SUPPLEMENTAL INFORMATION

Striking individual differences in color perception uncovered by 'the dress' photograph

Rosa Lafer-Sousa, Katherine L. Hermann, Bevil R. Conway
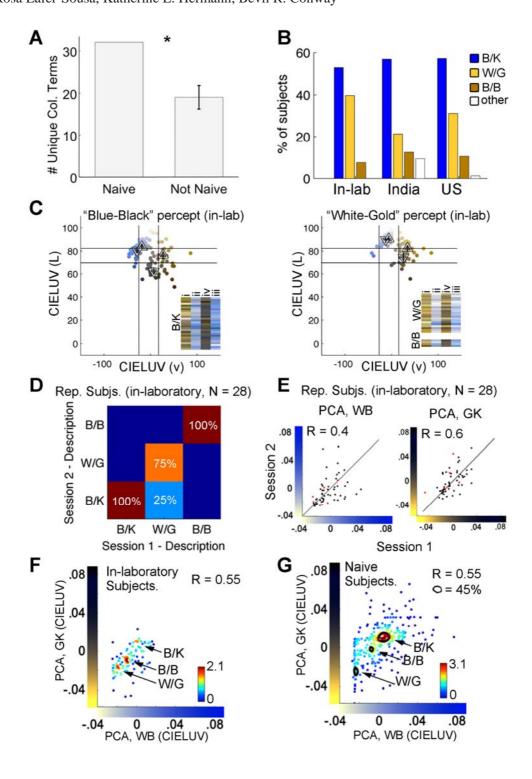
Figure S1. Analysis of the subset of subjects who performed the task in the laboratory under controlled viewing conditions (N=53), and naïve subjects tested on-line who reported having not seen the image before (N=311), also show striking individual differences in color perception of #TheDress that can be recovered both from the words used to describe the image and from the results of color matching. **(A)** Bar plot showing the number of unique color terms used by naïve (N=313) and non-naïve (N=1088 subjects). To control for the number of subjects, the bar for the non-naïve subjects shows the mean of 10,000 subsampling boot-straps of the non-naïve population. Error bar shows the 95% C.I. of the mean. The naïve subject group used more unique color terms ($p < 0.0001$). **(B)** Histogram showing the color descriptions of the dress for the in-laboratory subjects (N = 53; 51 had seen the image previously), the subset of subjects from India who were tested online (N = 160; 63 had seen the image previously), and the subset of subjects from the United States who were tested online (N = 1147; 945 had seen the image previously). **(C)** Color matches made by control subjects reporting B/K or W/G, for regions in the dress (see i, ii, iii, iv, main Figure 1, panel A). White triangles show population averages (upward triangles for regions i and ii, and downward triangles for regions iii and iv, and contain the 95% C.I. of the mean. Insets depict color matches for individual subjects in each row, sorted by category (B/K, W/G, B/B). Subjects used a chin rest to control viewing angle and distance, and a calibrated display. **(D)** A subset of in-laboratory subjects (N = 28) participated in a test-re-test reliability assessment 3-5 days after participating in the main experiment. Heat map values indicate color descriptions on re-test as a function of how subjects originally reported the dress. Results are predominantly along the x=y diagonal, showing strong test-retest reliability. **(E)** In-laboratory-re-test subjects' color matches for regions (left shows white/blue regions (ii, iii); right shows gold/black regions (i, iv) from experiment 1 ("Session 1") plotted against matches for those same regions on re-test ("Session 2"). Note that the color terms "blue" and "brown" refer to the colors that most people assign to the pixels when the pixels are presented in isolation on a neutral background. See Supplementary Experimental Procedures for the definition of the axes. All data (from the entire study) in which subjects were asked to match a blue region in the image (ii, iii) contributed to the PCA used to define the axes in the left panel; and all data in which subjects were asked to match a brown region (i, iv) were used to define the axes in the right panel. The x=y diagonal is plotted for reference. If subjects show perfect reliability, the results should fall along the diagonal. Subjects whose responses changed in the color-description task are indicated in red. Re-test reliability was significant (regions ii, iii, R = 0.4, p = 0.002; ii, iv, R = 0.6, p = $3*10^{-6}$). **(F)** In-laboratory subject color matches for regions (i and iv) plotted against matches for regions (ii and iii). The x value for each subject was defined as the weight along the first principal component for their color match; the y-axis values were determined similarly, using matches for the brown pixels of the image (i, iv, Figure 1A). Axes in all density plots shown in the present report (Figure 1F, S1G) are identical (R = 0.55, p = $7*10^{-10}$). **(G)** Naïve subject color matches for 'brown' region (i) plotted against matches for neighboring 'blue' region (ii). Contours contain 45% of the respondents within each category (B/K, W/G, B/B), defined independently by the color descriptions provided by the subjects. Axes defined as for D. (R = 0.55, p = $6*10^{-26}$)
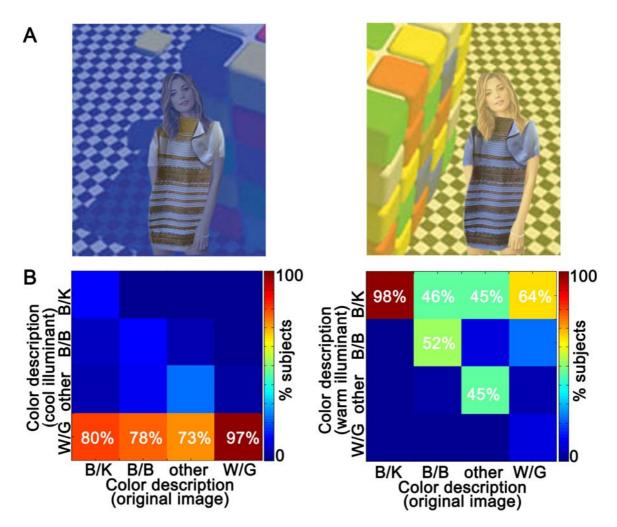
Figure S2. Embedding the image in a scene with overt cues to illumination causes most subjects to perceive the dress in a predictable way: W/G when seen under a cool illuminant; B/K when seen under a warm illuminant (N = 1127). Cues to the illuminant are provided by a global color tint applied to the whole scene, including the skin of the model. **(A)** The dress embedded within a scene simulating a cool illuminant (left), and the dress embedded within a scene simulating a warm illuminant (right) **(B)** Color descriptions under these conditions as a function of how subjects originally described the dress.

Table S1.
Results of logistic regression models described in **Supplemental Experimental Procedures**.
Odds ratio [95% C.I.]; p calculated by Wald chi-square test.

	Model 1	**Model 2**	**Model 3**
Description	white/gold (success) versus blue/black response	main category (white/gold or blue/black, success) versus non-main category response	same as Model 2, but with "country" removed as a predictor variable
Intercept	0.31 [0.17-0.56]; **p < 0.0001***	20.86 [9.81-45.14]; **p < 0.0001***	15.50 [7.60-32.12]; **p < 0.0001***
Condition: Main experiment, order B (reference: main experiment order A)	1.15 [0.88-1.51]; p = 0.32	1.04 [0.72-0.49]; p = 0.84	1.03 [0.72-1.47]; p = 0.89
Condition: In-laboratory control condition (reference: main experiment order A)	1.46 [0.78-2.70]; p = 0.23	1.09 [0.41-3.75]; p = 0.88	1.33 [0.52-4.56]; p = 0.60
Condition: Scale experiment, 36% (reference: main experiment order A)	0.74 [0.52-1.04]; p = 0.08	1.21 [0.77-1.93]; p = 0.41	1.19 [0.76-1.89]; p = 0.46
F.lux software installed	0.75 [0.50-1.10]; p = 0.14	1.10 [0.67-1.91]; p = 0.71	1.11 [0.68-1.91]; p = 0.68
Male	0.75 [0.58-0.95]; **p = 0.019***	0.92 [0.66-1.28]; p = 0.63	0.89 [0.64-1.24]; p = 0.48
India (reference: USA)	0.84 [0.54-1.28]; p = 0.42	0.48 [0.31-0.76]; **p = 0.0016***	
Other country (reference: USA)	0.82 [0.43-1.52]; p = 0.55	6.48 [1.38-115.75]; p = 0.07	
Had seen image prior to study	1.21 [0.88-1.68]; p = 0.24	1.27 [0.86-1.86]; p = 0.22	1.53 [1.06-2.17]; **p = 0.021***
Age	1.02 [1.01-1.03]; **p = 0.0035***	0.97 [0.95-0.98]; **p < 0.0001***	0.97 [0.95-0.98]; **p < 0.0001***

