The most reasonable explanation of “the dress”: Implicit assumptions about illumination

Christoph Witzel
Laboratoire Psychologie de la Perception, CNRS UMR 8242, Université Paris Descartes, Paris, France Justus-Liebig-Universität Gießen, Gießen, Germany

Chris Racey
School of Psychology, University of Sussex, Brighton, UK

J. Kevin O’Regan
Laboratoire Psychologie de la Perception, CNRS UMR 8242, Université Paris Descartes, Paris, France

Millions of Internet users around the world challenged science by asking why a certain photo of a dress led different observers to have surprisingly different judgments about the color of the dress. The reason this particular photo produces so diverse a variety of judgments presumably is that the photo allows a variety of interpretations about the illumination of the dress. The most obvious explanation from color science should be that observers have different implicit assumptions about the illumination in the photo. We show that the perceived color of the dress is negatively correlated with the assumed illumination along the daylight locus. Moreover, by manipulating the observers’ assumptions prior to seeing the photo, we can steer how observers will see the colors of the dress. These findings confirm the idea that the perceived colors of the dress depend on the assumptions about the illumination. The phenomenon illustrates the power of unconscious inferences and implicit assumptions in perception.

Introduction

Recent buzz in social media has brought up an example of a photo of a dress (Figure 1) in which different observers have surprisingly different judgments about the color of the dress. Many see the dress as striped gold and white, and many see it as black and blue (Swiked, 2015). Perception feels as though we are directly accessing the characteristics of the world surrounding us. Because we feel that colors are a fundamental property of the visual environment, it is shocking when other observers see completely different colors in this particular photo.

Given the attention of the broader public to this photo, it is no wonder that its effects on color appearance engendered massive debate concerning possible explanations of “the dress” among color scientists (Gegenfurtner, Bloj, & Toscani, 2015; Lafer-Sousa, Hermann, & Conway, 2015; Macknik, Martinez-Conde, & Conway, 2015; Winkler, Spillmann, Werner, & Webster, 2015). It has been proposed that the ambiguity of the perceived colors of the dress is directly related to the variation of the actual colors of the dress along the daylight locus (Gegenfurtner et al., 2015). The daylight locus is the curve that represents the variation of daylight from blue to reddish yellow in color space. Others have speculated that individual differences were due to different prior expectations toward illuminations along the daylight locus (Lafer-Sousa et al., 2015). Still others speculated that individual differences were due to different perceptions of bluishness for colors that vary between gray and blue (Winkler et al., 2015). Nevertheless, it is still not clear why observers see the colors of the dress in fundamentally and apparently irreconcilable ways (Brainard & Hurlbert, 2015).

We consider that a simple explanation provides a complete account of the striking individual differences in the perception of the dress (Witzel, 2015). The idea is that observers, in order to make a judgment about the colors in the dress, must make an assumption about the illumination in the photo. In the photo, the illumination is particularly ambiguous, leading to a large variety of color judgments for the dress (Figure 1). Because observers do not realize the extent to which their judgments depend on their implicit assumptions about illumination, they are surprised to find in this particular case that other

Citation: Witzel, C., Racey, C., & O’Regan, J. K. (2017). The most reasonable explanation of “the dress”: Implicit assumptions about illumination. *Journal of Vision*, *17*(2):1, 1–19, doi:10.1167/17.2.1.

doi: 10.1167/17.2.1 Received October 29, 2015; published February 1, 2017 ISSN 1534-7362 © 2017 The Authors

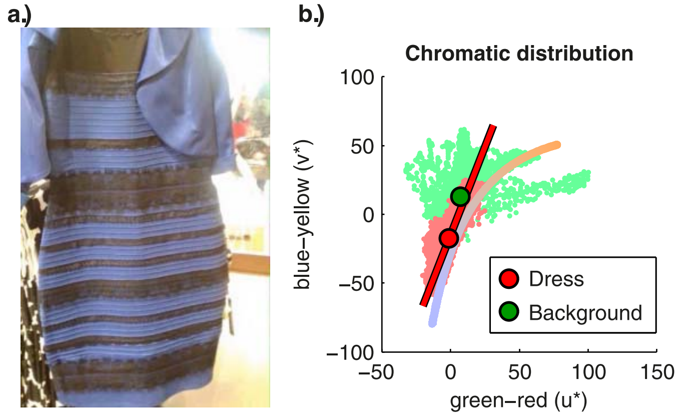

Figure 1. Photo of the dress. (a) Original photo of the dress (Swiked, 2015). Many observers see the dress in this photo as striped gold and white, and many others see it as black and blue. Disambiguating the illumination as in Figure 2 and Figure 10 influences the perception of dress colors. (b) Color distributions in the photo. Light red dots show the colors of the dress and light green dots the colors in the rest of the photo (i.e., the background). The saturated red and green disks show the average color of the dress and the background, respectively. The red line illustrates the main variation of the color on the dress through the first principal component. The colored curve shows the Planckian locus to illustrate the variation of daylight. Note that the colors of the dress vary approximately parallel to the daylight locus (Gegenfurtner et al., 2015), and the average color of the background is aligned with the variation of colors in the dress. Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

observers disagree fundamentally with their color judgments.

Unconscious inferences

The perceptual interpretation of a photo is not a conscious interpretation and does not involve an explicit decision about the interpretation. Instead, the interpretation occurs automatically when looking at a photo and in general when looking at a realistic scene or anything else in the natural environment.

The idea of automatic interpretation goes back to the notion of "unconscious inference," proposed by Hermann von Helmholtz (1924 [1867]), and it is supported by research on color constancy (Foster, 2011). Color constancy is the ability to perceive and recognize colors independently of changes in illumination (Hurlbert, 2007). In color constancy, observers automatically perceive the colors of objects and surfaces, which requires an interpretation of the illumination. At the same time, observers are largely unaware of the color of the illumination, a fact evidenced by the difficulty in estimating illuminations (e.g., Granzier, Brenner, & Smeets, 2009; for review, see Foster, 2011). For this reason, lay observers are often unaware of the fact that the perceived color of an object depends on the implicit perceptual interpretation of the illumination.

This is true, at least, for most everyday life situations. There are a few exceptions, in which color constancy breaks down, and we become aware of the important role of the illumination. A case in point is when a recently purchased item of clothing does not look the same at home as it did in the shop. In such cases, the perceived color of the garment does not fit the color expected by the implicit interpretation of the scene and its illumination. In contrast, in most cases, color constancy provides a coherent, unambiguous perception of a color, so that observers are generally unaware of the important role of the illumination for the perception of color.

We share the idea with many other color scientists that the photo of the dress is a particular case of color constancy (Brainard & Hurlbert, 2015; Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Winkler et al., 2015). In particular, we consider that the uniquely ambiguous illumination in the photo allows two interpretations of the same scene, which are both completely coherent with the particular properties of the scene (depicted objects and background) and with the principles of color constancy. This ambiguity depends on particular properties of this photo of the dress.

Particularities of the photo

The photo of the dress has attracted unprecedented attention all over the world because such high individual differences are not found for most photos or images (Macknik et al., 2015; Swiked, 2015). We propose that the photo of the dress is a special case because it combines an unusually high degree of ambiguity with photorealism, which convinces observers that they are seeing a real scene on the photo.

Following previous suggestions (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015), we think that the unique ambiguity of the photo is due to very particular properties of the color distributions in the photo. These particularities are illustrated in Figure 1b.

First, the dress has two kinds of stripes that differ in their colors: the *body* (which has the same color as the jacket) and the *lace* (which includes the stripe directly around the neck). As illustrated by the light red dots in Figure 1b, the two color distributions that correspond to the two kinds of stripes vary in the same hue direction (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015). One common principal component alone explains 86% of the variation of the two color distributions of the dress stripes (red line in Figure 1b).

Another particularity of the photo is the color of the background. The overall color of the background is indicative of the illumination in the background. In the case of this photo, the background appears to be under a yellow illumination, and the average of the color distribution of the background is very close to the principal component of the stripe colors (green disk in Figure 1b).

Because of these two particularities, the information about the colors of the dress and about the colors of the illumination are confused in the photo. As a result, the photo allows for two coherent interpretations: The color contrast between the dress and the background can be attributed either to the color of the dress or to a difference between the illumination in the background and the illumination that reaches the dress.

A third particularity of the photo increases the uncertainty about the possible interpretations. The principal component that represents the hue direction of the dress colors and its relation to the color of the background (red line in Figure 1b) closely follows the daylight locus (black curve in Figure 1b). It has been shown that observers are particularly uncertain about the estimation of colors along the daylight locus (Pearce, Crichton, Mackiewicz, Finlayson, & Hurlbert, 2014; Witzel, Valkova, Hansen, & Gegenfurtner, 2011). Consequently, the coincidence of the colors in this photo with the daylight locus further increases the ambiguity of the photo.

Gegenfurtner and colleagues (2015) have shown that the alignment with the daylight locus is fundamental to the ambiguity of the image and the divergence of percepts. Rotating the color distribution of the image to the orthogonal color direction disrupts the phenomenon: In this case, all observers reported similar colors of the dress (e.g., pink for the body when the color distribution was rotated by 90°).

However, the color distribution alone does not completely account for the ambiguity of the dress. This is evident by the observation that images with spatially scrambled colors of the dress do not lead to the same individual differences in color perception as the original dress color (Hesslinger & Carbon, 2016). Hence, the spatial arrangement and the meaningfulness of the photo seem to be important for its ambiguity.

A fourth peculiarity of the photo is that observers are firmly convinced about the colors they see on the photo. Only very few observers can flip their perception of the dress colors from blue-black to white-gold and vice versa (Lafer-Sousa et al., 2015). We think this is because the photo looks sufficiently realistic to convince observers that it displays a real dress and is not artwork. This is a fundamental difference with many visual illusions and bistable images that also allow for two or more interpretations. For such visual illusions, most observers flip between the different interpretations once they understand the principle. This is not the case for the dress. As a consequence of the realism of the photo, observers believe that their initial interpretation of the photo is veridical, in the sense that it captures the presumed reality depicted on the photo. This explains why only a few observers can flip their perception of the dress: We can believe in only one reality.

Objective

According to our explanation, the ambiguity and the realism of the photo compel observers to interpret the ambiguous illumination in one or the other way to disambiguate the photo and make sense of the depicted scene. This interpretation requires implicit assumptions about which aspects of the tints in the photo are due to the illumination and which are due to the dress. These implicit assumptions disambiguate the illumination in the photo and determine the colors observers perceive on the dress.

Figure 2 illustrates the idea by disambiguating the cues about the illumination. It shows the same image of a dress as in Figure 1, cut out and pasted into the shadow. In this context, the cues about the illumination suggest that the light of the sunset comes from behind the woman, and hence the front of the woman and the dress are in the shadow. The illumination in the shadow is darker and more bluish. The observer will automatically attribute low lightness and bluishness to the shadow, and the dress will appear comparatively light and less blue in this context. As a consequence, the dress in Figure 2 should appear gold and white to most observers. Figure 10 provides a complementary image that disambiguates the illumination so that the colors of the dress appear blue and black.

The idea that the individual differences in the perception of the dress are related to color constancy and the estimation of the illumination is not unique to our account but common to most explanations proposed in color science (Brainard & Hurlbert, 2015; Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Macknik et al., 2015; Winkler et al., 2015). However, there is also an important difference. Previous speculations about the dress (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Winkler et al., 2015) had suggested that the ambiguous perception of the dress is due to general differences in how observers perceive achromatic colors (gray and white) and the color of the illumination. In particular, one study claimed that differences in perceived dress colors are due to different general priors about the illumination. According to this idea, "some people favor a cool illuminant (blue sky), discount shorter wavelengths, and perceive white/gold; others favor a warm illuminant (incandescent light), discount longer wavelengths, and see blue/black" (Lafer-Sousa et al., 2015, p. R545).

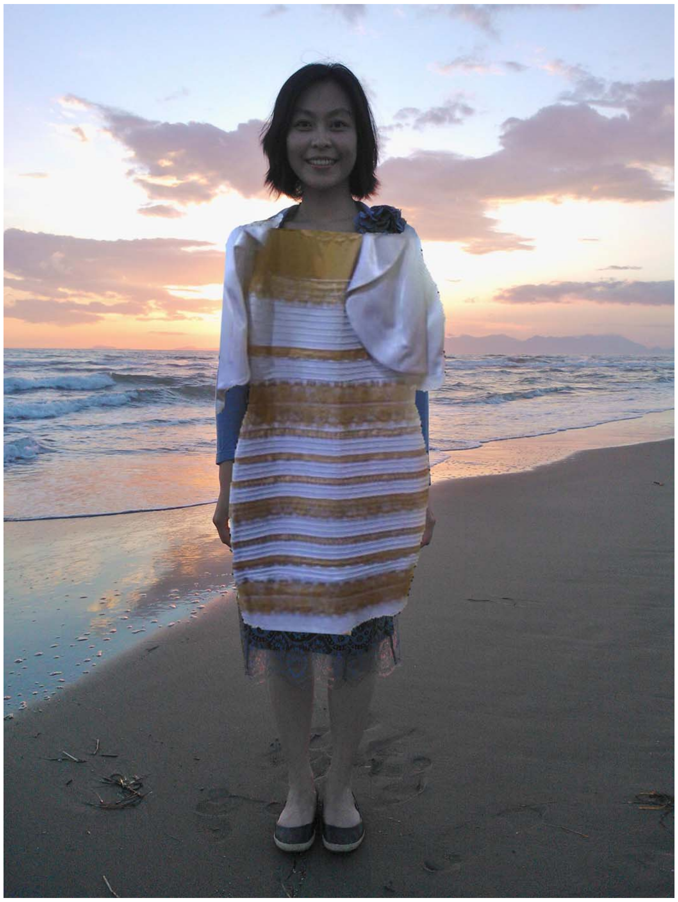

Figure 2. Dress in shadow. The same picture of the dress as in Figure 1 cut out and pasted into in the shadow where it appears gold and white to most observers. Figure 10 shows a context in which the dress appears blue and black to most observers. To see the differences in color perception between Figure 2 and Figure 10 more clearly, they are shown at separate locations in the text. Ideally, it is best not to have previously seen the other photos. Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

Another study claimed that "an important contributing component of the color appearance of the dress, and why it varies across observers, is the relatively greater ambiguity in the blue-white boundary, which may increase the tendency to perceptually discount the blue" (Winkler et al., 2015, p. R548). This tendency to discount blue implies a *blue bias* that is a tendency to perceive gray as slightly bluish (Pearce et al., 2014; Wuerger, Hurlbert, & Witzel, 2015).

