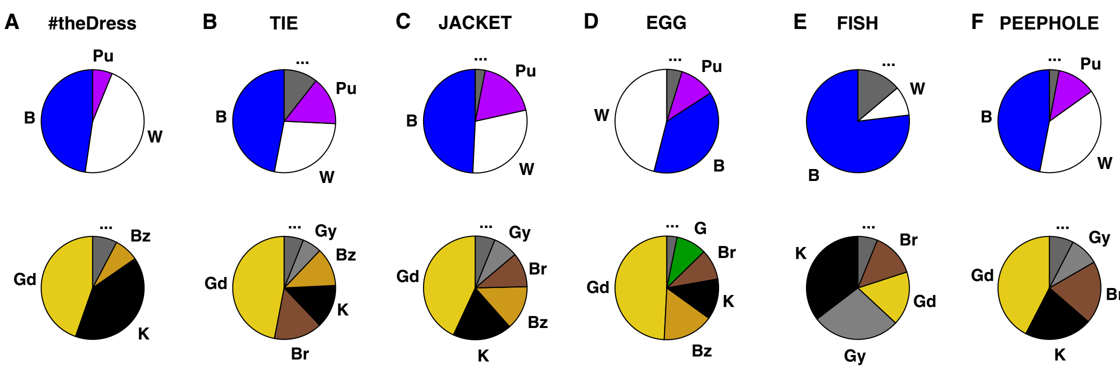

Fig. 6. Color naming in the laboratory experiment. The pie chart expresses the relative naming frequency for each color term observers chose. Each panel corresponds to one stimulus. The upper and lower parts of each panel correspond to the light and dark parts of each object. Initials of color terms are given on the side of each pie: B, blue; W, white; Pu, purple; Gd, gold; K, black; Bz, bronze; Br, brown; Gy, gray; G, green; and . . . , other, i.e., color terms that have been chosen fewer than five times.

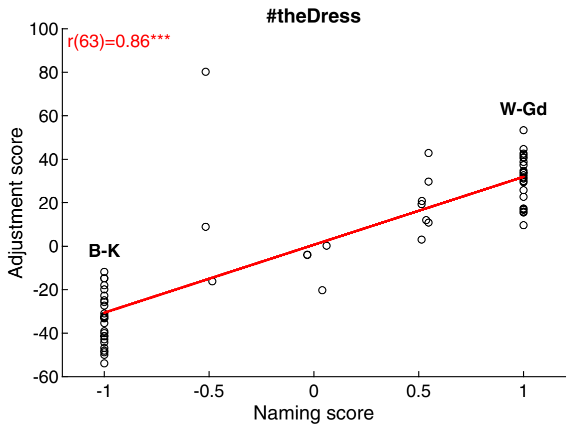

Fig. 7. Correlation between naming and adjustment scores for #theDress in the laboratory experiment. The x axis represents the naming score varying between −1 (B–K = blue–black) and 1 (W–Gd = white–gold). The y axis corresponds to the adjustment scores. Each black circle is the data for one participant. The red line is the regression line illustrating the correlation given in the upper left corner.

scores for all images [all $P < 0.001$; see blue bars in Fig. 8(B)]. These correlations suggest that naming scores convey similar information as the adjustment scores.

We tested for similarities between #theDress and the new images by correlating naming scores of the new images with those of the dress. All correlations were highly significant ($P < 0.001$) and confirmed those observed with the adjustments [green bars in Fig. 8(C)].

C. Online Survey

The answers to #theDress of one of the 72 observers were excluded because he gave the same answer for body and lace (gray). We calculated naming scores for the naming data from the online survey in the same way as above for the laboratory naming data. We correlated the online naming scores with the adjustment and naming scores from the laboratory for each image [see green and yellow bars in Fig. 8(B)]. All correlations were highly significant ($P < 0.001$), indicating similarity between responses in the online survey and those in the laboratory experiment. The only exception was the correlation between online naming and adjustments for the egg, which was much lower than the other correlations [$r(66) = 0.32$,