	Model 4
Description	white/gold (success) versus blue/black response; non-naïve participant group
Intercept	0.52 [0.30-0.91]; **p = 0.02***
Condition: Main experiment, order B (reference: main experiment order A)	1.09 [0.81-1.48]; p = 0.56

Condition: In-laboratory control condition (reference: main experiment order A)	1.54 [0.81-2.92]; p = 0.19
Condition: Scale experiment, 36% (reference: main experiment order A)	0.76 [0.51-1.12]; p = 0.17
F.lux software installed	0.81 [0.53-1.24]; p = 0.34
Male	0.74 [0.56-0.97]; **p = 0.03***
India (reference: USA)	0.83 [0.43-1.52]; p = 0.55
Other country (reference: USA)	0.85 [0.40-1.71]; p = 0.65
Had seen image switch colors prior to study	0.60 [0.46-0.79]; **p = 0.0003***
No response to switching question	0.97 [0.28-3.02]; p = 0.95
Age	1.01 [1.00-1.03]; **p = 0.032***

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Experimental Design
The majority of participants were recruited and tested through Amazon's Mechanical Turk using a combination of template (Morris Alper's Turk Suite Template Generator 2014, available online at *http://mturk.mit.edu/template.php*) and custom HTML and JavaScript code. A smaller number of subjects (N=53) were recruited from the MIT University and Wellesley College campus through word of mouth and social media, and also tested using the M-Turk platform on a calibrated display in the laboratory. We adhered to the MIT Committee on the Use of Humans as Experimental Subjects policies on using Amazon's Mechanical Turk for research purposes. Informed consent was obtained for those subjects who performing the study in the laboratory study. Procedures were approved by the institutional review board of Wellesley College.

We conducted three experiments: 1, a main experiment; 2, a follow-up experiment to assess the role of image size in determining what colors people report; 3, an in-laboratory, controlled experiment. Data were pooled from the various experiments depending on the analysis performed.

Subjects were between 18 and 69 years of age. To control for subject quality among the Mechanical Turk participants, we required that subjects have Mechanical Turk approval ratings of 95% or higher, and have completed at least 1000 human intelligence tasks (HITs) on Mechanical Turk previously.

Experiment 1: Main
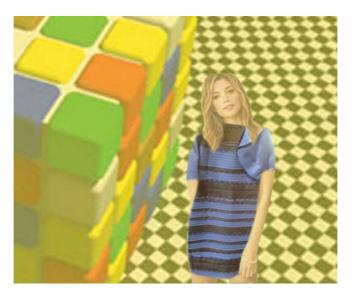
Of the M-Turk participants who enrolled in the study, we rejected 70 as repeat participants. Of the remaining 1175, we rejected 20 subjects who indicated they were or might be colorblind, and 11 subjects who provided non-color word responses to our free response questions about dress color, leaving 1074 participants. Sample was gender balanced: $N_{males} = 563$, $N_{females} = 510$, $N_{other} = 1$. We estimate subjects spent approximately twenty minutes completing the survey (subjects who performed the tasks in the laboratory took about the same amount of time).

We assigned subjects to one of two conditions (A and B). In both conditions, we presented three images and asked questions about the dress color and lighting in each. The first image shown was always the same, but the order of the second two images differed between the two conditions. In addition, all subjects were queried on their perception of a blurry version of the first image.

To ensure that stimuli were the same size across displays for on-line subjects, we specified the sizes of stimulus images in absolute pixels in the HTML experiment code. There is some variability from display to display in terms of the actual physical size of a pixel. We measured the images on a typical monitor in the laboratory, to provide a reasonable estimate for how the pixel values correspond to degrees of visual angle. These values are given in the legend of Figure 1, and below. We estimate that among different displays the variance in actual display size was ~+/-10% of the size measured in laboratory.

The stimulus images were:

(1) the original dress image that circulated on the internet ("original") was presented at 36% of its original size so that the entire image would be visible on all displays. We restricted the study to subjects who performed the tests on monitors large enough to display the entire 36% image; this precluded participants who might do the study using a hand-held device. The dress image was presented at an absolute size of 226 x 337 pixels. This corresponded to 7.2° of visual angle width on the 18.65" wide iMac (21.5" diagonal) used for the in-laboratory experiment, using a 40-cm viewing distance.

(2) A cut-out of the original dress superimposed on a picture of a model, embedded in a scene with a simulated *warm illuminant (*model was placed so as to appear in direct illumination); the dress portion of the stimulus was 76% of the size of the dress in the original image. The scene was provided, with permission, by Beau Lotto and Dale Purves. **Figure S2** shows a cropped version of what was shown. The complete picture was 518x429 pixels (~16.5° of visual angle on the horizontal axis) is shown below.

(3) A cut-out of the original dress also superimposed on a picture of a model, but embedded in a scene with a simulated *cool illuminant (*model was placed in the shadow cast by a scene element). The scene was provided, with permission, by Beau Lotto and Dale Purves. **Figure S2** shows a cropped version of what was shown. The complete picture was 518x429 pixels (approximately 16.5° of visual angle on the horizontal axis), shown below.

The skin and hair of the model in (2) and (3) were tinted to reinforce the perception that the model is within a particular scene, illuminated by the same light source that is illuminating the rest of the scene. The chromaticity of skin-tint was matched to the overall illuminant (estimated from the "white" patches in the original Lotto cubes), and the lightness of the tint was varied to indicate either shadow or direct illumination. The dress in all three images (1, 2, and 3) was physically identical.

(4) A blurry version of the original image was presented at 41% of the source size, with a Gaussian blur radius of 3.3 pixels (0.11°). The image was 8.3° of visual angle along the horizontal axis.

In condition A, subjects saw the cool-illuminated image (3) first and then the warm-illuminated image (2). In condition B, subjects saw image (2) first and then image (3). All subjects saw image (1) before seeing either (2) or (3). And all subjects saw (4) last.

Each image was shown for fifteen seconds, and then we asked the subjects to make color and lighting judgments. The task was programmed so as to require that the subjects were exposed to the image for at least 15 seconds before they could answer the questions. The image was on the screen continuously while the subjects answered the questions (this was not designed as a test of color memory). Between the presentation of the first and second images, we collected basic demographics information. Between the presentations of the last two images, we asked subjects about the environment in which they were completing the study. We also asked subjects whether they had viewed the original dress image prior to this study, and if so, whether they had experienced multiple percepts of the dress (i.e. "switching"). We inquired about the rate of switching of the dress colors in pre-study viewings of the image. The full survey is given in the Appendix below.

Experiment 2: Scale
To explore the effect of image size on percept, we performed a second experiment in which subjects viewed the original dress image at one of four sizes, defined as a % of the original: 10% (63 x 94 pixels, or 2.0° of visual angle on a 21.5-inch iMac), 36% (the image size we presented in our main experiment; 226 x 337 pixels, or 7.2° of visual angle), 100% (628 x 937 pixels, or 20.0° of visual angle; and 150% (942 x 1405 pixels, or 30.0° of visual angle). The stimuli in the 10% and 36% conditions fit fully in all browser windows. For the 100% and 150% images, only part of the image was visible at once in the height direction, but was completely visible in the horizontal dimension. As a result, subjects had to scroll over the stimulus image, from top to bottom, to view it. Scrolling to see the entire image was required in order for subjects to access the buttons to move through the study, ensuring that all subjects saw the entire image even in the 100% and 150% displays. The rest of the scale experiment consisted of a subset of the questions included in the main experiment; see full surveys given in the Appendix below.