Unlike these attempts to explain the dress, we do not think it is necessary to assume fundamental differences in color perception across individuals that are independent from the interpretation of the photo. Instead, we propose that the perceived colors of the dress are determined by the unconscious inferences and assumptions observers spontaneously come up with to make sense of the scene on this particular photo. These assumptions may well be specific to the interpretation of the photo and do not necessarily imply some fundamental individual differences in color priors.

To test our explanation, we first examined the impact of assumptions about the illumination on the perception of the dress colors. For this purpose, an experiment was conducted to test whether perceived dress colors depended on the assumed illumination. Preliminary results of this experiment have been presented in conference papers (Witzel, Hurlbert, & Wuerger, 2016; Witzel & O'Regan, 2015; Wuerger et al., 2015) and communicated in a television broadcast on visual illusions (M6, 2015).

Second, we examined the idea that the assumptions about the illumination in the scene are produced spontaneously out of the need to disambiguate the photo and make sense of the scene. In this case, it should be possible to bias which colors observers will perceive on the dress by suggesting one or the other interpretation prior to seeing the ambiguous photo of the dress. For this purpose, we conducted a questionnaire-based experiment.

Illumination and dress colors

This first experiment tested whether the perceived dress colors depend on observers' assumptions about the illumination. For this purpose, we measured the perceived colors of the dress, the estimated color of the illumination in the photo, and the subjective gray point of the dress and a neutral disk.

To determine the perceived colors of the dress and the observers' estimations about the illumination, we used (a) a color-naming and (b) a color-adjustment task. In the color-adjustment task, we asked observers to match the color of a disk presented on a computer monitor to each of the two stripes of the dress (the blue/white dress body and the white/gold dress lace) and—in a later part—to the color of the illumination that they think reaches the dress. Then we tested whether the observers' estimations about the illumination of the dress predicted the observers' perceived colors of the dress, as predicted by our explanation.

Moreover, we examined those previous claims according to which the individual differences in the perceived colors are related to general differences in how observers perceive gray-scale (achromatic) colors (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Winkler et al., 2015). To make this argument, previous observations of individual differences in gray adjustments along the daylight locus (Chauhan et al., 2014; Witzel et al., 2011) have been taken as support for the idea that there are individual differences in priors about

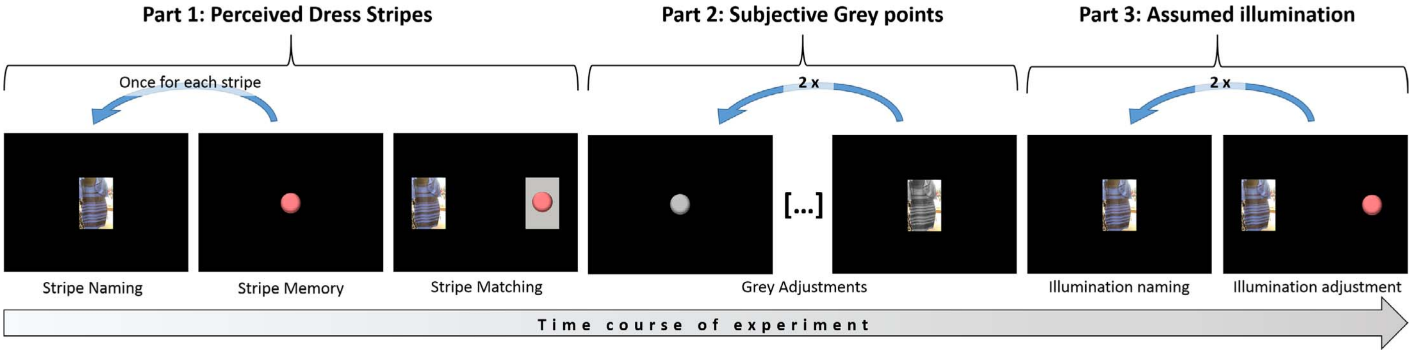

Figure 3. Overview of the experiment. Each frame illustrates the stimulus display of one particular task. Blue arrows indicate loops of iterations/blocks. Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

the illumination (Brainard & Hurlbert, 2015; Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015). These gray adjustments reflect the *subjective gray point*. A subjective gray point is the color that observers subjectively perceive as achromatic gray. The subjective gray point reflects the chromaticity of the illumination and determines the perceived chromaticity of all achromatic (gray-scale) colors in a scene. Because achromatic surfaces reflect the chromaticity of the illumination, the gray adjustments also indicate what observers assumed to be the chromaticity of the illumination.

To measure the chromaticity of the subjective gray point, we asked observers to adjust the dress and the disk, respectively, so that they looked gray to them. We distinguish between a *generic subjective gray point* and a *specific subjective gray point*.

The generic subjective gray point reflects the general perception of gray-scale colors, independent of a particular context and image. The gray adjustments of the disk provide estimations of the *generic subjective gray point*, which reflects the chromaticity of the generic subjective white point. Moreover, these measurements are equivalent to those from figure 6b of Witzel et al. (2011) that Gegenfurtner et al. (2015, see their figure 1) considered when speculating about the perceived colors of the dress. If individual differences in the blue bias (Pearce et al., 2014; Winkler et al., 2015) or in prior expectations toward the illumination (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015) matter for the perception of the dress, then the perceived dress colors should be related to the individual differences in the generic subjective gray point.

In contrast to the generic subjective gray point, we call the specific subjective gray point the subjective appearance of gray in the special case of the dress photo. Because of the particularities of the dress and its background, different observers might assume a different gray point for the scene on the photo. The gray adjustments of the dress within the context of the photo are supposed to assess the *specific subjective gray point*. Because the subjective gray point of the photo should reflect the observers' assumptions about the illumination in the photo, the gray adjustments of the dress should produce similar results as the direct measurement of the assumed illumination.

Method

All measurements were done in one session as a single experiment with three parts. Hence, all measurements involved the same participants and the same setup.

Participants

Thirty-one observers (24 women, seven men; 30.7 ± 10.8 years old) participated in the study. Red-green color vision deficiencies were excluded through Ishihara plates (Ishihara, 2004) and in a few cases by self-report.

Apparatus

Stimuli were displayed on a ViewSonic PN5f+ CRT monitor driven by a NVIDIA GeForce 8400 GS graphics card (NVIDIA Corporation, Santa Clara, CA) with a color resolution of eight bits per channel, a spatial resolution of 1,280 × 1,024 pixels (at a size of 36.5 × 27 cm), and a refresh rate of 85 Hz. The CIE1931 chromaticity coordinates and luminance of the monitor primaries were R = [0.615, 0.351, 14.4], G = [0.295, 0.600, 43.5], and B = [0.144, 0.076, 5.16]. Gamma corrections without bit loss were applied based on the measured gamma curves of the monitor primaries. All measurements were done in a dark experimental booth, and observers looked at the screen from a distance of about 50 cm.

Overview of stimuli and procedure

Figure 3 gives an overview of all tasks, presents the corresponding stimulus displays, and illustrates the time course of the experiment. The first part of the experiment measured the perceived stripe colors of the dress through color naming (*stripe naming*) and adjustments (*stripe adjustments*). The second part determined subjective gray points through *gray adjustments*. The measurements of the assumed illumination were done by color naming (*illumination naming*) and adjustments (*illumination adjustments*) in the third and last part. Overall, the experiment took about 10–20 min.

All three parts of the experiment included the original photo of the dress shown in Figure 1 and a disk with uniform chromaticity and a luminance texture that made it appear three-dimensional and bumpy (cf. Figure 4). In all tasks, instructions and stimuli were presented on a black background (cf. Figure 3). Instructions were presented in the color of the monitor white point ($xyY1931 = [0.3109\ 0.3475\ 63.1\ cd/m^2]$). This white point was also used to represent the color distributions of the images in CIELUV space.

In the color-naming tasks (stripe and illumination naming), the photo of the dress was presented in the center of the otherwise black screen. Observers were asked to describe the colors of each stripe of the dress and the illumination of the dress, respectively, through a color term. The 13 available color terms included the 11 basic color terms (pink, red, orange, yellow, green, blue, purple, brown, black, gray, and white) as well as gold and bronze.

To illustrate the color-adjustment method and results, the Supplementary Material provides videos of mock trials for each part of the experiment. Color adjustments were implemented through a polar adjustment technique (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006; Witzel et al., 2011). Observers could press one of four keys to add yellow, blue, green, and red in the images. These changes were translated into polar coordinates in CIELUV space. Changes in azimuth imply that the color distribution is rotated and hence changes in hue. Changes in radius mean that the color distribution is rescaled (compressed or expanded) and hence changes in saturation. This technique reproduces the same color distributions in a different hue direction. At the same time, it allows for adjusting the images to be completely achromatic by canceling opponent colors, which implies a scaling of saturation to zero (for illustration, see Supplementary Figure S1).

In the stripe-adjustment task, observers could also adjust the lightness (L*) by pressing one of two keys. In the gray adjustments and in the adjustments of the illumination, lightness was fixed and only chromaticity could be adjusted. Observers could switch between a coarse and fine adjustment mode. In the coarse adjustment mode, they could keep a key pressed to surf continuously along a color dimension. In the fine adjustment mode, one key press corresponded to a shift of 1 CIELUV unit along the respective CIELUV dimension. Observers had to do fine adjustments before confirming the adjusted color and continuing with the next trial.

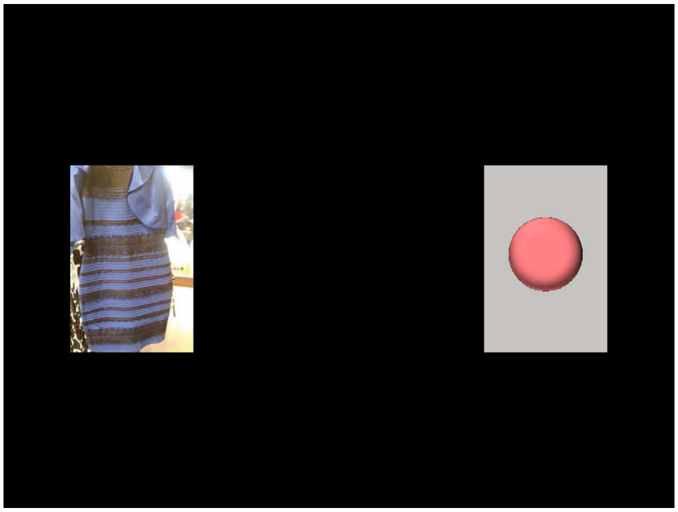

Figure 4. Stimulus display in the stripe-matching task. The color of the disk (here pink) was initially random. The display in the illumination adjustment task was the same, except for the disk being presented on a black rather than a gray local background. Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

In all color adjustments, the adjustable image was presented in a random color at the beginning of a trial. When adjustments reached the monitor gamut, this was indicated through a sound and a message on the screen. In this case, adjustments were locked until the observers pressed keys to shift the colors back into the monitor gamut (which were indicated in the computer message). To become familiar with the color-adjustment procedure, observers completed a practice trial before starting the main parts of the experiment.

Measurements of perceived stripe colors

In the stripe adjustments, observers were asked to adjust the color of the disk to match the color of the dress body and of the lace in separate trials. For these adjustments, the disk was placed in the center of a gray rectangle, which had the chromaticity of the white point ($xyY1931 = [0.3109, 0.3475, 31.5\ cd/m^2]$) and the size of the dress photo. The gray local background of the rectangle was used to control local contrast and make stripe adjustments more precise.

There were two versions of stripe adjustments. In one version, the original photo of the dress was presented simultaneously with the adjustable disk (*stripe-matching task*). Here, the adjustable disk was shown in a random color on one side of the monitor and the image of the dress on the other side (cf. Figure 4). In the other version, the original photo was presented first, followed by the adjustable disk so that observers had to do the adjustments by memory (*stripe memory task*).

The first part started with the stripe-naming task. There were two blocks: the first one for the lace and the second one for the body of the dress. A block started by presenting the original photo of the dress shown in Figure 1. Observers were asked to memorize the color of the stripe (body or lace) that corresponded to the respective block. To continue, participants had to name the color of that stripe. Then the randomly colored isolated disk was presented, and observers adjusted the color of the disk to match the respective stripe from memory. After the two blocks followed two blocks of stripe matching, one for each of the two stripes.

Measurements of subjective gray

In the gray adjustments, only one image was presented in the center of the screen. Observers were asked to adjust the random initial color of the respective image so that it looked gray to them (Hansen et al., 2006; Witzel et al., 2011).