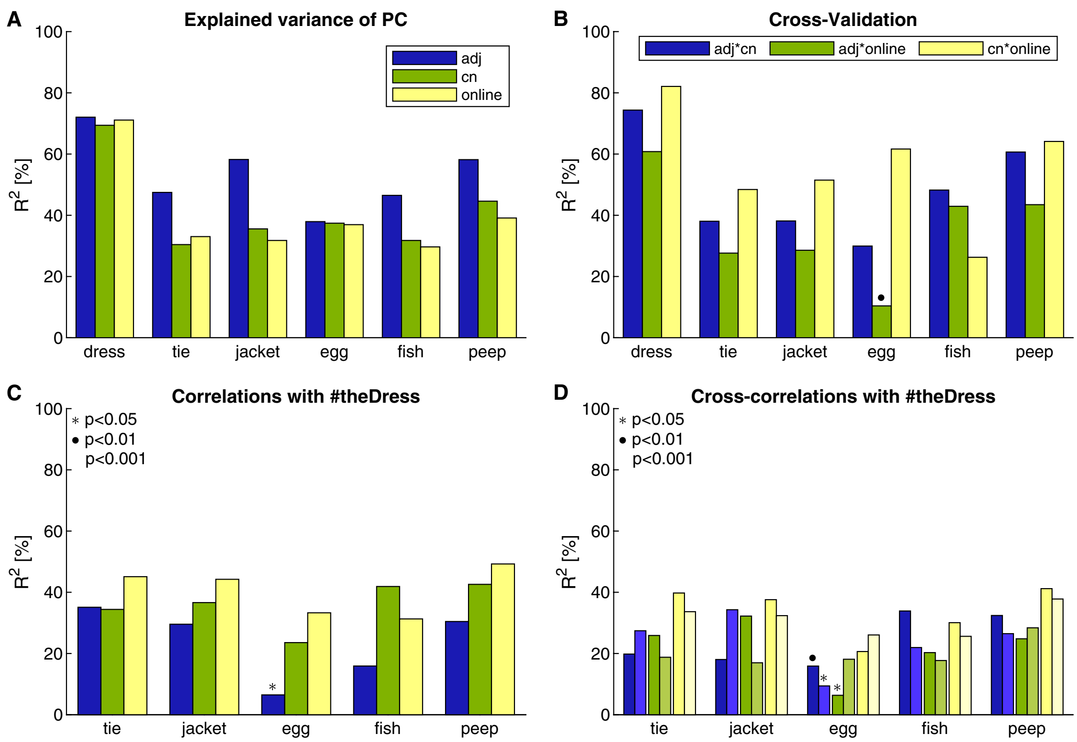

Fig. 8. Overview of results. Panel (A) shows the variance explained (y axis) by the first principal component for each object and each measurement, i.e., adjustments (blue), naming (green), and online survey (yellow). Panel (B) illustrates the correlations between scores for each stimulus in two measurements, i.e., between adjustments and naming (blue), adjustments and online survey (green), and between laboratory and online naming (yellow). Panel (C) summarizes the correlations between the scores for #theDress and those for the five other stimuli in each kind of measurements, i.e., for adjustments (blue), naming (green), and online survey (yellow). Panel (D) illustrates correlations between the scores of #theDress in one measurement and those for the other stimuli in another measurement. The dark blue bars correspond to correlations between the adjustment scores for #theDress and the laboratory naming for the other images; the second blue bars show the inverse, i.e., correlations between laboratory naming of #theDress and adjustments of the other images. The dark and light green bars refer to correlations between online #theDress scores and adjustments of other images, and between adjustments of #theDress and online naming of the other images, respectively. The yellow bars correspond to correlation between online naming of #theDress and laboratory naming of the other images, and vice versa. In panels (B)–(D), symbols above bars refer to P-values for correlations: $^*P < 0.05$, $\bullet\ P < 0.01$. To avoid clutter, no symbol is shown for $P < 0.001$.

$P = 0.007$; see black dot in Fig. 8(B)] and did not reach significance with the uncorrected data.

The online survey reproduced the results from the laboratory, with naming scores for all new images being significantly correlated with those for #theDress (all $r(69) > 0.55$, all $P < 0.001$). In fact, for all but the fish, the correlations were highest for the online naming than for the other measurements [yellow bars in Fig. 8(C)]. These results were the same with the uncorrected data (all $P < 0.001$).

As previously [5], we tested whether the questions on the assumptions about the scene and the illumination were related to #theDress. The assumptions concerned #theDress, not the new images. We calculated correlations between the naming scores and the ratings (0 to 10) of illumination brightness and color (yellow–blue). #theDress was negatively correlated with brightness ($r(69) = -0.30$, $P = 0.01$) and positively with yellow–blue judgments ($r(69) = 0.31$, $P = 0.008$).