The same subject screening and rejection criteria were used as in the main experiment. In addition, we rejected subjects who had previously participated in the main experiment, leaving N = 1127 (N = 288 in 10% condition, N = 274 in 36% condition, N = 292 in 100% condition, and N = 273 in 150% condition). In the analysis of image size, we pooled results from the main and in-laboratory experiment (described in the next section) to estimate the fraction of people who saw the dress as W/G at the 36% size (see **Figure 1G**).

In-laboratory experiment
One possibility is that the individual differences reported for the dress image are simply the result of differences in the conditions in which different subjects viewed the image, rather than differences in how the brains of different people assign colors to the same stimulus. To test the extent to which individual differences observable in our main experiment were also present in subjects who performed the survey under controlled viewing conditions on the same display, we tested 53 subjects ($N_{males} = 22$, $N_{females} = 31$) in laboratory. Subjects viewed the display at a distance of 40cm. We re-tested 28 of these subjects 3-5 days after their first experimental session in order to estimate intra-subject repeat reliability.

All in-laboratory subjects performed the experiment through the same Mechanical Turk interface used by subjects in the main experiment. The experiment was performed on a calibrated 21.5-inch iMac computer with 1920 x 1080 pixel resolution in a windowless room with LED overhead lighting (C.I.E. xyY = .4814 .4290 4.3587, measured off the MacBeth color checker standard

white, held at the same location and viewing angle on the monitor at which we presented the dress image). Normal color vision was confirmed with Ishihara plates.

Analysis Methods
The analysis of free response and color-matching (**Figure 1,C-F**) was conducted using the combined data from the main experiment (N = 1074), the 36% condition of the scale experiment (N = 274), and the in-laboratory control experiment (N = 53) (pooled N = 1401; sample was gender balanced: N_{males} = 731, $N_{females}$ = 668, N_{other} = 2). Of the pooled data (online and in-laboratory), 1190 subjects reported being from the US (1147 were online subjects), 161 from India (160 were online subjects), and 50 were from another country (41 were online).

Logistic regression models
To identify variables that significantly predicted subjects' free response answers (for the original image, image (1) described above), we created logistic regression models of two binary pairs of interest: reports of white/gold (success) versus blue/black (N = 1221; "Model 1") and reports of white/gold or blue/black (success) versus any other response (N = 1399; "Model 2"). Models were implemented in the R statistical software package using the function "glm" with the logit link function. In both models, we included the following predictor variables: study condition (main experiment order condition A, main experiment condition B, scale experiment 36% condition, in-laboratory experiment), gender (male or female; we removed 2 subjects who indicated "other" for the purposes of the models), country (United States, India, or other), age, whether subject reported having viewed the original dress image prior to this study ("prior view", naïve or non-naïve), and whether the subject reported having the software f.lux installed on their computer. P values for predictors were calculated using the Wald chi-square test. See **Table S1** for model results.

In addition to the full group models, we separately modeled the responses of the subset of subjects who had seen the original image prior to our study and responded with a main category (N = 962; "Model 4" of white/gold [success] versus blue/black responses). We included as predictor variables study condition, gender, country, age, f.lux, and whether subject reported having seen the image change colors at any point prior to our study (**Table S1**).

Logistic regression results documenting the impact of previous view
Overall, 57.0% (95% C.I. = [54.4%-59.5%]) of subjects described the dress as blue/black (B/K); 30.3% [28.0%-32.8%] as white/gold (W/G); 10.6% [9.1%-12.4%] as blue/brown (B/B); and 2.1% [1.4%-3.0%] gave some other description. Non-naïve subjects used a smaller number of unique color terms than naïve subjects (Figure 1C; **Figure S1A**, p < 0.0001, one-sample T-test). Residents of India were less likely than residents of the US to use a main category (B/K, W/G) rather than B/B or other (p = 0.0016, Wald chi-square; OR=0.48, 95% C.I.=[0.31-0.76]; **Table S1, Figure S1B**). The Indian population contained relatively more naïve subjects than the US population (60% versus 17%). To explore the possibility that country masked an important difference between naïve and non-naïve participants, we created "Model 3," in which we included all original predictors except for country. When country was removed from the logistic regression model ("Model 3"), prior experience became a significant predictor: non-naïve subjects were more likely to choose B/K or W/G, over B/B or other (p = 0.021, Wald chi-square; OR= 1.53 [1.06-2.17]) (see **Table S1**). The description B/B (or other) would seem to be categorically different from the descriptions W/G and B/K insofar as B/B comprises two chromatic colors (as opposed to a chromatic color and an achromatic term). That subjects with prior experience of the image are less likely to report B/B is consistent with the conclusion that some subjects actually undergo a change in their perception of the dress following experience.

Together, these results suggest that the media coverage of the image (#TheDress) changed the language people used to describe the image and possibly their perception of it.

Size experiment
In **Figure 1D**, we show the percentage of main category respondents who responded white-gold as a function of age. Data was based on subject responses in each condition of the scale experiment, and pooled with the main and in-laboratory experiment data for the 36% image size data point. To explore the effects of image size on percept, we created a logistic regression model of white/gold versus blue/black responses; we included image size as the only predictor.

Confidence intervals
Unless otherwise noted, all reported 95% confidence intervals were calculated analytically in the R statistical package using the modified Wald method [S1].

Analysis of individual differences: variability in lightness (L) and chromaticity (u,v) of color-matching data
Subjects made color matches using the color picker tool from the Raphael JavaScript library (http://raphaeljs.com/license.html). This tool allowed subjects to freely vary the hue, saturation and lightness to obtain the best match. Color matches were recorded as the HEX values generated by the tool, and converted to Luv [S2]. We chose the Luv color space for the analysis because this space is more-or-less perceptually uniform. The color match made by a subject for a given patch of the dress (i, ii, iii or iv, Figure 1A) was defined in the three dimensions of Luv color space. **Figure 1E** and **Figure S1C** show color matches sorted by the description given by the subject, B/K or W/G, for regions in the dress (see i, ii, iii, iv, **Figure 1A**). (Color matches for remaining subjects not shown, except in insets, **Figure 1E, Figure S1C**.) White triangles show population averages (upward triangles for regions i and ii; downward triangles for regions iii and iv; symbols contain the 95% C.I. of the mean). Insets depict color matches for individual subjects in each row, sorted by category (B/K, W/G, B/B, other). The 95% C.I. of the mean was defined as: mean + (T-score x standard error).

For a given region, the average color match made by W/G perceivers differed in both lightness (L) and hue (u, v) from the match made by B/K perceivers (Figure 1E). The magnitude of difference in the lightness dimension (L) was 20, 13.5, 13.3, and 20.7 (for regions i, ii, iii, and iv respectively), and this difference was significant for each region (p values < 0.0001, t-test). The magnitude of difference in the 'u' chromaticity dimension was 9.2, 6.6, 6.5, and 9.9 (for regions i, ii, iii, and iv respectively), and this difference was significant for each region (p values < 0.0001, t-test). The magnitude of difference in the 'v' chromaticity dimension was 23.9, 24.6, 26.6, and 25 (for regions i, ii, iii, and iv respectively), and this difference was significant for each region (p values < 0.0001, t-test).

The same analysis was also conducted separately for subsets of subjects: in-laboratory and online. The differences in magnitude of the lightness dimension and the chromaticity dimensions (u, v) were significant for the online data (and comparable in magnitude to the numbers provided above). The differences in magnitude of the lightness dimension and the chromaticity dimensions of the in-lab data were also significant (although the absolute magnitudes were smaller): The magnitude of difference in the lightness dimension (L) was 5.8, 6.8, 9.1, and 12.4 (for regions i, ii, iii, and iv respectively), and this difference was significant for each region (ps < 0.01, t-test). The magnitude of difference in the 'u' chromaticity dimension was 3, 6, 6.7, and 4.5 (for regions i, ii, iii, and iv respectively), and this difference was significant for all regions except for 'i' (ps < 0.02, t-test). The magnitude of difference in the 'v' chromaticity dimension was 7.2, 12.5, 16.2, and 18.5 (for regions i, ii, iii, and iv respectively), and this difference was significant for each

region (ps < 0.01, t-test). Differences in the demographics of the on-line and in-laboratory subjects may partially account for these differences (the in-laboratory group was more homogenous, mostly undergraduates and graduate students).