The images in the gray adjustments included the disk and three versions of the dress. The problem with the original photo of the dress is that it has a bimodal color distribution due to its two kinds of stripes. This means that chromaticities cluster in the opponent blue and brown regions of color space. For this reason, deviation from gray (as defined by the white point above) is signaled by the coloration in the opponent color directions of the two stripes, and hence observers have much stronger cues about deviations from the gray point. This might go counter to the objective of our measurements because it might reduce the possibility of measuring the observer's uncertainty about the chromaticity of the white point. For this reason, we included a version of the dress with uniform hue, in which only saturation and luminance varies. Adjustments of this version of the dress imply uniform changes in colors, such as giving the whole dress a red tint (for illustration see Supplementary Figure S2).

In these two versions, the adjustable dress was shown within the fixed background of the original photo. A third version featured the dress without the background from the photo: It was cut out from the photo and pasted on a uniform black background. There were two blocks, and in each block, all four objects were adjusted one after another in random order.

In this task, all images, including the disk, were shown on a completely black background. This was done because the gray background around the disk in the stripe adjustments would allow for matching the disk to the background color. This would counteract our aim of measuring observers' generic subjective gray point.

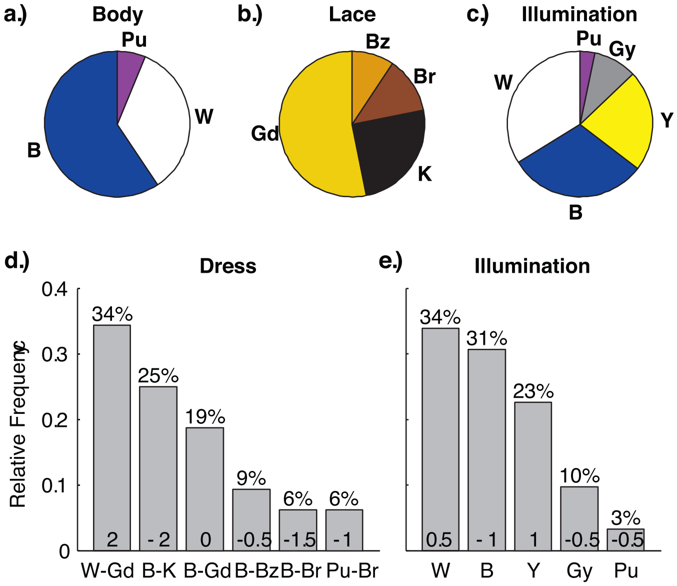

Figure 5. Color-naming data. (a, b) Dress stripe naming. (c) Description of the illumination by naming. (c, d) Relative frequencies of each color term combination and dress and illumination scores at the bottom of the bars. B = blue; Pu = purple; W = white; Gd = gold; Bz = bronze; Br = brown; K = black; Gy = gray; Y = yellow.

Measurements of assumed illumination

In the third part, observers completed two blocks, each with one trial of illumination naming and illumination adjustment.

At the beginning of a block, the photo of the dress was presented. By choosing one of the 13 color terms, observers were asked to describe the color of the light that they thought shines on the dress.

Then followed the illumination adjustment. The same stimulus display as in the stripe adjustments was presented (Figure 4), but the disk was shown on a completely black background (i.e., without the gray rectangle). Observers were asked to adjust the color of the disk to match the colors of the illumination that they thought reached the dress. The adjustments of the assumed illumination were done on a disk while the original photo was presented simultaneously.

Results and discussion

For comparison with other data, Figure 5a, b, and d illustrates the stripe-naming data. In this task, naming for each of the two stripes (dress body and lace) was assessed independently in separate trials. In most cases, the stripes of the dress were called "white-gold" (34%). Then followed "blue-black" (25%), and a combination of both, "blue-gold" (19%). The latter combination is understandable, if we consider that blue includes light blue because there was no response option for light blue. Observers rarely called the stripes of the dress "blue-bronze" (9%), "blue-brown" (6%), and "purple brown" (6%). Interestingly, these proportions do not completely agree with those observed by Lafer-Sousa et al. (2015), who reported a higher proportion of blue-black than white-gold.

Principal components of adjustments

The colors observers estimated for the dress stripes and the illumination precisely reflect the particular properties of the photo. Figure 6a, b illustrates the average adjustments for the dress body (blue dots) and the dress lace (brown dots) by simultaneous matching and by memory, respectively. When lumping the adjustments of the body and the lace together, the first principal component (red line) explained 91% and 89% of the variance of the adjustments for both stripes. The main variation captured by the principal component closely follows the variation of daylight (colored curve in Figure 6a). Consequently, the variation of perceived dress colors across observers has the same pattern as the color distribution in the photo. This is in line with previous findings (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Winkler et al., 2015).

However, the alignment of the principal component with the daylight locus is a necessary implication of the fact that the colors of the body and the lace differ along the daylight locus. What is important for the individual differences in perception is whether the observers' variation in perception of each of the stripes may be represented by one principal component that varies along the daylight locus. In other words, the question is: Do the adjustments of the blue stripes covary with the adjustments of the brown stripes? To show this, we need to consider the adjustment of each of the stripes separately and independently of each other. So, we calculated a six-dimensional principal component decomposition, that is, with three dimensions (L*, u*, v*) for each stripe (body and lace). It turns out that the first principal component still explained 62% of the variance for both the adjustments by simultaneous match and by memory. These principal components intersected all six dimensions of the matchings (loadings for L*, u*, and v* of lace: [0.25, 0.22, 0.33]; body: [0.15, 0.15, 0.86]) and the adjustments by memory (lace: [0.31, 0.25, 0.55]; body: [0.18, 0.16, 0.69]). This shows that body and lace adjustments covary with 62% common variance along a single dimension in that six-dimensional space.

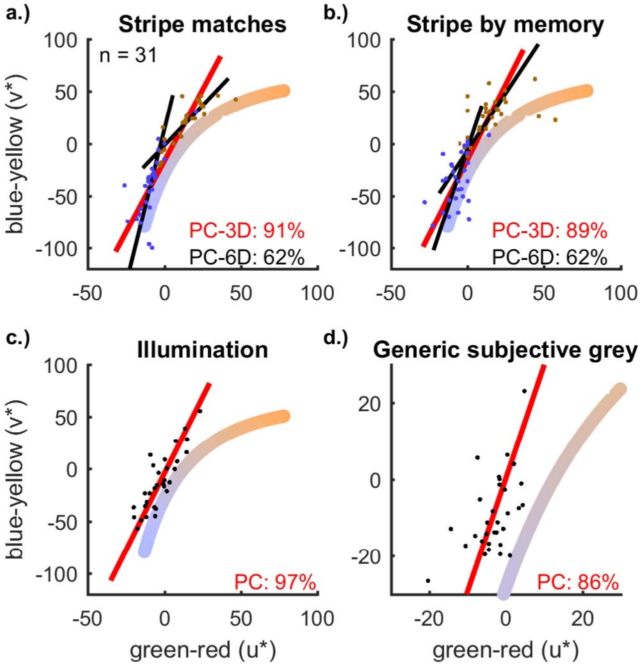

Figure 6. Adjustments of colors matching the dress (a, b), illumination (c), and gray (d). (a, b) Average adjustments of the colors of the dress body (blue dots) and lace (brown dots) for each observer in the simultaneous matches and the adjustments by memory, respectively. (c, d) Average adjustments of the assumed illumination and the gray adjustments of the disk, respectively (black dots). In all panels, the x- and y-axes represent the green-red (u*) and blue-yellow (v*) axis of CIELUV color space. The red lines and the colored curve in these panels show the first principal components of the adjustments and the daylight locus, respectively. In panels a and b, the red line refers to the three-dimensional and the thin black lines to the six-dimensional principal component decomposition. The percentages refer to the variance explained by the principal component. Note that all adjustments vary parallel to the daylight locus.

This six-dimensional principal component is illustrated by the thin black lines in Figure 6a, b. These lines are actually two parts of the same principal component. Because both adjustments of the lace and the body are done in the same color space (CIELUV), we can plot the second and third dimension (u* and v* of the body) and the fifth and six dimension (u* and v* of the lace) into one graphic, resulting in the two black lines of the six-dimensional principal component. These two black lines nicely follow the curvature of the daylight locus. The location of colors along the daylight locus is captured by the correlated color temperature: A high correlated color temperature implies colors toward the blue ("lower") end and a low correlated color temperature toward the yellow end of the daylight locus. Hence, the higher the correlated color temperature of the body adjustments, the higher the correlated color temperature of the lace adjustments. In other words, the observers' perception of the whole dress shifts along the daylight locus.

Further analyses of the simultaneous matches clarify the common pattern of adjustments that is represented by that first principle component. The lightness adjustments (L* axis) of the (blue/white) dress body were strongly correlated with all dimensions of the adjustments of the (black/gold) lace, $r(30) = 0.62$, $p = 0.0002$; $r(30) = 0.54$, $p = 0.001$; $r(30) = 0.66$, $p < 0.0001$. Moreover, the yellow-blue adjustments (v* axis) of the dress body were strongly correlated with the lightness adjustment (L*) of the lace, $r(30) = 0.62$, $p = 0.0002$. The important role of lightness in the estimation of the dress colors is also in line with previous findings (Gegenfurtner et al., 2015).

Figure 6c illustrates the average adjustments of the illumination that observers believed reaches the dress (black dots). The principal component of these adjustments (red line) explained 97% of the variance and closely followed the daylight locus (colored curve). This result clearly shows that the observers assumed a color of the illumination in this photo somewhere along the daylight locus. Consequently, the interobserver variability of the estimated colors of the dress (see above) and of its illumination directly reflect the variation of color distributions (Figure 1b) that are assumed to be the source of the ambiguity of the photo.

Figure 5c, e illustrates the relative frequencies of the illumination naming. The illumination that reached the dress was called "white" most frequently (34%), closely followed by "blue" (31%) and then "yellow" (23%). In some cases, it was also called "gray" (10%) and rarely "purple" (3%). Observers did not use any other color term to describe the color of the illumination in this task. The color terms used closely reflect the variation of color adjustments along the daylight locus, too.

Based on these observations, we recoded the naming data into a *dress score* that is more useful for the main analyses below. The dress score is a quasi-metric index of color naming that indicates whether observers' color naming was closer to black-blue or to white-yellow on the blue/dark versus white/light dimension, along which the perception of the dress varies (cf. Figure 5d). Hence, white-gold and blue-black were considered as the two extrema, and these combinations were coded as 2 (1 for white and 1 for gold) and −2 (−1 for blue and −1 for black). Because "gold" corresponds to 1 and "blue" to −1, the combination "blue-gold" results in a value of 0. Because purple includes bluishness and bronze shares similarity with gold, these answers were coded as +0.5. Inversely, because brown is relatively dark but not yet black, it was given a value of −0.5. As a result, "blue-bronze," "purple-brown," and "blue-brown" corresponded to values of −0.5, −1, and −1.5 (cf. Figure 5d).

A similar approach was used to convert the illumination naming data into *illumination scores* (cf. Figure 5e). Illumination scores varied between −1 for blue and +1 for yellow. White was coded as −0.5 because it refers to bright light but not exactly to the bright yellow light in the background. Gray and purple were coded +0.5 because they were in line with the idea that the dress is in a shadow or a bluish illumination. All intermediate values result from averaging the above values (−1, −0.5, 0.5, and 1) across the two repeated measurements.

Finally, all gray adjustments (disk and three versions of dress) varied along an axis approximately parallel to the daylight locus. In particular, we measured the chromaticity of the generic subjective gray point through gray disk adjustments. When observers adjusted the color of the disk (Supplementary Figure S1a) to gray, the adjustments varied mainly along a principal component (86% of variance) close to the daylight locus (Figure 6d). This is in line with previous measures of generic subjective gray points (Chauhan et al., 2014; Witzel et al., 2011).

The same pattern could be observed for the three versions of the dress: The first principal component of the adjustments of the dress in the context of the photo (Supplementary Figure S1b), the dress with uniform hue (Supplementary Figure S1b and Figure S2), and the dress without background (Supplementary Figure S1c) explained 86%, 92%, and 84% of variance across observers, respectively (cf. panel a of Supplementary Figures S5–S7). These results reconfirm those shown for gray adjustments of objects in general (figure 6 in Witzel et al., 2011) and support previous speculations that the white point of the dress is uncertain in chromaticity along the daylight locus (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015).

Main results

To disambiguate and interpret the photo, observers have to make an assumption about the illumination. If this assumption is the source of the striking differences in the perceived colors of the dress, the perceived colors of the dress depend on the assumed illumination. In particular, the more blueness observers attribute to the illumination, the yellower they should perceive the dress and vice versa. To test this hypothesis, we examined whether the adjustments of the dress and of the illumination were negatively correlated. Results confirmed our predictions.