We tested the relationship between color perception and qualitative questions about assumptions by calculating point-biserial correlations (the statistics of which correspond to t-tests for independent samples). Observers who assumed the dress was in the shadow tended to answer color terms with a higher score, i.e., towards white–gold [$r(63) = 0.48$, $P < 0.001$; six observers answered they did not know]. Conversely, observers who answered the photo was overexposed tended to answer blue–black [$r(57) = -0.34$, $P = 0.009$]. The results for the questions about shadow and overexposure could be reproduced when using the dress score from naming in laboratory [shadow: $r(58) = 0.39$, $P = 0.002$; overexposed: $r(51) = -0.31$, $P = 0.02$], and the adjustment scores [$r(60) = 0.27$, $P = 0.04$; overexposed: $r(54) = -0.30$, $P = 0.03$]. The other questions, such as those about the flash and the direction of illumination (see Method section) did not yield any significant correlation. The results support the idea that the assumptions about the dress being in the shadow and the photo being overexposed are strongly related to the perceived colors of #theDress.

D. Cross Validation

A final open question is whether the observed correlations between #theDress and the new images can partially be explained by a response bias: Observers might tend to give similar responses across trials, even with different images, when conducting the measurements in one experimental session. To address this issue, we compared the responses between the laboratory and the online measurements.

Figure 8(D) reports correlations between responses to #theDress and the other images across the three types of measurements. The online measurements of #theDress were done before seeing the other images. In contrast, the adjustments of the new images were done before adjusting #theDress. For this reason, the correlations between these two measurements [dark green bars in Fig. 8(D)] are of particular interest. The lowest of these correlations occurred for the egg, and was still significant [$r(66) = 0.25$, $P = 0.04$]. However, all combinations across measurements yielded significant correlations [Fig. 8(D)]. Again, results were all reproduced with the uncorrected data except for two nonsignificant correlations involving the egg. Overall, the positive correlations show that responses to the new images are related to the perception of #theDress, and that this main result is independent of sequencing and robust to measurement noise across the different types of measurements.

4. DISCUSSION

Figure 8 summarizes the results of all three measurements, the adjustments and naming in the laboratory, and the naming in the online survey. For all images, a high amount of variance in the individual differences of the adjustments could be explained by one principal component [Fig. 8(A)]. All three kinds of measurements provided evidence for a correlation between the perception of #theDress and the new images [Figs. 8(C) and 8(D)]. Yet, principal components and correlations do not fully explain the total variance, and the egg and maybe also the fish seemed not to correlate as well with #theDress as the other images [Fig. 8(D)].

A. Adjustment and Naming Scores

Figure 8(A) shows how much variance was explained by the first principal component, indicating how representative the adjustment and naming scores were for the respective measurements. The measurements of #theDress (left group of bars) tended to follow most closely the linear trend captured by the principal component [see Fig. 4(A)]. For the other images, the adjustments (blue bars) seem to be better represented by the scores than the naming data (green and yellow bars). This is likely due to the data format and the approach to producing the naming scores.

Figure 8(B) illustrates the similarities of the three kinds of measurements. Responses to #theDress (left group of bars) are most consistent across the three measurements. This might be related to the higher variance explained by the principal component for responses to #theDress [see Fig. 8(A)]. Correlations involving adjustment scores (blue and green bars) were lower than those between laboratory and online naming (yellow bars) for all images except for the fish.

Adjustments and color naming capture different aspects of the individual differences and have different advantages and disadvantages. Adjustments, relying on direct perceptual matches, are better suited to capturing the individual differences in perceived color than color naming, which confounds individual differences in perception with differences in naming. Color naming varies across observers (e.g., [27–29]). General variation in color naming adds to the similarities across images. For example, an observer who has a larger blue category would be more likely to call all images blue than an observer with a smaller blue category. As a result, color naming might confound two kinds of individual differences, one concerning the individual difference due to #theDress effect and one due to general differences in naming. In this way, correlations of naming scores might potentially overestimate individual differences in perception.

In addition, adjustments provide more precise color specifications in three-dimensional color-space than color terms, which do not distinguish between the many perceivable colors within each category. At the same time, color adjustments in three-dimensional color space constitute a technical challenge to naïve observers. Most naïve observers are unaware of color-opponency and have difficulties navigating through the two-dimensional chromatic plane. The fact that lightness had to be adjusted independently (to avoid effects of the asymmetric gamut) added to the complexity of the adjustment task. In contrast, color naming is rather straightforward and does not require any understanding of color space. This might produce additional noise in the adjustment as compared to the naming task.