Principal component analysis (PCA) for color-matching data and test-retest reliability
The color match made by a subject for a given patch of the dress (i, ii, iii or iv, **Figure 1A**) was defined in the three dimensions of Luv color space. In order to obtain a single metric for the color match made by each subject, for a given patch in the dress, we performed PCA on the 3-D cloud of data points for the entire population. We ran two principal component analyses: first, using all data obtained from color matches to the "blue" pixels (ii, iii, **Figure 1A**); second, using all data obtained from color matches to the "brown" pixels (i, iv, **Figure 1A**). Note that the color terms "blue" and "brown" refer to the colors that most people assign to the pixels when the pixels are presented in isolation on a neutral background. The first (white to blue, "WB") component of the PCA produced a vector in the 3-D color space that captured most (86.4%) of the variance in the data. The second vector (gold to black, "GK") also captured a sizable amount of the variance (71.2%). The (x,y) values for each subject in the density plots shown in **Figure 1F**, **Figure S1D, E, F, and G** were defined as the weight of that subject's color match along the first principal components along the WB and GK vectors. Conducting the PCA on the entire set of color matches enabled us to define a common set of axes on which to display color-matches made by the different populations (naïve, non-naïve, in-laboratory). Analyzing the results using PCA conducted on each population separately (not shown) did not alter the main conclusions. When defined only using data collected in laboratory, the WB component explained 91.6% of the variance, while GK component accounted for 77.3% of the variance.   Using the PCA weights, we were able to: (1) test the reliability of the color matching in a subpopulation (N=28/53) that was re-tested in the laboratory (testing sessions were 3-5 days apart); (2) assess the extent to which color matches for 'blue' regions predicted matches to 'brown' regions; (3) assess the extent to which the distribution of percepts in the population was continuous or categorical.

(1) Test-retest reliability:
We sought to assess the reliability of the individual differences, to compare with individual differences measured in other circumstances [S3]. The challenge posed by this goal, however, is that the image of the dress is not only seen differently by different people, but also seen differently by the same person at different points of time –i.e. the image is multistable for some subjects. We re-tested about half (N=28) of the in-laboratory subjects in a separate session, 3-5 days after the first testing session. **Figure S1D** shows the intra-subject reliability of color descriptors used in sessions 1 and 2 (11% of subjects switched between sessions). We also compared the color matches that these subjects made for the identical patches seen during the two sessions (**Figure S1E**). Because of multistability, the intra-subject reliability measure provides a low estimate for the true reliability. Nonetheless, the reliability scores were significant (linear correlation: regions ii, iii, R = 0.4, p = 0.002; regions ii, iv, R = 0.6, p = $3*10^{-6}$).

(2) Color matches for one region predict another:
Figure 1F shows a density plot in which we plot the color matches (PCA weights) for 'brown' region i versus 'blue' region ii for all subjects irrespective of the color description they used for the dress (**Figure S1G** shows the naïve subjects only). Comparing color-match judgments for neighboring regions (i vs ii) controlled for subtle shifts in illumination across the image. Importantly, the results and conclusions did not change when we plotted iii vs iv, or when we plotted both 'brown' regions (i and iv) against both 'blue' regions (ii and iii) (data not shown). In the case of the in-laboratory data, we did not have sufficient data to plot only regions i and ii, so **Figure S1F** shows a density plot in which we plot the color matches for both 'brown' regions (i and iv) against those for both 'blue' regions (iii and iv).

In each density plot, the color scale bar shows the number of subjects at each location in the plot. The use of a density plot is necessary because with over 1400 subjects, data points for many subjects overlap. The maximum color in the scale bar is in **Figure 1F** 2.3 subjects (2.1 in **Figure S1F**, and 3.1 in **Figure S1G**). In addition, the plots were smoothed to reveal any underlying structure. The smoothing was a Gaussian of radius 0.0068 PC units in Figure 1F (0.0027 in **Figure S1F**; 0.0059 in **Figure S1G**). Importantly, the color-matching data are obtained independent of the color descriptions used by the subjects. The density plots clearly show two features: first, color matches to region i are highly predictive of color matches to region ii (**Figure 1F**, R=0.59; p=10^{-264}; **Figure S1F**, R = 0.55, p = $7*10^{-10}$; **Figure S1G**, R = 0.55, p = $6*10^{-26}$, where the correlation in **Figure S1G** shows how well color matches for both 'blue' regions (ii and iii) predict matches for both 'brown' regions (i and iv)). And second, there is inhomogeneity along the line of maximum correlation: showing evidence for three distinct clusters, discussed next.

(3) Peaks in color match density plots correspond to the three main categories of color descriptions:

We sought to determine the extent to which these clusters in the density plots correspond to the three main groups of subjects, those reporting B/K, W/G or B/B. Using the independent color description data provided by the subjects, we sorted the color matches into four sub-populations defined by main-category description (W/G, B/K, B/B, and other). We then generated a density plot for each sub-population and determined contours that captured the peak density (containing at least 25% of the data). The black squiggle contours in **Figure 1D and Figure S1G** show these contours superimposed on the density plot obtained for the whole population. The contours are coincident with the three clusters in the population data, showing that the color-matching data uncovers the three main descriptive reports. **Figure S1F** does not have squiggles because the N is low; nonetheless, that panel also shows three peaks.

RGB (HEX) to Luv conversion
The white point used to convert RGB (HEX) to CIE values (**Figure 1B**) and Luv values (**Figure 1E,F**) was the standard CIE 1931 F7 illuminant typical of many displays [95.047 100 108.883, XYZ] (Westland, p 65). The white point for the monitor used to collect data in the laboratory was different [34.49 37.67 35.03, XYZ], but we analyzed the in-lab data using the F7 illuminant to facilitate comparison with the MTurk data (**Figure S1C,E,F,G**). The conclusions are no different if we analyze the in-lab data using the lab-specific white point (data not shown). Similarly, when the in-lab white point was defined using the measured mean of the dress image [16.38 16.6 19.41, XYZ], the conclusions remained the same (data not shown).

Histograms for unique color terms
Subjects who had previously seen the image used a smaller number of unique color terms than naïve subjects (p<0.0001, one-sample T-test). In Figure S1A we show the number of unique color terms used by the subset of subjects who had never seen the dress image before the study ("Naïve", N = 313) and the subset of subjects who had seen the image previously ("Not-naïve", N = 1088). Both populations show evidence for three dominant categories: B/K, W/G and B/B. But there are subtle differences in the number of unique color terms used and, as described above, subjects' perception of the dress. In order to quantify the number of unique terms, controlling for the difference in absolute number of subjects in each population, we subsampled the not-naïve population, without replacement 10,000 times. The number of samples drawn on each subsample was 313, and the number of unique color terms was tabulated. A one-sample t-test was performed to determine whether the number of unique color terms derived from the not-naïve population could have come from a distribution whose mean was equal to the number of color terms used by the naïve population (p < 0.0001).

Many of the unique color terms appeared to refer to the same perception. For example, "white-golden", "white-goldish", and "white-gold" were deemed redundant. Color descriptions deemed redundant were binned. Four categories emerged: Blue/Black (B/K), White/Gold (W/G), Blue/Brown (B/B), or Other. These categories were distinguished as "Main" categories (B/K or W/G) or not-main categories (B/B or other). This distinction reflects the use of one achromatic and one chromatic color term for the Main categories, as opposed to the use of two chromatic color terms for the not-Main category.