Because the systematic variation of the adjustments could be represented by their principal components, we first calculated the correlations between the scores of the principal components of the two kinds of adjustments (i.e., the projections on the black and red lines in Figure 6a–c). Here, we focus on the six-dimensional

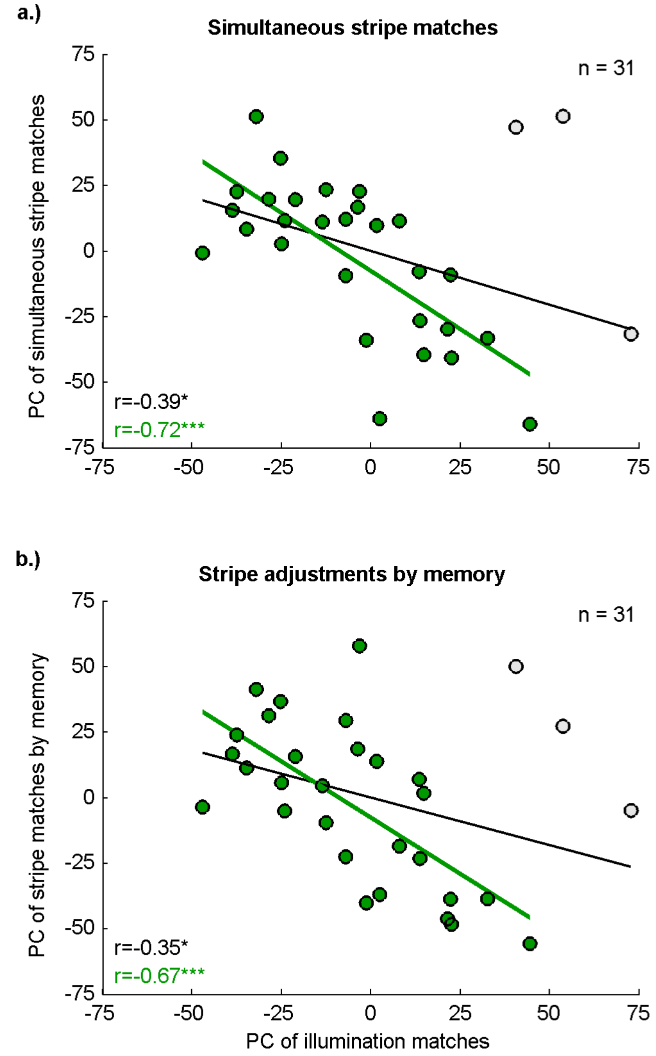

Figure 7. Correlations between the adjustments of dress and illumination colors. (a) Results involving simultaneous matches of stripes. (b) Results for stripe adjustments by memory. The axes correspond to principal component scores, that is, the projections of the adjustments on the principal components (black and red lines) in Figure 6a and Figure 6b. White disks indicate scores that were identified as outliers in robust correlation analyses. The black line shows the regression including all data points (white and green disks); the green line shows the regression excluding outliers (only green disks). Corresponding correlation coefficients are shown in the lower left corner. The number of participants is given in the top right corner. To report the variability of the illumination adjustment scores, Supplementary Figure S3 provides the same graphics but with error bars for the two measurements of estimated illumination.

principal components of the dress adjustments, but similar observations were made when using averages across lace and body along the three-dimensional component. Figure 7a illustrates the relationship between the principal component scores for the assumed illumination (x-axis) and for the perceived dress colors measured by simultaneous matching (y-axis). As predicted, the scores of the dress's colors were negatively correlated with scores of the illumination ($r = -0.39$, $p = 0.01$, one-tailed; black line in Figure 7a).

At first look, this correlation is only moderate. To guarantee the reliability of this observation, we calculated robust correlation indices, which provide better estimates of the true relationship between variables (Pernet, Wilcox, & Rousselet, 2012). All robust correlation analyses were significant (see Table S1 in the Supplementary Material for details). Most important, these analyses identified outliers based on objective criteria (white disks in Figure 7a). A *skipped Pearson correlation* that controls for outliers (i.e., involving only the green disks in Figure 7a) indicates a very strong correlation that explains more than 50% of the variance, $r(29) = -0.72$, $p < 0.001$, one-tailed.

Moreover, the relationship between the estimated illumination and the perceived dress color could be reproduced in other measurements. First, we observed similar negative correlations when using the principal component scores of the stripe adjustments by memory (Figure 6b) instead of the simultaneous matches, $r(29) = -0.35$, $p = 0.03$, and skipped correlation, $r = -0.67$, $p = 0.0001$, both one-tailed. We also found similar results when using the dress and illumination scores of the color naming instead of the color adjustments. As for the adjustments, the dress scores were negatively correlated with the illumination scores, $r(29) = -0.43$, $p < 0.01$, one-tailed. Similar correlations—that is, between $r(29) = -0.45$ and $r(29) = -0.34$ (all $p < 0.05$)—were also found when comparing naming data with adjustment data (for details, see the "Naming" section of the Supplementary Material). These additional results show that the negative correlation between assumed illumination and perceived dress colors reliably occurs for different measures and data sets.

The observed correlations also do not depend on the use of principal components to represent the variation of the adjustments. The correlation of the principal component scores was due to the common variance of the yellow-blue adjustment (v* axis) of the dress body and the lightness adjustment (L* axis) of the lace (see above). Each of these dimensions separately was correlated to the yellow-blue dimension (v*) of the adjustment of the illumination, $r(29) = -0.40$, $p = 0.01$, and $r(29) = -0.36$, $p = 0.03$.

To test the idea that the perceived colors of the dress directly depend on the variation along the daylight locus (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015), we calculated the correlated color temperature of the adjustments. The correlated color temperature corresponds to the projection of colors on (a model of) the daylight locus. The correlated color temperatures of the two kinds of adjustments (dress and illumination) reproduced the negative correlation found with the principal components, $r = -0.43$, $p = 0.01$.

Overall, our observation that the estimated illumination predicts the perception of the dress is robust and reliable. The moderate size of observed correlations is due to outliers. These outliers can be attributed to the fact that it is very unusual and difficult for observers to adjust the color of an illumination in a photo on a disk displayed on a computer screen. In particular, participants may sometimes misunderstand the task and adjust the color of background illumination, rather than the color of the illumination that reaches the dress. Such misunderstandings produce the kind of outliers observed in Figure 4c (white disks). Moreover, the moderate size of the uncorrected correlation coefficients can be explained by the difficulty observers had in making their assumptions of the illumination explicit, in particular because it is well known from many studies that observers have difficulties in estimating illuminations (Foster, 2011). Participants' reports and comments on the experiment support the idea that the illumination adjustment is a particularly difficult task. For this reason, the true relationship between assumed illumination and perceived colors in the dress is most probably better reflected by the very large skipped correlation than by the moderate uncorrected correlation.

Subjective gray point

If general priors about the illumination or the blue bias were the reason for the perceived colors of the dress (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Winkler et al., 2015), the generic subjective gray point should predict perceived dress colors. Hence, the gray adjustments of the disk should be negatively correlated with the perceived colors of the dress: The bluer the generic subjective gray point of an observer, the more likely this observer should be to see the dress in white-gold.

However, this is not the case. Although those speculations were correct about the role of the illumination and the daylight locus in general (see above), the results of the present study show that the concrete predictions derived from those speculations do not hold. Here, we had measured the chromaticity of the generic subjective gray point through the gray adjustments of the disk. These adjustments did not at all predict the perceived color of the dress: simultaneous matches, $r(29) = -0.02$, $p = 0.45$; memory matches: $r(29) = -0.04$, $p = 0.41$ (Supplementary Figure S4b–c).

Moreover, because prior expectations refer to probability distributions for individual observers, we looked more closely at the distributions of gray adjustments. To predict the differences in dress perception, prior distributions must differ systematically between observers who perceive the dress as blue-black and those who see it white-gold. By probing data for different observers, the resulting distribution for

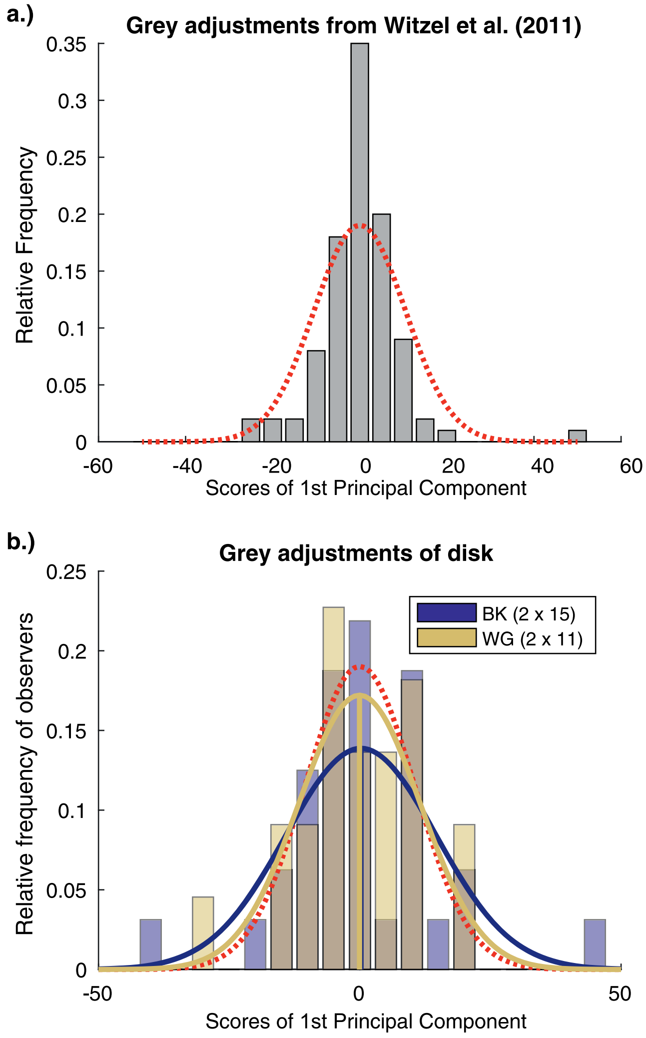

Figure 8. Generic subjective gray point. (a) Gray adjustments of color-neutral objects in Witzel et al. (2011). The y-axis represents the relative frequency of observers and the x-axis the first principal component of the adjustments. Gray bars show binned scores on the principal component, and the red dotted curve shows a normal distribution fitted to the data. (b) Binned scores of the principal components for single trials of disk adjustments are shown for observers with dress scores below (blue bars; BK = blue-black) and above (yellowish bars, WG = white-gold) zero. Gray areas show overlapping of bars. The blue and yellowish curves depict a normal distribution fitted to the blue and yellowish bars, and the vertical lines show the averages of these distributions. The dotted red curve is the same as the curve in panel a. Numbers in the legend report the size of the data sets (number of measurements × number of observers). Note that gray adjustments are approximately normally distributed (a), and there is no difference between the averages for observers with negative and positive dress scores.

each of the two groups of observers should reflect that systematic difference in priors.

The data from figure 6 in Witzel et al. (2011) allow for determining more precisely the distribution of generic subjective gray points. Figure 8a shows the binned scores along the first principal component of those gray adjustments, which also correspond to the aggregated adjustments in figure 1 of Gegenfurtner et al. (2015). There were 300 data points from 25 observers, with four images of objects shown on a uniform gray background (a uniformly colored disk, a disk with luminance noise pattern, a golf ball, and a sock) and with each image being adjusted three times. As illustrated by the red line in Figure 8a, the gray adjustments varied symmetrically around the mean and were approximately normally distributed.

To examine the relationship between dress colors and generic subjective gray points, we compared the gray adjustments of observers with a dress score below (blue-black) and above (white-gold) zero (observers with dress scores of zero were excluded). There were 15 observers in the blue-black and 11 in the white-gold group, each providing two measurements. The binned principal component scores for blue-black seers are shown as blue bars and those for white-gold seers as yellowish bars in Figure 8b. In particular, the white-gold distribution was very similar to the one for the data of Witzel et al. (2011), as shown by the red dotted line. Most important, there was barely any difference between the averages of the two distributions (vertical yellowish line covers vertical blue line).

It may be that the priors of the two groups differ in a more complicated way that cannot be predicted by average gray adjustments and correlations with stripe adjustments. However, our results undermine the simple idea that white-gold observers have a stronger general tendency to discount for bluish colors than blue-black observers.

In contrast to the generic subjective gray point, evidence was found for a correlation between perceived dress colors and the specific subjective gray point for the photo: Gray adjustments of the dress with uniform hue were negatively correlated with simultaneous stripe matches, $r(30) = -0.38$, $p = 0.02$ (Supplementary Figure S5b). This correlation was also statistically robust (all $p < 0.05$; cf. Supplementary Table S1). The correlation with the stripe matches by memory did not reach significance but went to the same direction ($r = -0.27$, $p = 0.06$; Supplementary Figure S5c). A multiple regression with the gray adjustments of the dress and the illumination adjustments as factors predicted together $R^2 = 34\%$ and 23% of variance of the stripe adjustments by simultaneous matching and by memory, respectively, $F(28) = 6.9$, $p < 0.01$, and $F(28) = 4.0$, $p = 0.03$).

These correlations are particularly interesting in the light of the correlations found for the illumination adjustments: Although the gray adjustments of the disk refer to a general bias in assumed illumination, the gray adjustments of the dress are an indirect measure for the chromaticity of the illumination observers assumed in this particular photo. Hence, the results for the gray adjustments of the disk further confirm the results found with the illumination adjustments.

The gray adjustments of the dress with the original color distribution and of the dress without background did not yield any significant negative correlation (Supplementary Figures S6–S7). We suppose that this is because in both cases, the color distribution of the dress is bimodal and limits the variation of the dress adjustments. This suspicion is supported by the fact that gray adjustments of the dress without background varied much less than for the disk and the dress with uniform hue (Supplementary Figure S7b, c). This lack of variation undermines a systematic correlation with the perceived dress colors.

In sum, these additional results show that the individual differences in the perception of the dress are not due to prior expectations about illuminations or a blue bias in general. What really matters for the differences in the perception of the dress are the individual differences in the interpretation of the illumination in this particular photo. This conclusion is further supported by the findings of another study conducted in parallel to ours (Chetverikov & Ivanchei, 2016). Based on online surveys, that study showed that the perceived colors of the dress depended on assumptions about the light source but not on the generic subjective gray point.