For all measurements, some observers reported pressing the wrong key by accident. We corrected data that was obviously swapped between the dark and light parts and deleted data when the same answer was given for the dark and the light part. The raw data provided very similar results, hence confirming that the main results do not depend on the noise produced by erroneous responses. Nevertheless, there were certainly some erroneous responses left in the data that we could not correct or delete. These erroneous responses might well have added noise and reduced the variance explained by principal components and correlations.

In any case, all results provide evidence for systematic individual differences in color perception in the new images similar to those observed in #theDress. The cross validation between laboratory measurements and online survey further shows that those correlations cannot be attributed to effects of presentation sequences. We cannot completely exclude the possibility that observers memorize their answers and try to be consistent across the two sessions (online and laboratory); however, we think this is very unlikely, given the time between the sessions. The differences across objects also contradict a general tendency to give consistent answers because such a tendency should not vary across objects.

B. Main Determinants of Individual Differences

Our findings allow us to pinpoint three sufficient conditions to produce #theDress effects. First, our algorithm focuses on the major blue–yellowish hue direction of #theDress and excludes the role of the chromatic distribution away from that color direction. Following preliminary measurements [20], we used a modified version of #theDress that differed from the original in two respects: it was a cutout of the dress pasted on a black background, and its chromatic distribution was projected to the major, yellow–blue hue direction. However, the appearance of #theDress barely changes when its chromaticities are projected to one hue direction [see Fig. 1(C)]. In addition, our version produces very similar individual differences as the original. In previous studies, we measured individual differences for #theDress with background and complete chromatic distribution. The first component of the adjustments in the present study (72%) explained a higher amount of variance than in a previous study with the original #theDress [62%, cf. Fig. 6(A) in Ref. [12]]. This suggests that the systematic variability across observers was higher when projecting the distribution to the major hue direction. In addition, naming results were also very similar in this [Fig. 5(A)] and in previous studies [Fig. 5 in Ref. [12]; Figs. 4(A)–4(C) in Ref. [5]].

Hence, neither the projection to one hue direction nor the background seems to play a major role for the individual differences in perception. Previous studies have shown that a specific background can bias the perception of #theDress. This is the case when the background is unambiguous and specifies the light that illuminates the dress [7,8,12], or when it changes color perception due to local contrast [30]. However, the observations made here are in line with previous findings that the background of #theDress is largely irrelevant for the individual differences as long as the background is ambiguous [3,22,30]. They also explain why only assumptions about the illumination of the dress, but not assumptions about the background are related to the perception of #theDress (Fig. 11 in Ref. [5]).

Second, our algorithm requires input images to have a dark and a light chromatic part. Lightness might play an important role for #theDress effects. Perceived lightness varies across observers for #theDress, and lightness has a positive weight on the first principal component of all images except the lighter part of the egg (see Table 1). In addition, lightness induction modulates the colors seen in #theDress [30,31]. The role of lightness in #theDress effects could be explained by an earlier finding, according to which observers infer different lightness based on different assumptions about the illumination [32]. However, a gray-scale version of #theDress does not produce individual differences [9,19], suggesting that lightness variation alone cannot explain #theDress effects.

It seems to be the combination of a dark part with a brownish hue and a light part with a bluish hue that is essential for the ambiguous interpretation of the images. This is shown by the observation that individual differences in perception break down when rotating the color distribution of the dress by 180 deg [9,19,20]. The 180 deg rotation does not allow for attributing the bluish tint on the dress to shadow because the relationship between lightness and chromaticities does not allow for that [9].

Evidence for the important role of implicit assumptions about lighting and shadow has also been replicated in this study. The correlations between #theDress and lightness and blue–yellow ratings, shadow answers, and overexposure answers replicate those observed previously [5–12]. The fact that the question about the flash did not reproduce the significant effect observed by Witzel *et al.* [5] may be attributed to the much lower sample size ($N = 500$ in Ref. [5]). Answers to the question about the direction of the illumination did not yield any correlation. This finding is at odds with the one of Chetverikov and Ivanchei [2]. They found that people who perceive the dress as blue-and-black are likely to consider the light source as frontal. Given the replicability of the question about shadow across different studies (see also [4]), we think that assumed shadow might be the most important assumption behind those individual differences in color perception.