The complete list of unique color terms, and how they were binned is as follows:

Blue/black category: 'blue, black', 'blue, gray', 'light blue, gray', 'light blue, black', 'pale blue, black', 'blue, dark gray', 'dark blue, black', 'deep periwinkle blue, black', 'light blue, light black', 'periwinkle, dark gray', 'black, navy blue'

White/gold category: 'white, gold', 'white, brown', 'white, golden', 'white, tan', 'white, yellow', 'copper, white', 'gold, off-white', 'gold, white (white light blue)', 'gray, brown', 'gray, gold', 'white, golden brown', 'white, khaki', 'white, chocolate', 'white, sandy brown'

Blue/brown category: 'blue, brown', 'blue, gold', 'brown, light blue', 'gold, light blue', 'blue, golden', 'blue, tan', 'blue, dark bronze', 'blue, dark brown', 'dark brown, light blue', 'gold, baby blue', 'light blue, brownish gold', 'periwinkle, gold', 'sky blue, dark brown', 'sky blue, brown', 'dark blue, golden black', 'white-blue, gold', 'blue, bronze', 'blue, gold/black'

Other category: 'purple, brown', 'purple, black', 'ash, black', 'blue, olive', 'golden, lilac', 'lavender, black', 'lavender, gold', 'light violet, black', 'pink, black', 'purple, gold', 'purple, tan', 'violet, black', 'violet, gold', 'white, cream', 'yellow, gold', 'white, pink', 'purple green, light blue', 'gray, black', 'blue, white', 'white, black', 'gold, black'

Figure S1B shows the breakdown of color term usage for in-lab, India (online), and US (online) subjects. The distributions are similar, but the population from India shows more non-main-category terms. Consistent with the hypothesis that experience impacts the likelihood of using a main-category descriptor, the Indian population contained relatively more naïve subjects than the US population (60% versus 17%).

Compelling a particular perception of the dress (W/G or B/K) by embedding it in a scene, on a model
We surveyed color descriptions for the dress embedded two scenes with different simulated lighting conditions (image 2 and 3 described above; N = 1127, 1047 from the main experiment and 53 from the in-laboratory controlled experiment) (**Figure S2**). Of subjects who described the dress originally as white/gold, 64% switched to seeing it as blue/black under the simulated warm light; of those who saw the dress as blue/black originally, 80% switched to seeing it as white/gold under the simulated cool light. Subjects who described the original image as blue/brown (or blue/gold) were just as likely to maintain their description when shown the dress under the warm illuminant, consistent with the idea that they are doing a local pixel-color analysis (or assuming a neutral illuminant) (although these subjects were more likely to switch to white/gold given the cool illuminant).

SUPPLEMENTAL DISCUSSION

Source of the individual differences

Many hypotheses have been proposed regarding the source of the individual differences. Our study shows that age and gender are the two most significant predictors. But these correlations do not reveal the causes of the individual differences. A plausible hypothesis is that different people have different priors about the spectral content of the illuminant, with the population of priors varying along the daylight locus [S4]. Consistent with this hypothesis, observers show greater variability along the daylight locus when adjusting a patch on a screen to be "white" (our unpublished results; see also [4 S5], but see [6]). We have not found a strong correlation between the chromaticities subjects report to be "white" and the likelihood that they see the dress as W/G or B/K (data not shown). Nonetheless, it remains possible that people who experience more daylight are more likely to develop a "blue" illuminant prior, since the sky is blue; whereas, people who spend more of their time awake at night are more likely to develop a "warm" illuminant prior, since (historically) most artificial lights are biased for long wavelengths (i.e. "warm"). Other subjects may favor a neutral illuminant (and see the dress as B/B). If the illumination-discounting theory is correct, and different people have different priors on the illumination determined by their exposure, we would predict that "larks" (daytime chrontotype) will show a bias for W/G (discounting the blue in the dress); and "night owls" (night-time chronotype) will show a bias for B/K (discounting the orange in the dress). Does the extent to which one is exposed to a given illuminant (blue-sky daylight versus warm incandescent light) predict the colors one sees the dress? We attempted to answer this question by asking subjects about their sleep-wake patterns. We found a slight, but insignificant trend: subjects reporting more awake time during the day were slightly more likely to see the dress as W/G. Although the results provided by directly surveying sleep-wake patterns are at present inconclusive (we may need to sort data by longitude etc.), our results provide some support for the chronotype hypothesis. Older subjects and women are more likely to have a daytime chronotope [6]. Consistent with this, we found that these two variables significantly predicted a higher likelihood of reporting the dress to be W/G. Lens density, corneal coloration, and other optical properties of the eye vary with age. It is also a possibility that these factors contribute to the individual differences observed presently.

Cultural influences on perception of 'the dress'?
In the group of all male and female subjects (N = 1399), age (p < 0.0001 Wald chi-square) and a subject's country (India versus USA) (p = 0.0016) were significant in predicting a main category response versus a non-main category response (blue/brown or other): older subjects and people from India were less likely to report a main category (see **Table S1** for full model results). These results may provide some evidence for cultural influences in how the dress image is perceived. Alternatively, the Indian population contained many more naïve subjects relative to the USA population. When we removed country from the logistic regression model, prior experience of the image significantly predicted response (p = 0.021, Wald chi-square): the odds of reporting a main category increased by a factor 1.53 [1.06-2.17] for subjects who had seen the image previously.

Individual differences in multistability
Non-naïve subjects who saw the dress in the main experiment as W/G rather than B/K were less likely to have reported a switch in the color of the dress prior to our study. Given the results of our size experiment, we can conclude that the size of the image used in the main experiment biased subjects towards reporting B/K. We hypothesize that the subjects who continue to see the dress as W/G at the smaller sizes used in the main experiment were less likely to experience 'the dress' as multistable—these subjects, on average, seemed to be less susceptible to switching. This hypothesis is supported by the data from the experiments using the images embedded in scenes with simulated illuminants. Fewer W/G respondents (64%) switched their perception to B/K when the dress was embedded in a warm-illuminated scene, compared to the fraction of B/K respondents (80%) who switched their perception when the dress was embedded in a cool-

illuminated scene (Figure S2). These results provide evidence of striking individual differences not only in the perceived color, but also in the multistability of the image.

In the Main experiment, subjects performed the color matching for the various regions of the dress (i, ii, iii, iv, **Figure 1A**), after they had been queried about the terms used for their perception of the dress under the different contexts (images 2, 3). We found that the particular image (2 or 3) that was shown before the color matching was conducted biased the color-matching results (data not shown), providing evidence of hysteresis. We are investigating the extent to which prior exposure, over various time scales, impacts perception of the dress (experiments in progress).

Effects of image size
In the Scale experiment, we found that a reduction in image size biased the population who reported a main category (B/K or W/G) to see the dress as B/K (with a unit increase in image size, the odds of reporting white/gold increased by a factor 1.004 [1.002 – 1.007]; $p < 0.0001$). As described in the main text, we hypothesize that this difference may be attributable to a change in spatial frequency that renders less material information available. Alternatively the impact of size on perceived color may reflect differences in how the image is perceived when more or less of it is seen at a given time. In addition, it is possible that when only a portion of the image is seen at a given time, the particular portion of the image that is seen first has an impact. In the 100% and 150% conditions of the Scale experiment, only part of the image was visible at once: the top of the image was the first part of it seen. Subjects then had to scroll over the image, from top to bottom, to view it in full; in the 10% and 36%, the full image was visible on the screen at once. The pixel colors of the regions at the top of the dress are lighter than those at the bottom of the dress (**Figure 1A,B**). A subject integrating over the full image at once might arrive at a different percept than a subject who sees the lighter portion first, and then scrolls down over the rest of the image. If true, it is possible that a subject's first fixation on the dress image might influence their report of its color (experiments in progress).

APPENDIX: FULL SURVEYS

MAIN EXPERIMENT.

Page 1
In this study, we will show you a number of images. We want you to look at these images carefully, and then answer some questions about them. If you're not sure about an answer, take your best guess. Please answer all questions before proceeding to the next page.

Please try to complete the study in a single sitting, without taking breaks.

Page 2
Stimulus: dress on original background (presented for 15 seconds)

Please look carefully at the dress. A "continue" button will appear just below this text after 15 seconds.

Page 3
Stimulus: dress on original background

This is a picture of a _____ and _____ dress. (Fill in the colors that you see
Two textboxes were provided.