Questions about the scene

Finally, to get an idea about how observers interpret the photo, we also asked some of them what kind of light reaches the dress in the depicted scene of the photo. In particular, we asked them whether the illumination of the dress is the same as the one that reaches the background, whether the dress is in sunlight, whether it is in the shadow, and whether it is illuminated by the flash of the camera (for detailed questions, see the "Questions About Illumination" section in the Supplementary Material). Results are illustrated in Figure 9. Not one of these questions was answered consistently across observers. This suggests that different observers give very different answers in the case of this photo, confirming that the cues about illumination in this photo are highly ambiguous. So, we wondered whether we could suggest to observers how to interpret the illumination in the scene so as to influence what colors they ultimately perceive on the dress.

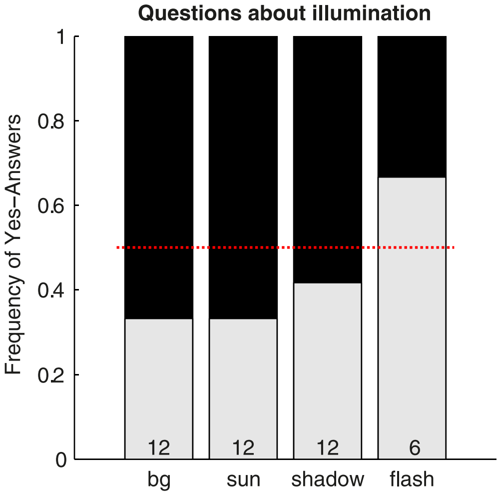

Figure 9. Questions about illumination. Each bar corresponds to one of the four questions about the lighting of the dress (bg = "same as background"; sun = "in sunlight"; shadow = "in shadow"; flash = "exposed to flash"). The proportion of white bars corresponds to yes and black to no answers to those questions. The dotted red line illustrates chance level ($p = 0.5$). The numbers at the bottom of the bars correspond to the number of observers (the flash question was added later on; hence, the lower number of observers).

Biasing perception

According to our explanation, observers engage spontaneously with one of the two possible interpretations to disambiguate and make sense of the photo. In this case, we should be able to steer which interpretation observers "choose" by suggesting to them one of the two interpretations prior to showing them the original ambiguous photo. Once observers have a ready-made interpretation at hand that allows them to make sense of the photo, there is no need to find another perceptual interpretation. In this case, they should stick to the suggestion and perceive the photo of the dress accordingly.

To influence the observers' interpretation of the photo, we showed them either one or the other disambiguated image of the dress (Figure 2 or Figure 10) prior to viewing the original photo (Figure 1). Figure 2 suggests the interpretation that the dress is in the shadow and hence should appear white-gold. In contrast, Figure 10 suggests that the dress is in the sun and should be seen as blue-black. Then we tested whether the prior view of the disambiguated photos biases the observer's perception of the original photo (Figure 1) toward the one suggested in the disambiguated photo.

We would like to emphasize that the disambiguation of the dress colors in these images (i.e., Figure 2 and Figure 10) is due to the interpretation of the scene, not to low-level perceptual mechanisms, such as color contrast. Previous proposals of disambiguating images have not made this distinction.

Figure 10. Dress in the sun. The same picture of the dress as in Figure 1 and Figure 2 cut out and pasted in the sun where it appears black and blue (note that it is better not to have previously seen Figure 1 and Figure 2 to see this). Showing observers one or the other disambiguated photo in Figure 2 and this figure prior to seeing the original photo in Figure 1 determines how observers see the colors of the dress in the original photo of Figure 1 (see text for detailed results). Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

For example, Lafer-Sousa et al. (2015; see figure S2 there) cut and pasted the dress into the yellow and blue version of the color contrast cube images of Lotto and Purves (Lotto, 2008; Lotto & Purves, 2002; Purves, 2015). The scene in these color contrast cube images shifts the color appearance of any object toward either yellow or blue depending on whether they are shown in a blue or yellow context, respectively. In the original color contrast cube images, this shift in color appearance has been shown for a uniform gray patch on the cubes. This can be done with any object and is not specific to the ambiguity of the dress.

These shifts in color appearance are at least in part due to the strong color contrast between the objects and their surround in the respective image. However, color contrast cannot explain why different observers see different colors in the original photo of the dress. In the case of the original photo of the dress, color contrast never changes across observers as the photo remains exactly the same. In contrast, the interpretation of the scene can change across observers, and our explanation is that the differences in initial interpretation are responsible for the differences in perception.

The images of Lafer-Sousa et al. (2015) do not disentangle whether the stable perception of the dress in either of the two versions (blue and yellow context) is due to color contrast, scene interpretation, or both. Hence, it is unclear to what extent the influence of those blue and yellow contexts on the perceived dress color has to do with the interpretation of the illumination or whether it is a simple effect of color contrast, which cannot explain the phenomenon at all. Similar issues exist for other images (e.g., Macknik et al., 2015) that show the dress once in a dark shadow and once in bright sunlight, so that potential disambiguating effects could potentially be due to lightness contrast.

Here, we wanted to disentangle the effects of scene interpretation, which we assume to be at the origin of the dress phenomenon, from effects of color contrast, which cannot play a role for this phenomenon. For this reason, we created the images in Figure 2 and Figure 10, in which local color contrast contradicts scene interpretation and cannot explain the appearance of the dress colors.

In our images, local contrast and scene interpretation were dissociated: The interpretation of the illumination in the scene produced the opposite perception than what would be expected from local contrast. We did this to show that any observed effects were due to interpretation rather than to low-level sensory effects of local contrast (see the Method subsection of the Biasing perception section for details). We first tested whether the disambiguation in these images works by asking observers to describe the colors of the dress in either Figure 2 or Figure 10. If the interpretation of the illumination influences the perceived color, the dress should be seen as white-gold in Figure 2 due to the shadow, and as blue-black in Figure 10 due to the direct sunlight.

After observers saw one or the other disambiguated photo, we showed them the original photo (Figure 1) and asked them how they see the colors in that photo. Because the prior view of one of the disambiguated photo suggests a clear interpretation of the illumination, we expected observers to transfer this interpretation to the image of the dress in the original photo of Figure 1 to make sense of the illumination conditions in that photo. If the prior view of the disambiguated photo biased the observers' interpretation of the illumination in the original photo, observers should see the dress in the original photo of Figure 1 as white-gold after seeing the image in Figure 2 and as blue-black when they saw the image of Figure 10 before. In other words, the image in Figure 2 is used to induce white-gold and the one in Figure 10 to induce blue-black perception in the original photo of Figure 1. In addition to this main question, we also assessed the observers' assumptions about the illumination and the scene configuration to double-check and further specify the findings of the experiment above. Preliminary results of this experiment have been presented in a conference paper (Witzel, Racey, & O'Regan, 2016).

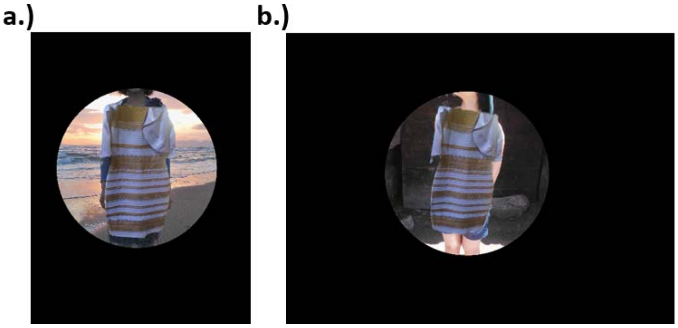

Figure 11. Estimation of local contrast. Note that the cutout in panel a is lighter ($L^* = 55$) and slightly less bluish ($u^* = -3.4$) than the one in panel b ($L^* = 33$, $u^* = -4.5$). Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

Method

Participants

Twenty observers (16 women, 39.3 ± 16 years) took part in the questionnaire. Nine observers took part in the white-gold inducing version, which involved the image of Figure 2; 11 observers participated in the blue-black inducing version with the image of Figure 10.

Apparatus and stimuli

The stimuli were the original image in Figure 1 and the disambiguated images in Figure 2 and Figure 10. The disambiguated images were made so that the appearance of the dress could not be explained by local color or lightness contrast. In both images, the dress is embedded in a dark local surround. In fact, in the blue-black inducing image of Figure 10, the local surround is actually darker than the rest of the image because of the shadow in the background. In the white-gold inducing image of Figure 2, the local surround is lighter because of the sunset behind the dress.

To give a rough estimation of local contrast, CIELUV lightness (L^*) and bluishness (v^*) were estimated for the circular cutouts in Figure 11. The cutout of the dress in the shadow in Figure 11a was lighter ($L^* = 55$) and slightly less bluish ($u^* = -3.4$) than the cutout of the dress in the sun in Figure 11b ($L^* = 33$, $u^* = -4.5$). Hence, the same cutout of the dress was contrasted to a lighter local background in Figure 11a than in Figure 11b. This implies that local contrast alone should produce the impression of darker blue-black colors for the dress in Figure 2 and lighter white-gold colors in Figure 10.

In contrast, the interpretation of the illumination in the images suggests the inverse. The interpretation of

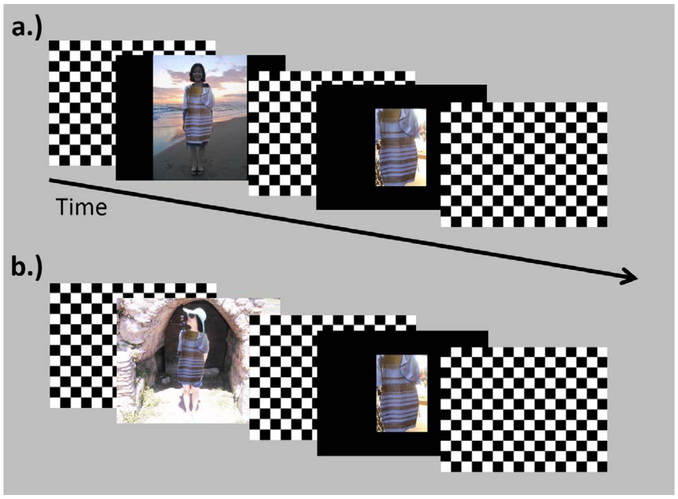

Figure 12. Stimulus displays and time course in the second study. Panel a and b show the different conditions (white-gold vs. blue-black inducing). Photograph of the dress used with permission. Copyright Cecilia Bleasdale.

the illumination is determined by the interpretation of the spatial relationships in the scenes. Because the light of the sunset is behind the dress, the dress is in the shadow in Figure 2. Hence, the dress itself should appear lighter (i.e., white-gold) because the darkness must be due to the illumination conditions. In Figure 10, the legs and the head of the woman are in the bright sunlight, which suggests that the woman is standing in the sun in front of the cave and not in the shadow in the cave. Because of this important clue about the spatial location of this woman, the image in Figure 10 suggests that the dress is comparatively dark, namely, blue-black, because the lightness must be attributed to the direct sunlight.

Procedure

Stimulus presentation was simply implemented with PowerPoint slides, and responses were collected through a written questionnaire. Figure 12 illustrates the stimulus presentation and time course of the measurement.

Observers were shown one or the other of the two disambiguated images in Figure 2 and Figure 10 and described the colors of the dress in the respective photo through unconstrained color naming. After a checkerboard mask, the original photo was shown, and observers described the colors of the dress in that photo through color naming. Subsequently, participants were also asked to describe the illumination in the original dress image (Figure 1) and to answer the questions about the illumination from the preliminary survey shown in Figure 9. At the end, we also asked whether the observer had seen the original photo before the experiment and whether the perceived colors of the dress ever changed in perception (or "switched") between white-gold and blue-black. The form to enter answers to the questions is provided in Supplementary Figure S11.

Results and discussion

Figure 13 illustrates the main results. As in the first experiment, dress scores were calculated based on the dress color-naming data. For *t* tests, we report one-tailed statistics because the direction of effects are defined, and we provide the corresponding correlation coefficient to appreciate effect sizes.

Disambiguation

First, the left bars in Figure 13e–f illustrate how observers saw the colors of the dress in the disambiguating images. The bars show the average dress scores for the dress in the white-gold disambiguating image (Figure 2) and in the blue-black disambiguating image (Figure 10), respectively. To evaluate the disambiguating effect of these images, we compared the dress scores between participants who saw one and the other image through independent *t* tests.

The type of disambiguation (white-gold in Figure 2 vs. blue-black in Figure 10) strongly influenced how observers saw the colors of the dress in the disambiguated images, $r = 0.76$, $t(18) = 5.0$, $p < 0.0001$. These results show that the disambiguating background of the images influences how observers see the same cutout of the dress. This result cannot be explained by low-level sensory effects of color contrast, as in Lafer-Sousa et al. (2015). Instead, it suggests that the perceived colors of the dress can be determined by the interpretation of the illumination conditions, mainly due to spatial cues.

Main results: Induction effects

Most important, in case of induction effects, the perception of the original photo should depend on the type of disambiguated photo seen before. This also implies that the perceived colors of the dress in Figure 1 should be positively correlated with the colors perceived in the disambiguated photo seen before. To test this, we calculated correlations between the dress scores for the disambiguating and the original image. Moreover, using independent *t* tests we compared the dress scores for the original image between the two groups of participants who saw different disambiguating images.