Third, the peephole image showed a cutout of a photo of sandals that had a similar chromatic distribution as #theDress. Previously, we showed that individual differences in the perception of those sandals are strongly correlated with perceptions of #theDress [20]. Here, we presented only the cutout. This makes the content of the image barely recognizable. The observed correlations between peephole and #theDress suggest that the individual differences in perception do not depend on recognizability, i.e., whether the observer recognizes the objects in the scene.

At the same time, we know from previous studies [3,22] that the color distribution alone, without any meaningful content, is not sufficient to produce individual differences. Hesslinger and Carbon [3] dissected the dress in squares of different sizes and scrambled the squares. The smaller the square, the less the image looked like a real material, and the smaller the individual differences. From those previous observations, together with our observation for the peephole, we take that the objects do not need to be recognizable to produce individual differences; but they must look like real material.

In sum, the hue direction of the chromatic distribution, the dark and light components, and the realism of the images were the only similarities between the new images and #theDress. The individual differences in the perception of the new images and the correlations with #theDress show that these features are sufficient to produce #theDress effect. So, the stripes and body shape of the dress in #theDress photo are not critical for the individual differences in color perception. Instead, the distribution of the chromaticities from bright bluish colors to dark brownish colors is the critical feature of this phenomenon.

Our algorithm can be of great use for other studies. By now, #theDress is world-famous. In this study, most observers (67 of 72, i.e., 93%) responded that they knew the image before completing our online survey (cf. 78% in Ref. [7], 73% in Ref. [5], 89% in Ref. [22]). It is possible that this prior knowledge influences what observers see or lead observers to answer according to their knowledge rather than their perception [8]. Such effects of prior knowledge might reduce or interact with #theDress effects, and studies most likely want to avoid confusion between perception and knowledge. In this case, studies can use our algorithm for producing new sets of stimuli to elicit #theDress effects while making sure that observers did not see those images ever before.

C. Other Determinants

However, the correlations between #theDress and the new images were not complete [see Fig. 8(D)]. In particular, the egg produced lower correlations [see Figs. 8(C) and 8(D)], and the fish yielded different naming patterns [Fig. 6(E)] than tie, jacket, and peephole. These observations imply that the effectiveness of our algorithm depends on properties of the images beyond the color distribution, dark–light components, and realism (see Section 4.B). Still other factors might modulate the magnitude of #theDress effects.

It is generally known that many properties of photos can affect color and material appearance [32–37]. For example, #theDress, the tie, and the jacket are all fabrics. The fish and the egg might have different material properties than fabrics. One such material property is gloss. Perceived gloss has been shown to be related to the perceived color of #theDress [5]. Our algorithm does not map the gloss of #theDress to the new images, and we did not control for gloss in the stimulus sampling. Therefore, the gloss of our new images is likely to vary. On visual inspection [Fig. 2(C)], the fish seems to lack gloss, and that might be the reason for smaller #theDress effects with the fish. Another candidate material property is translucency. According to Fig. 2(C), the fish seems to feature some translucency at the fins. Much of the yellowishness from #theDress seems to be mapped to the translucent part of those fins. Still another candidate property is the spatial distribution of chromaticities. In Fig. 2(B), the blue spots in the lower, shaded area of the egg look particularly saturated. This concentration of saturated blue does not seem as visible in #theDress and the other images [Figs. 1(A) and 2]. The high saturation could contradict an effect of shadow on a white egg. These particularities of the fish and the egg might undermine the ambiguity of illumination assumptions and #theDress effect. In addition, quail eggs are typically white with brown spots. Knowledge about typical colors (*memory colors*) affects color appearance (for review, see [13]), which might also counteract individual differences.

Here, the role of gloss, translucency, saturation, memory colors, or still other properties of the images remain hypothetical because our study does not allow for identifying determinants beyond the color distribution. Starting from our observations, future studies may quantify and systematically control the effects of other image properties (e.g., [32–37]) and evaluate their role in #theDress effects as we have done it for the color distribution in this study.

Finally, while our results show that the color distribution is sufficient, newest findings suggest that it might not be a necessary condition for individual differences in color perception. Those findings show that individual differences can be found for images with color distributions that do not align with the blue–yellow hue direction of #theDress [38,39]. Interestingly, those observations contrast with previous ones, according to which #theDress effects reduce for colors away from the blue–yellow hue direction [19,20]. For the moment, how far the individual differences revealed by the new images are related to those of #theDress remains an open question. Our approach of correlating individual differences across images [5,20,26] could help clarify the relationship between the phenomena.