Stimulus: dress on original background, with arrows pointing to regions labeled A, B, C, and D**

**regions A, B, C, and D noted here correspond to Figure 1A regions i, ii, iii, and iv, respectively

Now we would like you tell us about the material properties (fabric) of the dress. For each of the indicated regions (A, B, C, and D) in the diagram above, what kind of material do you think the dress is made of? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*1. How would you describe the material at (A)?
    Pull-down menu with options as follow:
- o   semi-transparent, sheer; chiffon, tulle, organza
- o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o   semi-gloss, med-luster, med-sheen; rayon, polyester
- o   hi-gloss, hi-sheen; silk, glossy-satin

*2. How would you describe the material at (B)?
    Pull-down menu with options as follow:
- o   semi-transparent, sheer; chiffon, tulle, organza
- o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o   semi-gloss, med-luster, med-sheen; rayon, polyester
- o   hi-gloss, hi-sheen; silk, glossy-satin

*3. How would you describe the material at (C)?
    Pull-down menu with options as follow:
- o   semi-transparent, sheer; chiffon, tulle, organza
- o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o   semi-gloss, med-luster, med-sheen; rayon, polyester
- o   hi-gloss, hi-sheen; silk, glossy-satin

*4. How would you describe the material at (D)?
    Pull-down menu with options as follow:
- o   semi-transparent, sheer; chiffon, tulle, organza
- o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o   semi-gloss, med-luster, med-sheen; rayon, polyester
- o   hi-gloss, hi-sheen; silk, glossy-satin

Stimulus: dress on original background

*1. Does the lighting on the dress look different than the lighting in the background?
    Pull-down menu with options as follow:
- o   Yes
- o   No

*2. The lighting in the background is...Check all that apply:

Check boxes appeared next to the following options:

o warm
o bright
o cool
o dim
o dark
o yellowish
o bluish
o reddish
o greenish
o purplish
o glaring
o iridescent
o washed out
o blown out

*3. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the lighting conditions in the background.

Pull-down menu with options as follow:

o 1
o 2
o 3
o 4
o 5

*4. The light illuminating the dress is...Check all that apply:

Check boxes appeared next to the following options:

o warm
o bright
o cool
o dim
o dark
o yellowish
o bluish
o reddish
o greenish
o purplish
o glaring
o iridescent
o washed out
o blown out

*5. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the light illuminating the dress.

Pull-down menu with options as follow:

o 1
o 2
o 3
o 4
o 5

DEMOGRAPHICS

Please help us out by answering the following questions about yourself. You must answer all of the questions marked with an asterisk (*).

*1. What is your gender?
> *Pull-down menu with options as follow:*
- o   Male
- o   Female
- o   Other

*2. How old are you?
> *Pull-down menu with options 18-100*

*3. Is English your mother tongue?
> *Pull-down menu with options as follow:*
- o   Yes
- o   No

4. If English is not your mother tongue, what is your mother tongue?
> *A pull-down menu contained a comprehensive list of languages, including Tamil and other languages spoken in India. List of languages available on request.*

*5. What country are you from?
> *A pull-down menu contained a comprehensive list of countries. List available on request.*

5 (A). If you're from the United States, what part of the country are you from?
> *Pull-down menu with options as follow:*
- o   Northeast - New England
- o   Northeast - Mid-Atlantic
- o   Midwest - East North Central
- o   Midwest - West North Central
- o   South - South Atlantic
- o   South - East South Central
- o   South - West South Central
- o   West - Mountain

*6. Which levels of education have you completed? Check all that apply.
> *Check boxes appeared next to the following options:*
- o   Elementary school (primary school)
- o   High school
- o   College/university
- o   Advanced graduate work

*7. Do you have normal or corrected-to-normal vision?
> *Pull-down menu with options as follow:*
- o   Yes
- o   No

*8. Are you colorblind?

Pull-down menu with options as follow:
- o Yes
- o No
- o Unsure

*9. On most days, what time do you typically wake up
   Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"

*10. On most days, what time do you typically go to bed?
   Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"

11. Do you have a background in the visual arts? Check all that apply:
   Check boxes appeared next to the following options:
- o Professionally
- o Hobby
- o Painter
- o Graphic designer
- o Art historian
- o Architecture
- o Illustrator
- o Photographer

Page 7
Stimulus: dress on first colored background (warm or cool illuminant depending on condition; presented for 15 seconds)

Please look carefully at the dress. A "continue" button will appear just below this text after 15 seconds.

Page 8
Stimulus: dress on first colored background

This is a picture of a _____ and _____ dress. (Fill in the colors that you see
Two textboxes were provided.

Page 9
Stimulus: dress on first colored background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. For each of the indicated regions (A, B, C, and D) in the diagram above, what kind of material do you think the dress is made of? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*1. How would you describe the material at (A)?
   Pull-down menu with options as follow:
- o semi-transparent, sheer; chiffon, tulle, organza
- o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin

- o semi-gloss, med-luster, med-sheen; rayon, polyester
- o hi-gloss, hi-sheen; silk, glossy-satin

*2. How would you describe the material at (B)?
  Pull-down menu with options as follow:
- o semi-transparent, sheer; chiffon, tulle, organza
- o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o semi-gloss, med-luster, med-sheen; rayon, polyester
- o hi-gloss, hi-sheen; silk, glossy-satin

*3. How would you describe the material at (C)?
  Pull-down menu with options as follow:
- o semi-transparent, sheer; chiffon, tulle, organza
- o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o semi-gloss, med-luster, med-sheen; rayon, polyester
- o hi-gloss, hi-sheen; silk, glossy-satin

*4. How would you describe the material at (D)?
  Pull-down menu with options as follow:
- o semi-transparent, sheer; chiffon, tulle, organza
- o flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o semi-gloss, med-luster, med-sheen; rayon, polyester
- o hi-gloss, hi-sheen; silk, glossy-satin

Page 10
Stimulus: dress on first colored background

*1. Does the lighting on the dress look different than the lighting in the background?
  Pull-down menu with options as follow:
- o Yes
- o No

*2. The lighting in the background is...Check all that apply:
  Check boxes appeared next to the following options:
- o warm
- o bright
- o cool
- o dim
- o dark
- o yellowish
- o bluish
- o reddish
- o greenish
- o purplish
- o glaring
- o iridescent
- o washed out
- o blown out

*3. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the lighting conditions in the background.

Pull-down menu with options as follow:
o 1
o 2
o 3
o 4
o 5

*4. The light illuminating the dress is...Check all that apply:
     Check boxes appeared next to the following options:
o warm
o bright
o cool
o dim
o dark
o yellowish
o bluish
o reddish
o greenish
o purplish
o glaring
o iridescent
o washed out
o blown out

*5. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the light illuminating the dress.
     Pull-down menu with options as follow:
o 1
o 2
o 3
o 4
o 5

Page 11
Tell us a bit about your environment.

*1. What device are you using to view this webpage?
     Pull-down menu with options as follow:
o Laptop
o Computer monitor (flat screen
o Computer monitor (rounded screen
o Tablet
o Other

*2. How big is the screen on which you're viewing this webpage? (Take your best guess if you're not sure.)
     Pull-down menu with options as follow:
o 10-13 inches wide
o 14-17 inches wide (regular laptop)
o 14-17 inches wide
o More than 17 inches wide

*3. Are you using a Mac or PC?
>   *Pull-down menu with options as follow:*
>   o   Mac
>   o   PC
>   o   Unsure

*4. What time is it right now? (select the approximate hour)
>   *Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"*

*5. Are you indoors or outdoors?
>   *Pull-down menu with options as follow:*
>   o   Indoors
>   o   Outdoors

*6. If you're indoors, is there natural light entering the room through a window?
>   *Pull-down menu with options as follow:*
>   o   N/A
>   o   Yes
>   o   No

*7. What are the lighting conditions where you are right now, on a scale of 1 (darkest) to 10 (brightest)?
>   *Pull-down menu with options as follow:*
>   o   1 (darkest; as dim as a room without lights at night
>   o   2
>   o   3
>   o   4
>   o   5
>   o   6
>   o   7
>   o   8
>   o   9
>   o   10 (brightest; as bright as the beach on a sunny day

*8. Is the software f.lux installed on the computer you're using to do this HIT?
>   *Pull-down menu with options as follow:*
>   o   Yes
>   o   No

Page 12
Stimulus: dress on second colored background (warm or cool illuminant depending on condition; presented for 15 seconds)

Please look carefully at the dress. A "continue" button will appear just below this text after 15 seconds.