The perception of the colors in the disambiguated images was strongly correlated with the perceived dress colors in the original image of Figure 1, $r(18) = 0.83$, $p < 0.0001$. Thus, the colors seen in the original photo of

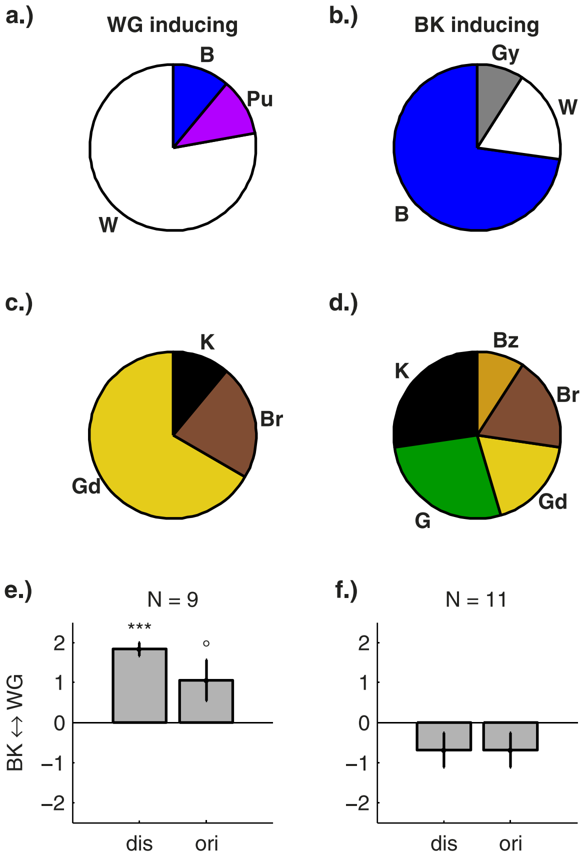

Figure 13. Induction effects. (a–d) How observers name the colors of the dress in the original photo (Figure 1) after seeing either the white-gold inducing (a, c) or the blue-black inducing (b, d) images of Figure 2 and Figure 10, respectively. The first row (a, b) illustrates the color naming of the dress body and the second row (c, d) the naming of the lace. (e, f) Dress scores in the white-gold (e) and blue-black (f) inducing condition. The first bars ("dis") show the dress scores for naming the colors in the disambiguating images (Figure 2 and Figure 10). The second bars correspond to the naming data in panels a–d and indicate how observers name the colors in the original photo of the dress. Error bars represent standard errors of mean, and symbols above the bars indicate significance in paired *t*-tests across participants (°$p < 0.1$; ***$p < 0.001$). Note the essential result that the disambiguated images did influence not only how observers see the colors of the dress in these images (first bars in panels e and f) but also how they see the colors subsequently in the original image (second bars in panels e and f).

the dress (Figure 1) depended on which of the two disambiguated images (Figure 2 or Figure 10) observers saw before, $r(18) = 0.52$, $t(18) = 2.6$, $p < 0.01$.

Figure 14 (fifth bar) shows the frequency of observers who had not seen the original dress image (Figure 1) before or who could not remember it. Prior experience with the original image could have influenced the effect of the disambiguated image on the perceived colors in the original image. However, the correlation between the dress scores and between the dress scores and the group of observers was similar when controlling for prior experience (seen before or not) in a partial correlation, $r(18) = 0.78$, $p < 0.0001$, and $r(18) = 0.56$, $p < 0.01$.

Taken together, these findings suggest that the disambiguated images influence the observers' assumptions about the illumination in the ambiguous, original photo of the dress, which determines their perception of the dress colors. However, this study involved comparatively few participants (only nine and 11 per group). It might be that the group with the white-gold inducing image happens to include more white-gold seers and the other group more blue-black seers by pure chance. Moreover, there is a difference between influencing observers who never saw the image of the photo before and changing the perception of observers who have already seen the photo and have strong assumptions about how to interpret it. To show changes in perception, we would need to better control whether and how observers saw the original image before engaging in our questionnaire. Finally, the question arises as to whether the possibility of influencing the perceived colors of the dress is related to the observers' ability to spontaneously or intentionally "switch" their perception between blue-black and white-gold. For now, our findings provide the first evidence that the perception of the dress can be influenced at all. We are currently in the process of conducting an online study with a large number of observers that aims to further clarify this issue.

Additional findings

In this second experiment, we reproduced the finding from the first experiment. The description of the dress colors in the original image was negatively correlated with the description of the illumination in that image; however, this correlation was not high, $r(17) = -0.40$, $p < 0.05$. Observers generally struggled with the question about the color of the illumination, and one participant failed to answer the question about the color of the illumination. These reactions confirm again how difficult it is for observers to explicitly describe their assumptions about the illumination of depicted scenes.

Moreover, we examined which assumptions about the scene were related to the perceived dress colors. For this purpose, we compared the dress scores for participants who answered "yes" and "no" in the respective questions through independent t tests.

The first four bars of Figure 14 show the frequencies of answers to the qualitative questions about the illumination. Observers who saw the white-gold inducing image (Figure 2) before seeing the original image (Figure 1) responded more often that the dress in the original photo was in the shadow than observers

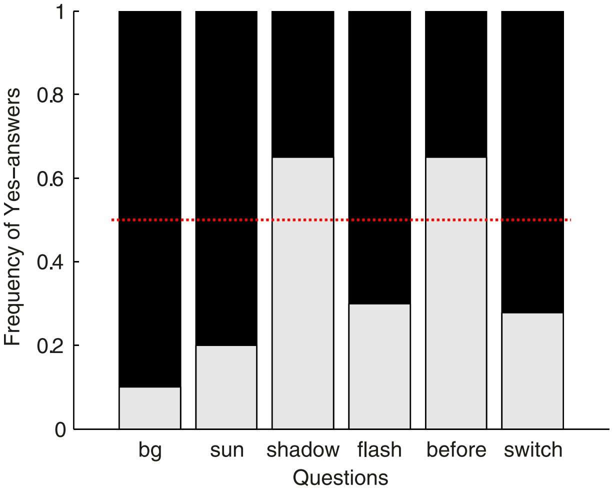

Figure 14. Interpretation of and prior experience with the original photo of the dress. Format as in Figure 9. The different questions along the x-axis are bg = "Is the dress illuminated by the same light as the background?"; sun = "Do you see the dress in sunlight?"; shadow = "Do you see the dress in a shadow?"; flash = "Is the dress illuminated by a flash (of the camera)?"; before = "Have you ever seen the (original) photo of the dress before?"; switch = "Did the appearance of the colors in the (original) photo ever switch"?

who saw the blue-black inducing image (Figure 10) prior to the original photo, $r = 0.45$, $t(18) = 2.2$, $p = 0.02$. In turn, observers saw the dress in the original photo more often as being in the sun if they were presented the blue-black inducing image (Figure 10) than if they were shown the white-gold inducing image, $r = 0.45$, $t(18) = -2.2$, $p = 0.03$.

Finally, the four questions of Figure 14 correspond to those of Figure 9. The disambiguated images in both Figure 2 and Figure 10 show the dress in a different light than the direct background, and they show it without a camera flash. A comparison between the results in Figure 14 and those in Figure 9 suggests that fewer participants saw the dress in the same color as the background and in the light of a camera flash after seeing the disambiguated images (10% and 30%; Figure 14) than without seeing the disambiguated images prior to the original image (33% and 66%; Figure 9).

These findings are completely in line with the idea that the white-gold inducing image (Figure 2) with the dress in the shadow led observers to interpret the dress in the original photo (Figure 1) as being in the shadow, whereas the blue-black inducing image (Figure 10) with the dress in the sun made observers think the dress in the original photo is in the sun, too. However, the results supporting this idea explained only small amounts of the variance (19% for the question about the shadow and 20% for the question about the sun). This means that they indicate only statistical tendencies rather than a deterministic relationship. Moreover, they did not allow for answering the question of whether the interpretation of the dress being in the shadow or in the sun is directly related to the perception of the dress.

Nevertheless, they provide some evidence that we can influence how observers interpret the ambiguous photo of the dress. To further clarify the impact of implicit assumptions on color perception, it would be good to provide further evidence that the perception of dress colors can be influenced, or even changed, by shaping how observers interpret the photo.

Conclusion

Taken together, the results of the present study explain why different observers see the dress in the photo in fundamentally different colors. The perceived colors of the dress are due to (implicit) assumptions about the illumination. This is exactly what would be predicted from classical color science, and no additional mechanisms need to be invoked to account for the surprising diversity in judgments of the dress's color (Chetverikov & Ivanchei, 2016; Witzel, 2015; see also Toscani, Gegenfurtner, & Doerschner, 2017, this collection).

Moreover, our results provide some evidence that prior experience with disambiguated images may push observers to interpret the ambiguous original photo in one or the other way. These observations suggest that the perception of the dress colors may be modulated by biasing observers toward one or the other interpretation of the photo. This idea indicates that the perception of the dress colors is not mainly determined by general factors or mechanisms that are independent of this particular photo, such as differences in perceptual mechanisms. The lack of correlation between estimations of generic subjective gray points and perceived dress colors further undermines the idea that general priors about the illumination or a "blue bias" in the perception of gray could account for the individual differences in the perception of the dress (Gegenfurtner et al., 2015; Lafer-Sousa et al., 2015; Winkler et al., 2015; Wuerger et al., 2015).

However, it remains an open question whether there are other factors that systematically influence which interpretation of the photo is preferred by different observers. This would explain why there are systematic differences between different groups of observers, as observed previously (Lafer-Sousa et al., 2015).

After the discovery of the dress more than 1 year ago, only one other photo, of a jacket, has been found that seems to produce comparable individual differences in color perception (poppunkblogger, 2016).

Here, we proposed that the photo of the dress is a special case because it displays the particular color distributions while still appearing sufficiently realistic. If this is true, the same features should be found in the jacket. However, the pictures of the dress and the jacket were accidentally taken photos that were discovered retrospectively to produce those individual differences in color perception. A complete understanding would allow for intentionally creating or modeling an artificial image that produces those individual differences in color perception. Our findings suggest that ambiguity about the illumination and the configuration of the scene must be the core features that need to be modeled in such an image.

More generally, the present findings show that color perception depends on the observers' assumptions and beliefs about the scene. The striking individual difference in the perceived colors of the dress cannot be due to bottom-up properties of the image because they occur when observers look at the same image, even when controlled under experimental conditions. Instead, our findings show that the differences in the interpretation of the scene in the photo produce those fundamental differences in perception and that prior experiences may influence this perception. This implies that the unconscious assumptions and beliefs about the reality represented in the dress photo influence in a top-down fashion which colors observers perceive.

In this way, the phenomenon of the dress nicely illustrates the power of unconscious assumptions and beliefs in perception in general. For this reason, it might be considered as an example for the cognitive penetration of perception, which refers to the idea that cognitive aspects, such as thoughts, knowledge, and beliefs, shape perception (Collins & Olson, 2014; Firestone & Scholl, 2015; Witzel & Hansen, 2015). Moreover, the example of the dress shows that the individual differences in perception that may arise from differences in assumptions and interpretations can lead to shocking disagreements across individual observers.

Keywords: color appearance, color constancy, individual differences, scene interpretation, surface color, color switching dress

Acknowledgments

Thanks to Ying Chen for posing for the photos in Figure 2 and Figure 10, Carlijn van Alphen for help with data collection, and Sophie Wuerger, Anya Hurlbert, David Peterzell, Jules Davidoff, Karl Gegenfurtner, and David Brainard for comments. The work was supported by ERC Advanced Grant "FEEL" No. 323674 to J. Kevin O'Regan. Chris Racey is funded by a European Research Council project "CATEGORIES," ref. 280635.

Commercial relationships: none.

Corresponding author: Christoph Witzel.
Email: cwitzel@daad-alumni.de.
Address: Justus-Liebig-Universität Gießen, Gießen, Germany.

References

Brainard, D. H., & Hurlbert, A. C. (2015). Colour vision: Understanding #TheDress. *Current Biology*, *25*, R551–R554, doi:10.1016/j.cub.2015.05.020.

Chauhan, T., Perales, E., Xiao, K., Hird, E., Karatzas, D., & Wuerger, S. (2014). The achromatic locus: Effect of navigation direction in color space. *Journal of Vision*, *14*(1):25, 1–11, doi:10.1167/14.1.25. [PubMed] [Article]

Chetverikov, A., & Ivanchei, I. (2016). Seeing "the Dress" in the right light: Perceived colors and inferred light sources. *Perception*, *45*(8), 910–930, doi:10.1177/0301006616643664.

Collins, J. A., & Olson, I. R. (2014). Knowledge is power: How conceptual knowledge transforms visual cognition. *Psychonomic Bulletin & Review*, *21*, 843–860, doi:10.3758/s13423-013-0564-3.

Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behavioural and Brain Sciences*, 1–77, doi:10.1017/S0140525X15000965.

Foster, D. H. (2011). Color constancy. *Vision Research*, *51*, 674–700, doi:10.1016/j.visres.2010.09.006.

Gegenfurtner, K. R., Bloj, M., & Toscani, M. (2015). The many colours of 'the dress.' *Current Biology*, *25*, R543–R544, doi:10.1016/j.cub.2015.04.043.

Granzier, J. J., Brenner, E., & Smeets, J. B. (2009). Can illumination estimates provide the basis for color constancy? *Journal of Vision*, *9*(3):18, 1–11, doi:10.1167/9.3.18. [PubMed] [Article]

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*, 1367–1368, doi:nn1794 [pii], 10.1038/nn1794.

Hesslinger, V. M., & Carbon, C.-C. (2016). #TheDress: The role of illumination information and individual differences in the psychophysics of perceiving white–blue ambiguities. *i-Perception*, *7*, 2041669516645592, doi:10.1177/2041669516645592.

Hurlbert, A. C. (2007). Colour constancy. *Current*

Biology, *17*, R906–R907, doi:S0960-9822(07)01839-8 [pii] 10.1016/j.cub.2007.08.022.

Ishihara, S. (2004). *Ishihara's tests for colour deficiency*. Tokyo, Japan: Kanehara Trading Inc.

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, doi:10.1016/j.cub.2015.04.053.