5. CONCLUSION

#theDress is not unique. We developed a simple algorithm that produces new images with color distributions similar to #theDress. Our findings showed that those images elicited individual differences in color perception similar to those observed with #theDress. These observations suggest that the object in the photo (i.e., the dress) is of little importance; instead, the color distribution is a sufficient condition to produce individual differences in color perception with other pictures. It seems unnecessary that the depicted objects be recognizable, but they need to look like real materials so that the viewer makes assumptions about the illumination to make sense of image. Other factors, such as gloss, translucency, spatial information, and memory colors might also affect individual differences. In future developments, these additional factors could be integrated into our algorithm to modulate the strength of #theDress effect. In any case, the current version of the algorithm is already usable in studies that want to exclude effects of prior knowledge about the famous #theDress when investigating #theDress effects.

Funding. Deutsche Forschungsgemeinschaft (DFG) (SFB TRR 135 C2).

Acknowledgment. We thank Julia Lahoda and Joanna Szczotka for their help with data collection.

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. J. M. Bosten, J. D. Mollon, D. H. Peterzell, and M. A. Webster, "Individual differences as a window into the structure and function of the visual system," Vis. Res. **141**, 1–3 (2017).
2. A. Chetverikov and I. Ivanchei, "Seeing 'the Dress' in the right light: perceived colors and inferred light sources," Perception **45**, 910–930 (2016).
3. V. M. Hesslinger and C.-C. Carbon, "#theDress: the role of illumination information and individual differences in the psychophysics of perceiving white–blue ambiguities," i-Perception **7**, 2041669516645592 (2016).
4. P. Wallisch, "Illumination assumptions account for individual differences in the perceptual interpretation of a profoundly ambiguous stimulus in the color domain: "The dress"," J. Vis. **17**(4), 5 (2017).
5. C. Witzel, J. K. O'Regan, and S. Hansmann-Roth, "The dress and individual differences in the perception of surface properties," Vis. Res. **141**, 76–94 (2017).
6. S. Aston and A. Hurlbert, "What #theDress reveals about the role of illumination priors in color perception and color constancy," J. Vis. **17**(9), 4 (2017).
7. R. Lafer-Sousa, K. L. Hermann, and B. R. Conway, "Striking individual differences in color perception uncovered by 'the dress' photograph," Curr. Biol. **25**, R545–R546 (2015).
8. R. Lafer-Sousa and B. R. Conway, "#theDress: categorical perception of an ambiguous color image," J. Vis. **17**(12), 25 (2017).
9. A. D. Winkler, L. Spillmann, J. S. Werner, and M. A. Webster, "Asymmetries in blue-yellow color perception and in the color of 'the dress'," Curr. Biol. **25**, R547–R548 (2015).
10. M. Toscani, K. R. Gegenfurtner, and K. Doerschner, "Differences in illumination estimation in #thedress," J. Vis. **17**(1), 22 (2017).
11. K. Uchikawa, T. Morimoto, and T. Matsumoto, "Understanding individual differences in color appearance of "#theDress" based on the optimal color hypothesis," J. Vis. **17**(8), 10 (2017).
12. C. Witzel, C. Racey, and J. K. O'Regan, "The most reasonable explanation of "the dress": implicit assumptions about illumination," J. Vis. **17**(2), 1 (2017).
13. C. Witzel and K. R. Gegenfurtner, "Color perception: objects, constancy, and categories," Ann. Rev. Vis. Sci. **4**, 475–499 (2018).
14. E. L. Dixon and A. G. Shapiro, "Spatial filtering, color constancy, and the color-changing dress," J. Vis. **17**(3), 7 (2017).
15. O. A. Mahroo, K. M. Williams, I. T. Hossain, E. Yonova-Doing, D. Kozareva, A. Yusuf, M. Oomerjee, T. Soorma, and C. J. Hammond, "Do twins share the same dress code? quantifying relative genetic and environmental contributions to subjective perceptions of 'the dress' in a classical twin study," J. Vis. **17**(1), 29 (2017).