Page 13
Stimulus: dress on second colored background

This is a picture of a _____ and _____ dress. (Fill in the colors that you see
Two textboxes were provided.

Stimulus: dress on second colored background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. For each of the indicated regions (A, B, C, and D) in the diagram above, what kind of material do you think the dress is made of? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*1. How would you describe the material at (A)?
   Pull-down menu with options as follow:
   o   semi-transparent, sheer; chiffon, tulle, organza
   o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
   o   semi-gloss, med-luster, med-sheen; rayon, polyester
   o   hi-gloss, hi-sheen; silk, glossy-satin

*2. How would you describe the material at (B)?
   Pull-down menu with options as follow:
   o   semi-transparent, sheer; chiffon, tulle, organza
   o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
   o   semi-gloss, med-luster, med-sheen; rayon, polyester
   o   hi-gloss, hi-sheen; silk, glossy-satin

*3. How would you describe the material at (C)?
   Pull-down menu with options as follow:
   o   semi-transparent, sheer; chiffon, tulle, organza
   o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
   o   semi-gloss, med-luster, med-sheen; rayon, polyester
   o   hi-gloss, hi-sheen; silk, glossy-satin

*4. How would you describe the material at (D)?
   Pull-down menu with options as follow:
   o   semi-transparent, sheer; chiffon, tulle, organza
   o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
   o   semi-gloss, med-luster, med-sheen; rayon, polyester
   o   hi-gloss, hi-sheen; silk, glossy-satin

Stimulus: dress on second colored background

*1. Does the lighting on the dress look different than the lighting in the background?
   Pull-down menu with options as follow:
   o   Yes
   o   No

*2. The lighting in the background is...Check all that apply:

Check boxes appeared next to the following options:

- o warm
- o bright
- o cool
- o dim
- o dark
- o yellowish
- o bluish
- o reddish
- o greenish
- o purplish
- o glaring
- o iridescent
- o washed out
- o blown out

*3. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the lighting conditions in the background.

Pull-down menu with options as follow:

- o 1
- o 2
- o 3
- o 4
- o 5

*4. The light illuminating the dress is...Check all that apply:

Check boxes appeared next to the following options:

- o warm
- o bright
- o cool
- o dim
- o dark
- o yellowish
- o bluish
- o reddish
- o greenish
- o purplish
- o glaring
- o iridescent
- o washed out
- o blown out

*5. On a scale from 1 to 5, where 1 is cool and 5 is warm, please rank the light illuminating the dress.

Pull-down menu with options as follow:

- o 1
- o 2
- o 3
- o 4
- o 5

Stimulus: dress on original background, with arrow pointing to region (A). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*1. What color is the dress at (A)?

Stimulus: dress on original background, with arrow pointing to region (B). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*2. What color is the dress at (B)?

Stimulus: dress on original background, with arrow pointing to region (C). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*3. What color is the dress at (C)?

Stimulus: dress on original background, with arrow pointing to region (D). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*4. What color is the dress at (D)?

Stimulus: dress on first colored background, with arrow pointing to region (A). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*1. What color is the dress at (A)?

Stimulus: dress on first colored background, with arrow pointing to region (B). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*2. What color is the dress at (B)?

Stimulus: dress on first colored background, with arrow pointing to region (C). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*3. What color is the dress at (C)?

Stimulus: dress on first colored background, with arrow pointing to region (D). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*4. What color is the dress at (D)?

Stimulus: dress on second colored background, with arrow pointing to region (A). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*1. What color is the dress at (A)?

Stimulus: dress on second colored background, with arrow pointing to region (B). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*2. What color is the dress at (B)?

Stimulus: dress on second colored background, with arrow pointing to region (C). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*3. What color is the dress at (C)?

Stimulus: dress on second colored background, with arrow pointing to region (D). Color-picker appeared to the right of the image.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*4. What color is the dress at (D)?

*1. Had you ever seen this image prior to this study?
> *Pull-down menu with options as follow:*
o   Yes
o   No

If you responded "yes" to questions #1 above, please answer the following questions:

2. When you first saw this image, what color did you see the dress? (Fill in the colors that you saw)
> *Two textboxes were provided.*

3. In your viewings prior to this study, did your perception of the dress colors ever change?
> *Pull-down menu with options as follow:*
o   Yes
o   No

4. If it did change, please indicate all combinations that you've saw. (Check all that apply)
> *Check boxes appeared next to the following options:*
o   blue and black
o   blue and olive
o   blue and brown
o   blue and orange
o   white and gold
o   white and orange
o   lavender and gold
o   light blue and gold

5. How often did you see it change?
> *Pull-down menu with options as follow:*
o   Never
o   Infrequently

     o   Frequently

Over what time scale did it change?
       Pull-down menu with options as follow:
     o   N/A
     o   over the course of minutes
     o   over the course of hours
     o   over the course of days

Page 29
Stimulus: dress on original background, with alpha-numeric grid superimposed on top

*1. When you first saw the dress (whether in this study or prior), do you recall where you first looked? Give your best estimate by indicating a square in the grid below (e.g. D2).
       Two pull-down menus appeared, one with letters, and one with numbers, corresponding to the rows and columns of the grid.

2.When you've look at this image, where have you spent most of your time looking? (again, use the grid to indicate)
       Two pull-down menus appeared, one with letters, and one with numbers, corresponding to the rows and columns of the grid.

Page 30
Stimulus: blurred dress on original background (see Methods for stimulus description)

*3. What color is the dress in this blurry image?
       Two textboxes were provided.

*4. Do you know what color the dress is in real life? </p>
       Two textboxes were provided.

Page 31
Thank you very much for participating in this study.

Please enter any comments you have about this HIT in the box below:
       A textbox was provided.

Page 32
Thank you for your help!

SCALE EXPERIMENT.

Page 1
In this study, we will show you a number of images. We want you to look at these images carefully, and then answer some questions about them. If you're not sure about an answer, take your best guess. Please answer all questions before proceeding to the next page.

All questions appear at the bottom of the page. After answering questions, please click the "continue" button at the top of the page to proceed to the next page.

Please try to complete the study in a single sitting, without taking breaks.

Do not zoom in on the image. Please give your best estimate at this viewing size.

Stimulus: dress on original background (presented for 15 seconds)

Please look carefully at the dress. A "continue" button will appear just below this text after 15 seconds.

Stimulus: dress on original background

This is a picture of a ____ and ____ dress. (Fill in the colors that you see
Two textboxes were provided.

*Did you have to scroll over the image on the previous page to see all of it?
      Pull-down menu with options as follow:
   o   Yes, top to bottom
   o   Yes, bottom to top
   o   No

Do not zoom in on the image. Please give your best estimate at this viewing size.

Stimulus: dress on original background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. What kind of material do you think the dress is made of in the region to which the arrow is pointing? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*1. How would you describe the material where the arrow is pointing (neckline)?
      Pull-down menu with options as follow:
   o   semi-transparent, sheer; chiffon, tulle, organza
   o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
   o   semi-gloss, med-luster, med-sheen; rayon, polyester
   o   hi-gloss, hi-sheen; silk, glossy-satin

Do not zoom in on the image. Please give your best estimate at this viewing size.

Stimulus: dress on original background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. What kind of material do you think the dress is made of in the region to which the arrow is pointing? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*2. How would you describe the material where the arrow is pointing (sleeve?
     Pull-down menu with options as follow:
- o   semi-transparent, sheer; chiffon, tulle, organza
- o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o   semi-gloss, med-luster, med-sheen; rayon, polyester
- o   hi-gloss, hi-sheen; silk, glossy-satin

Page 7
Do not zoom in on the image. Please give your best estimate at this viewing size.

Stimulus: dress on original background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. What kind of material do you think the dress is made of in the region to which the arrow is pointing? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*3. How would you describe the material where the arrow is pointing (body material)?
     Pull-down menu with options as follow:
- o   semi-transparent, sheer; chiffon, tulle, organza
- o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o   semi-gloss, med-luster, med-sheen; rayon, polyester
- o   hi-gloss, hi-sheen; silk, glossy-satin

Page 8
Do not zoom in on the image. Please give your best estimate at this viewing size.