Lotto, R. B. (2008). Demo 12. Retrieved October 28, 2015, from http://www.lottolab.org/illusiondemos/Demo%2012.html

Lotto, R. B., & Purves, D. (2002). The empirical basis of color perception. *Consciousness and Cognition*, *11*, 609–629, doi:S1053810002000144 [pii].

M6. (2015). Trompe l'œil, camouflage, illusions d'optique: quand la réalité trompe nos yeux [TV broadcast]. In M6 (Producer), *E=M6*.

Macknik, S. L., Martinez-Conde, S., & Conway, B. R. (2015). Unraveling "The Dress." *Scientific American Mind*, *26*, 19–21, doi: doi:10.1038/scientificamericanmind0715-19.

Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D., & Hurlbert, A. (2014). Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. *PLoS One*, *9*, e87989, doi:10.1371/journal.pone.0087989.

Pernet, C. R., Wilcox, R., & Rousselet, G. A. (2012). Robust correlation analyses: False positive and power validation using a new open source matlab toolbox. *Frontiers in Psychology*, *3*, 606, doi:10.3389/fpsyg.2012.00606.

poppunkblogger. (2016). One year after #TheDress. Retrieved from http://wondirwin.tumblr.com/post/140009149046/poppunkblogger-dammitmichael

Purves, D. (2015). Color contrast: Cube. Retrieved October 28, 2015, from http://purveslab.net/see-for-yourself/

Swiked. (2015). guys-please-help-me-is-this-dress-white-and. Retrieved May 7, 2015, from http://swiked.tumblr.com/post/112073818575/guys-please-help-me-is-this-dress-white-and

Toscani, M., Gegenfurtner, K. R., & Doerschner, K. (2017). Differences in illumination estimation in #thedress. *Journal of Vision*, *17*(1):40, 1–14, doi:10.1167/17.1.40.

von Helmholtz, H. (1924 [1867]). *Treatise on physiological optic*. Vol. 2. J. P. C. Southall, Ed. Menasha, WI: The Optical Society of America.

Winkler, A. D., Spillmann, L., Werner, J. S., & Webster, M. A. (2015). Asymmetries in blue-yellow color perception and in the color of 'the dress.' *Current Biology*, *25*, R547–R548, doi:10.1016/j.cub.2015.05.004.

Witzel, C. (2015). The Dress: Why do different observers see extremely different colours in the same photo? Retrieved August 2015, from http://lpp.psycho.univ-paris5.fr/feel/?page_id=929

Witzel, C., & Hansen, T. (2015). Memory effects on colour perception. In A. J. Elliot, M. D. Fairchild, & A. Franklin (Eds.), *Handbook of color psychology* (pp. 641–665). Cambridge, UK: Cambridge University Press.

Witzel, C., Hurlbert, A. C., & Wuerger, S. M. (2016, May). *Variation of subjective white-points along the daylight axis and the colour of the dress*. Paper presented at the Annual Meeting of the Visual Sciences Society, St. Pete Beach, FL.

Witzel, C., & O'Regan, J. K. (2015, August). *The most reasonable explanation of the dress: Implicit assumptions about illumination*. Paper presented at ECVP 2015, Liverpool, UK.

Witzel, C., Racey, C., & O'Regan, J. K. (2016, May). *Perceived colors of the color-switching dress depend on implicit assumptions about the illumination*. Paper presented at the Annual Meeting of the Visual Sciences Society, St. Pete Beach, FL.

Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner, K. R. (2011). Object knowledge modulates colour appearance. *i-Perception*, *2*, 13–49, doi:10.1068/i0396.

Wuerger, S. M., Hurlbert, A. C., & Witzel, C. (2015, August). *Variation of subjective white-points along the daylight axis and the colour of the dress*. Paper presented at the ECVP 2015, Liverpool, UK.

The most reasonable explanation of “the dress”: implicit assumptions about illumination.

Christoph Witzel, Chris Racey, and J. Kevin O’Regan

Journal of vision, 2017, Special Issue: A Dress Rehearsal for Vision Science.

Supplementary material

Table of Content

Illumination and dress colors

Method

Stimuli at grey point

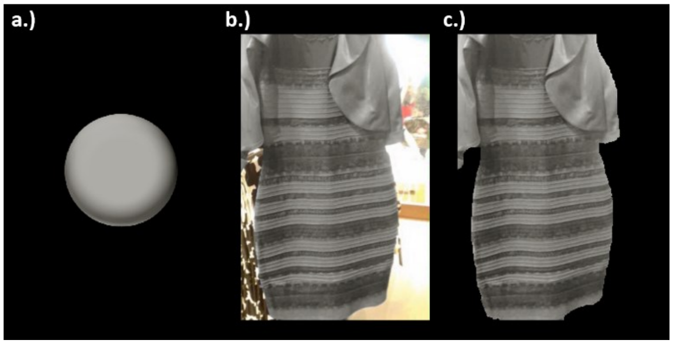

Figure S1. Illustration of completely grey stimuli. This is how the images looked in the grey adjustments, when they were set exactly to the chromaticity of the adapting white-point. *Note that the original version of the dress and the dress with uniform hue result in the same image at the adapting white-point (panel b).*

Dress with uniform hue

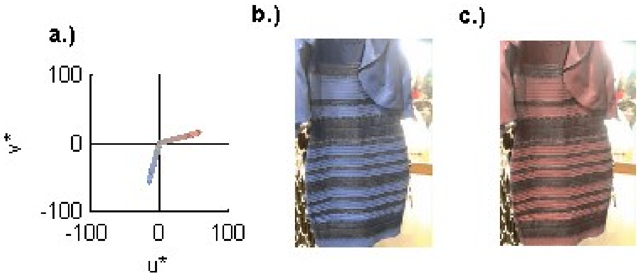

Figure S2. Dress with uniform hue. Panel a illustrates the color distribution in CIELUV space after compressing the hue of the color distribution (bluish hue direction) and after rotating this distribution by 120 degree (reddish hue direction). Panel b and c illustrate the stimulus images that correspond to the distributions in panel a. This version of the dress has been used as one of the stimuli in the grey adjustment task. *Note the difference between the color distribution and the image in this figure and those shown in Figure 1 of the main article.*

Questions about illumination

We asked observers the following 4 questions:

(1) Is the dress illuminated by the same light as the background?
☐ Yes ☐ No
(2) Do you see the dress in sunlight?
☐ Yes ☐ No
(3) Do you see the dress in a shadow?
☐ Yes ☐ No
(4) Is the dress illuminated by a flash (of the camera)?
☐ Yes ☐ No

Main Results

Error bars for PC scores of illumination adjustments

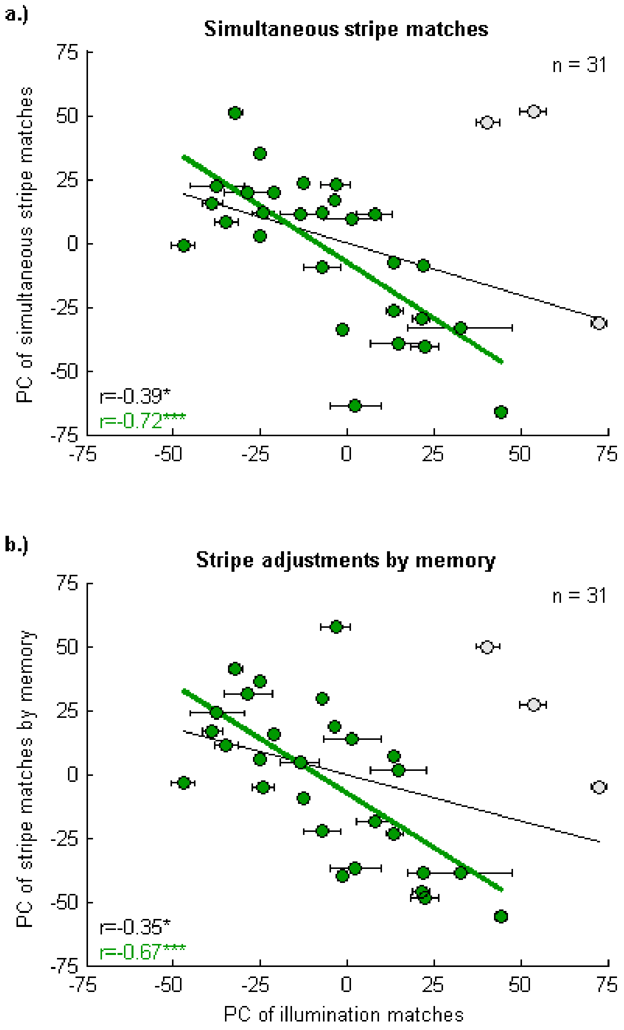

Figure S3. Correlations between assumed illumination and stripe adjustments with error bars. Horizontal error bars indicate standard errors of mean across the two illumination adjustments. Apart from that, format as in Figure 7 of the main article.

Robust correlations

Robust correlation analyses showed that the correlation between perceived stripe colors and assumed illumination are robust and reliable. As reported in the main article correlations were much higher when excluding outliers in a skipped correlation. This is true for both, adjustments by simultaneous matching and by memory (**Tables S1**).

Table S1. Robust correlations between stripe matches and illumination matches. * p < 0.05; ** p < 0.01; *** p < 0.001.

Type of correlation	r	p (left-tailed)	Bootstr. confidence interval	Sig.
Simultaneous stripe matches				
Pearson	-0.39	0.02	[-0.73, -0.07]	*
Bend	-0.48	0.003	[-0.77, -0.11]	**
Spearman	-0.44	0.007	[-0.72, -0.11]	**
Pearson Skipped	-0.72	< 0.001	[-0.85 -0.57]	***
Spearman Skipped	-0.68	< 0.001	[-0.81, -0.46]	***
Stripe matches by memory				
Pearson	-0.35	0.03	[-0.63 -0.07]	*
Bend	-0.44	0.01	[-0.71, -0.07]	**
Spearman	-0.42	0.01	[-0.68, -0.4]	**
Pearson Skipped	-0.67	< 0.001	[-0.82, -0.48]	***
Spearman Skipped	-0.67	< 0.001	[-0.82, -0.43]	***

Subjective Grey-Point

There were only correlations between grey adjustments and perceived dress colors when grey adjustments were done with the dress in the context of the photo (**Figure S5.b-c**).

There were no correlations between perceived dress colors and grey adjustments of a disk (**Figure S4.b-c**) or of the dress without background (**Figure S6.b-c**).

Grey adjustments of disk (Generic subjective grey-point)

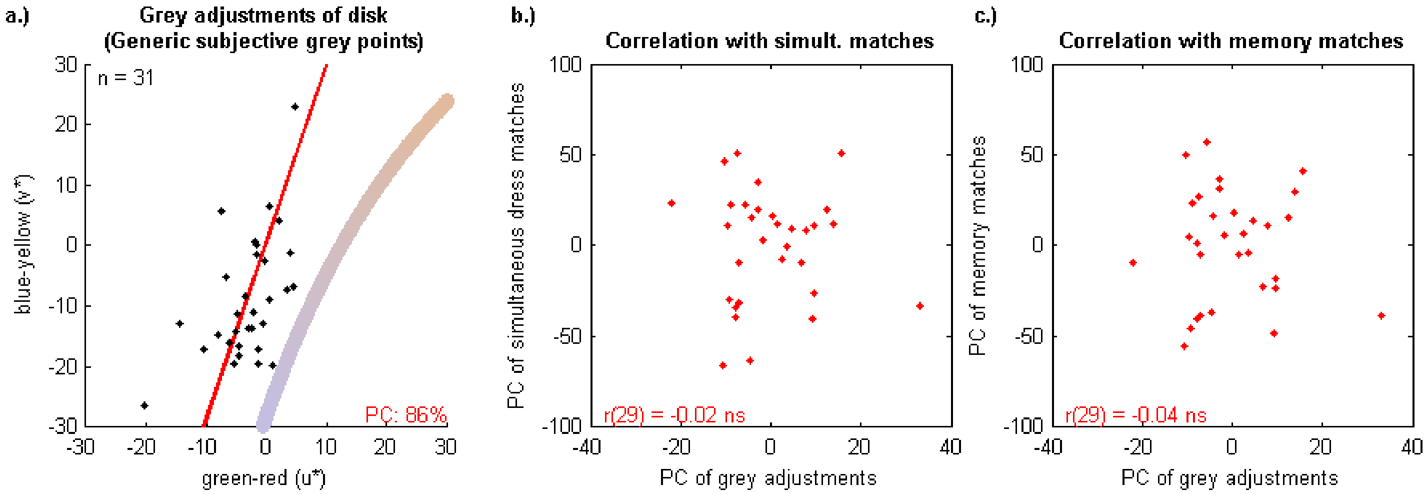

Figure S4. Grey adjustments of disk (*Generic subjective grey point*). Panel a shows the average grey adjustments (dots). For this figure, panel a is the same as Figure 6.d of the main article. Panel b and c illustrate potential correlations between the grey adjustments and the simultaneous (panel b) and memory matches (panel c) of the dress stripes. The axes show the scores along the respective principal components. The scores are plotted in red because the scores along the x-axis correspond to the projections onto the red line in panel a. Correlations are reported in the lower left corner (ns = not significant). *Note that there is no correlation for the grey adjustments of the disk (panels b & c)*.