16. J. Rabin, B. Houser, C. Talbert, and R. Patel, "Blue-black or white-gold? early stage processing and the color of 'The Dress'," PLoS One **11**, e0161090 (2016).
17. K. Vemuri, K. Bisla, S. Mulpuru, and S. Varadharajan, "Do normal pupil diameter differences in the population underlie the color selection of #thedress?" J. Opt. Soc. Am. A **33**, A137–A142 (2016).
18. C. Feitosa-Santana, M. Lutze, P. A. Barrionuevo, and D. Cao, "Assessment of #theDress with traditional color vision tests: perception differences are associated with blueness," i-Perception **9**, 2041669518764192 (2018).
19. K. R. Gegenfurtner, M. Bloj, and M. Toscani, "The many colours of 'the dress'," Curr. Biol. **25**, R543–R544 (2015).
20. C. Witzel, S. Poggemann, A. Jakob, K. R. Gegenfurtner, and M. Toscani, "Which image characteristics yield striking individual differences in perceived colour?" in *40th European Conference on Visual Perception (ECVP)*, Berlin, Germany, 2017.
21. L. H. Hardy, G. Rand, and M. C. Rittler, "HRR polychromatic plates," J. Opt. Soc. Am. **44**, 509–523 (1954).
22. D. Jonauskaite, N. Dael, C. A. Parraga, L. Chèvre, A. García Sánchez, and C. Mohr, "Stripping #theDress: the importance of contextual information on inter-individual differences in colour perception," Psychol. Res. doi:10.1007/s00426-018-1097-1 (2018).
23. C. Witzel and M. Toscani, "dressmaker," GitHub, 2019, https://github.com/christophWit/dressmaker.
24. B. Esposito and D. Nye, "What colors are these flip-flops?" BuzzFeed, 2016, https://www.buzzfeed.com/bradesposito/what-colors-are-these-flip-flops?bftw&utm_term=.dw79xjgGJ#.aj7eN8O5K.
25. C. Witzel and M. Toscani, Supplementary data for "How to make a #theDress," https://doi.org/10.5281/zenodo.3659308 (2020).
26. C. Witzel, "Striking individual differences in gloss perception," *60th Conference of Experimental Psychologists (TeaP)*, Marburg, Germany, 2018.
27. C. Witzel and K. R. Gegenfurtner, "Categorical sensitivity to color differences," J. Vis. **13**(7), 1 (2013).
28. M. Olkkonen, C. Witzel, T. Hansen, and K. R. Gegenfurtner, "Categorical color constancy for real surfaces," J. Vis. **10**(9), 1 (2010).
29. D. T. Lindsey and A. M. Brown, "The color lexicon of American English," J. Vis. **14**(2), 17 (2014).
30. L. D. Daoudi, A. Doerig, K. Parkosadze, M. Kunchulia, and M. H. Herzog, "The role of one-shot learning in #theDress," J. Vis. **17**(3), 15 (2017).
31. L. Hugrass, J. Slavikova, M. Horvat, A. A. Musawi, and D. Crewther, "Temporal brightness illusion changes color perception of "the dress," J. Vis. **17**(5), 6 (2017).
32. R. Robilotto and Q. Zaidi, "Limits of lightness identification for real objects under natural viewing conditions," J. Vis. **4**(9), 9 (2004).
33. I. Motoyoshi, S. Y. Nishida, L. Sharan, and E. H. Adelson, "Image statistics and the perception of surface qualities," Nature **447**, 206–209 (2007).
34. C. B. Wiebel, M. Toscani, and K. R. Gegenfurtner, "Statistical correlates of perceived gloss in natural images," Vis. Res. **115**, 175–187 (2015).
35. L. Sharan, Y. Li, I. Motoyoshi, S. Y. Nishida, and E. H. Adelson, "Image statistics for surface reflectance perception," J. Opt. Soc. Am. A **25**, 846–865 (2008).
36. I. Motoyoshi, "Highlight–shading relationship as a cue for the perception of translucent and transparent materials," J. Vis. **10**(9), 6 (2010).
37. R. W. Fleming and H. H. Bülthoff, "Low-level image cues in the perception of translucent materials," ACM Trans. Appl. Percept. **2**, 346–382 (2005).
38. P. Wallisch and M. Karlovich, "Disagreeing about Crocs and socks: creating profoundly ambiguous color displays," arXiv:1908.05736 (2019).
39. A. Werner, S. Fuchs, Y. Kersten, and M. Salinas, "#theShoe is the new #theDress-a colour ambiguity involving the red-green axis needs a new explanation," J. Vis. **18**(10), 891 (2018).