Stimulus: dress on original background, with arrows pointing to regions labeled A, B, C, and D

Now we would like you tell us about the material properties (fabric) of the dress. What kind of material do you think the dress is made of in the region to which the arrow is pointing? Below are some examples of each category. Take your best guess if you're not sure.

A panel showing gray-scale images of fabric swatches appeared at the center of the screen.

*4. How would you describe the material where the arrow is pointing (trim)?
     Pull-down menu with options as follow:
- o   semi-transparent, sheer; chiffon, tulle, organza
- o   flat matte, low-luster, low-sheen; cotton, linen, jersey, matte-satin
- o   semi-gloss, med-luster, med-sheen; rayon, polyester
- o   hi-gloss, hi-sheen; silk, glossy-satin

Page 9
DEMOGRAPHICS

Please help us out by answering the following questions about yourself. You must answer all of the questions marked with an asterisk (*).

*1. What is your gender?

Pull-down menu with options as follow:
- Male
- Female
- Other

*2. How old are you?
Pull-down menu with options 18-100

*3. Is English your mother tongue?
Pull-down menu with options as follow:
- Yes
- No

4. If English is not your mother tongue, what is your mother tongue?
A pull-down menu contained a comprehensive list of languages, including Tamil and other languages spoken in India. List of languages available on request.

*5. What country are you from?
A pull-down menu contained a comprehensive list of countries. List available on request.

5 (A). If you're from the United States, what part of the country are you from?
Pull-down menu with options as follow:
- Northeast - New England
- Northeast - Mid-Atlantic
- Midwest - East North Central
- Midwest - West North Central
- South - South Atlantic
- South - East South Central
- South - West South Central
- West - Mountain

*6. Which levels of education have you completed? Check all that apply.
Check boxes appeared next to the following options:
- Elementary school (primary school)
- High school
- College/university
- Advanced graduate work

*7. Do you have normal or corrected-to-normal vision?
Pull-down menu with options as follow:
- Yes
- No

*8. Are you colorblind?
Pull-down menu with options as follow:
- Yes
- No
- Unsure

*9. On most days, what time do you typically wake up

Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"

*10. On most days, what time do you typically go to bed?
Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"

11. Do you have a background in the visual arts? Check all that apply:
Check boxes appeared next to the following options:
o Professionally
o Hobby
o Painter
o Graphic designer
o Art historian
o Architecture
o Illustrator
o Photographer

Page 10
Tell us a bit about your environment.

*1. What device are you using to view this webpage?
Pull-down menu with options as follow:
o Laptop
o Computer monitor (flat screen
o Computer monitor (rounded screen
o Tablet
o Other

*2. How big is the screen on which you're viewing this webpage? (Take your best guess if you're not sure.)
Pull-down menu with options as follow:
o 10-13 inches wide
o 14-17 inches wide (regular laptop)
o 14-17 inches wide
o More than 17 inches wide

*3. Are you using a Mac or PC?
Pull-down menu with options as follow:
o Mac
o PC
o Unsure

*4. What time is it right now? (select the approximate hour)
Two pull-down menus appeared: one containing numbers 1-12, and one containing "A.M. (morning)" and "P.M. (night)"

*5. Are you indoors or outdoors?
Pull-down menu with options as follow:
o Indoors
o Outdoors

*6. If you're indoors, is there natural light entering the room through a window?
>   *Pull-down menu with options as follow:*
> o   N/A
> o   Yes
> o   No

*7. What are the lighting conditions where you are right now, on a scale of 1 (darkest) to 10 (brightest)?
>   *Pull-down menu with options as follow:*
> o   1 (darkest; as dim as a room without lights at night
> o   2
> o   3
> o   4
> o   5
> o   6
> o   7
> o   8
> o   9
> o   10 (brightest; as bright as the beach on a sunny day

*8. Is the software f.lux installed on the computer you're using to do this HIT?
>   *Pull-down menu with options as follow:*
> o   Yes
> o   No

Page 11
Stimulus: dress on original background, with arrow pointing to region (A). Color-picker appeared to the right of the image.

Do not zoom in on the image. Please give your best estimate at this viewing size.

Now we would like you to report on the color of the pixels you see in the image. We will provide a color-matching tool. Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*1. What color is the dress where the arrow is pointing (neckline)?

Page 12
Stimulus: dress on original background, with arrow pointing to region (B). Color-picker appeared to the right of the image.

Do not zoom in on the image. Please give your best estimate at this viewing size.

Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*2. What color is the dress where the arrow is pointing (sleeve)?

Page 13
Stimulus: dress on original background, with arrow pointing to region (C). Color-picker appeared to the right of the image.

Do not zoom in on the image. Please give your best estimate at this viewing size.

Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*3. What color is the dress where the arrow is pointing (body material)?

Page 14
Stimulus: dress on original background, with arrow pointing to region (D). Color-picker appeared to the right of the image.

Do not zoom in on the image. Please give your best estimate at this viewing size.

Please adjust the hue (color circle) and brightness (slider bar) to match the pixels you see in the image.

*4. What color is the dress where the arrow is pointing (trim)?

Page 15
Stimulus: dress on original background

*Is the dress in shadow?
    Pull-down menu with options as follow:
    o   Yes
    o   No

Page 16
Stimulus: dress on original background

*1. Had you ever seen this image prior to this study?
    Pull-down menu with options as follow:
    o   Yes
    o   No

If you responded "yes" to questions #1 above, please answer the following questions:

2. When you first saw this image, what color did you see the dress? (Fill in the colors that you saw)
    Two textboxes were provided.

3. In your viewings prior to this study, did your perception of the dress colors ever change?
    Pull-down menu with options as follow:
    o   Yes
    o   No

4. If it did change, please indicate all combinations that you've saw. (Check all that apply)
    Check boxes appeared next to the following options:
    o   blue and black
    o   blue and olive
    o   blue and brown

- o blue and orange
- o white and gold
- o white and orange
- o lavender and gold
- o light blue and gold

5. How often did you see it change?
    Pull-down menu with options as follow:
- o Never
- o Infrequently
- o Frequently

Over what time scale did it change?
    Pull-down menu with options as follow:
- o N/A
- o over the course of seconds
- o over the course of minutes
- o over the course of hours
- o over the course of days

<u>Page 17</u>
Stimulus: dress on original background

Do not zoom in on the image. Please make your best estimate at this viewing size.

*If you were forced to choose between one of these options, what color is the dress in this photograph?
    Pull-down menu with options as follow:
- o blue and black
- o white and gold

<u>Page 18</u>
*1. Do you know what color the dress is in real life?
    Pull-down menu with options as follow:
- o Yes
- o No

If you answered "yes" to question #1, please answer the following question

2. What is the true color of the dress?
    Two textboxes were provided.

<u>Page 19</u>
Thank you very much for participating in this study.

Please enter any comments you have about this HIT in the box below:
    A textbox was provided.

<u>Page 20</u>
Thank you for your help!

Supplemental References.

S1.   Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, *52*(2), 119-126.
S2.   Westland, S., and Ripamonti, C. (2004). Computational colour science using matlab. John Wiley and Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, England.
S3.   Wilmer, J.B., Germine, L., Chabris, C.F., Chatterjee, G., Gerbasi, M., and Nakayama, K. (2012). Capturing specific abilities as a window into human individuality: the example of face recognition. Cogn Neuropsychol *29*, 360-392.
S4.   Conway, B.R. (2015). Why do we care about the colour of the dress? The Guardian *http://www.theguardian.com/commentisfree/2015/feb/27/colour-dress-optical-illusion-social-media*.
S5.   Beer, R.D., Dinca, A., and MacLeod, D.I.A. (2006). Ideal white can be yellowish and bluish, but not reddish and greenish. Journal of Vision *6*, 417.
S6.   Adan, A., Archer, S.N., Hidalgo, M.P., Di Milia, L., Natale, V., and Randler, C. (2012). Circadian typology: a comprehensive review. Chronobiol Int *29*, 1153-1175.