Grey adjustments of dress with and uniform hue (and background)

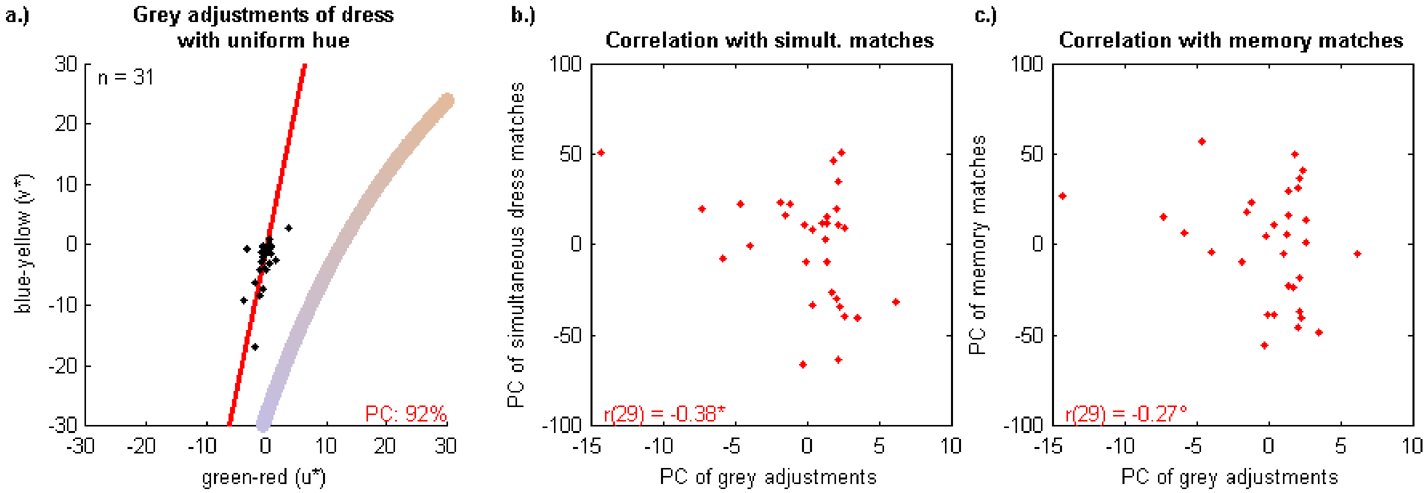

Figure S5. Grey adjustments of dress with uniform hue. Format is as in Figure S4. *There is a correlation between these grey adjustments and the dress stripes, reflecting the impact of the assumed illumination on the perceived colour of the dress.*

Grey adjustments of dress in the original photo

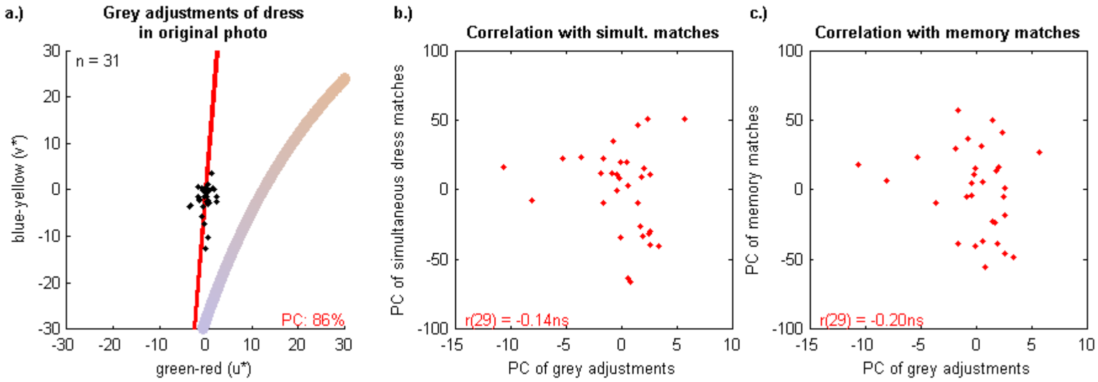

Figure S6. Grey adjustments of dress in the original photo. Format is as in Figure S4. *There is no significant negative correlation between these grey adjustments and the dress stripes.*

Grey adjustments of dress without background

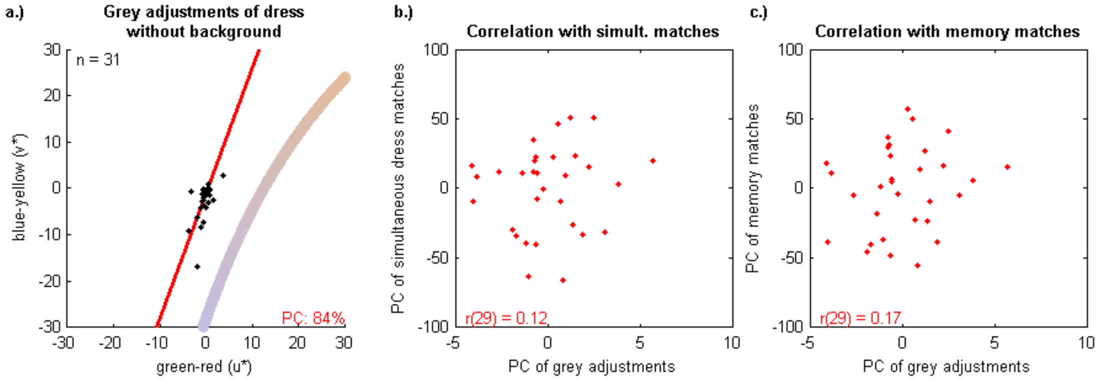

Figure S7. Grey adjustments of dress without background. Format is as in Figure S4. *There is no negative correlation (correlation coefficients are positive). At the same time, these grey adjustments (x-axes in panels b and c) varied much less than those for the disk (Figure S4) and the dress with uniform hue (Figure S5). We speculate that this might be due to the bimodal color distribution (bluish and brownish clusters) of the dress.*

Robust correlation analyses

Correlations between grey adjustments and simultaneous matching of stripe colours were also robust. However, skipped correlations were not as high as the ones that involved illumination adjustments. Moreover, there was no significant correlation between memory adjustments and grey adjustments (**Table S2**).

Type of correlation	R	p (left-tailed)	Bootstr. confidence interval	Sig.
Simultaneous stripe matches				
Pearson	-0.38	0.02	[-0.58, -0.11]	*
Bend	-0.35	0.03	[-0.61, -0.03]	*
Spearman	-0.35	0.03	[-0.62, -0.04]	*
Pearson Skipped	-0.38	0.02	[-0.58, -0.10]	*
Spearman Skipped	-0.35	0.03	[-0.60, -0.07]	*
Stripe matches by memory				
Pearson	-0.27	0.06	[-0.44, -0.05]	°
Bend	-0.21	0.13	[-0.45, 0.08]	ns
Spearman	-0.20	0.14	[-0.46, 0.08]	ns
Pearson Skipped	-0.11	0.28	[-0.31, 0.11]	ns
Spearman Skipped	-0.08	0.33	[-0.36, 0.21]	ns

Table S2. Robust correlations between stripe matches and grey adjustments of dress with uniform hue. ns = not significant ° $p < 0.1$; * $p < 0.05$. →

Naming

Naming data confirmed the correlations found with the adjustment data. First of all, illumination matches as well as grey adjustments of the dress (with uniform hue) also predicted whether observers named the stripes blue-black, or white-gold ($r(29) = -0.34$, $p = 0.02$, and $r(29) = -0.45$, $p < 0.01$; cf. **Figure S8**). Inversely, the illumination naming predicted the principal component scores of the adjustments in the simultaneous matching and in the memory task ($r(29) = 0.44$, $p < 0.01$, and $r(29) = 0.43$, $p < 0.01$; cf. **Figure S9**). Finally, the naming of the stripes could also be predicted by whether observers called the illumination by a color term that is rather light and yellow than dark and blue ($r(29) = -0.43$, $p < 0.01$; **Figure S10**).

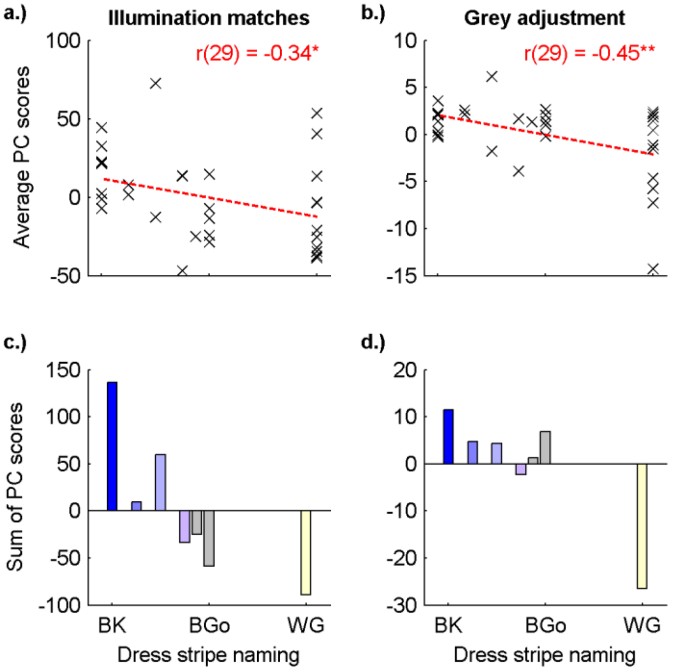

Figure S8. Correlation between dress stripe naming and illumination (left) and grey adjustments (right). Color name combinations are shown along the x-axis: BK = Blue-Black, BGo = Blue-Gold, and WG = White-Gold. The lower row is for illustration purposes only. It shows the sum (instead of the average) of the adjustment scores in order to combine the size of shift of adjustments to one or the other extreme of the principal component with the frequency of participants that made such a shift. Note however, that calculations of tests where based on the diagrams in the upper row. *These correlations replicate the negative correlations found for dress stripe matches.*

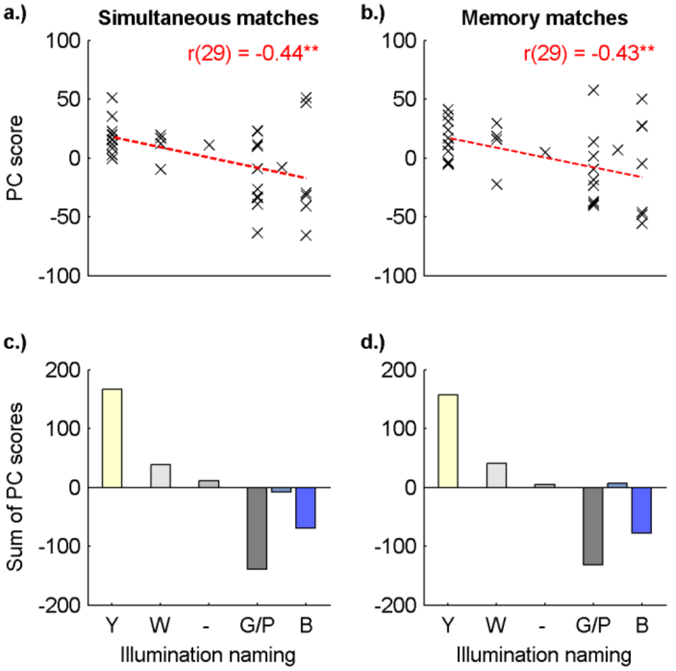

Figure S9. Correlation between illumination naming and stripe adjustments by matching (left) and by memory (right). The x-axis shows color descriptions of the illumination: Y = Yellow, W = White, – = ambiguous, G/P = Grey or Purple, and B = Blue. Apart from that, format is as in Figure S8. *Note again the negative correlations.*

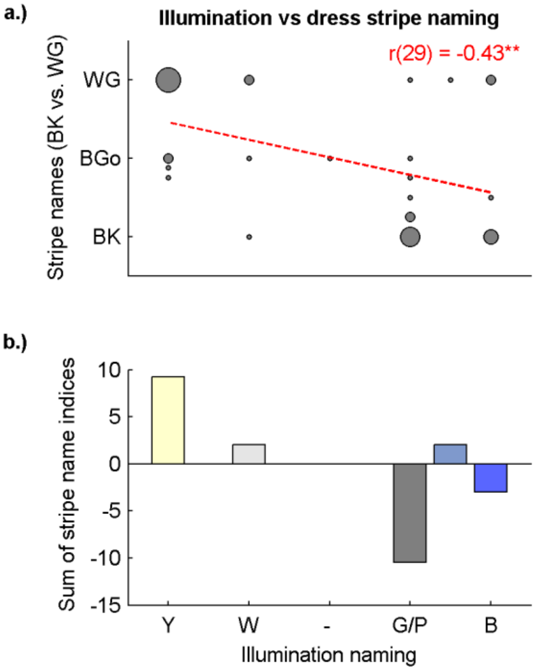

Figure S10. Correlations between dress stripe and illumination naming. Format is as in Figures S8-S9. The size of the disks in panel a indicates the frequency of observers with this combination of naming the dress and naming the illumination.

Biasing perception

Participant ID: __________ Gender: _________ Age: _________

(1) What colours are the stripes of the dress in photo A | B : ________________ and ________________

(2) What colours are the stripes in the original photo: ________________ and ________________

(3) What is the colour of the illumination that reaches the dress: ________________

(3.1) Is the dress illuminated by the same light as the background? ☐ Yes ☐ No

(3.2) Do you see the dress in sunlight? ☐ Yes ☐ No

(3.3) Do you see the dress in a shadow? ☐ Yes ☐ No

(3.4) Is the dress illuminated by a flash (of the camera)? ☐ Yes ☐ No

(4) Have you ever seen the (original) photo of the dress before? ☐ Yes ☐ No

(5) Did the appearance of the colors in the (original) photo ever switch? ☐ Yes ☐ No

Figure S11. Questionnaire form of the second study. Answers were collected and added to the questionnaire by the experimenter. In question 1, the experimenter selected which photo the observer saw (A or B).