

Color consistency in the appearance of bleached fabrics

Matteo Toscani*

Department of Psychology, Giessen University,
Giessen, Germany



Zarko Milojevic*

Department of Psychology, Giessen University,
Giessen, Germany



Roland W. Fleming

Department of Psychology, Giessen University,
Giessen, Germany



Karl R. Gegenfurtner

Department of Psychology, Giessen University,
Giessen, Germany



Human observers are remarkably good at perceiving constant object color across illumination changes. However, there are numerous other factors that can modulate surface appearance, such as aging, bleaching, staining, or soaking. Despite this, we are often able to identify material properties across such transformations. Little is known about how and to what extent we can compensate for the accompanying color transformations. Here we investigated whether humans could reproduce the original color of bleached fabrics. We treated 12 different fabric samples with a commercial bleaching product. Bleaching increased luminance and decreased saturation. We presented photographs of the original and bleached samples on a computer screen and asked observers to match the fabric colors to an adjustable matching disk. Different groups of observers produced matches for original and bleached samples. One group of observers were instructed to match the color of the bleached samples as they were before bleaching (i.e., compensate for the effects of bleaching); another, to accurately match color appearance. Observers did compensate significantly for the effects of bleaching when instructed to do so, but not in the appearance match condition. Results of a second experiment suggest that observers achieve color consistency, at least in part, through a strategy based on local spatial differences within the bleached samples. According to the results of a third experiment, these local spatial differences are likely to be the perceptual image cues that allow participants to determine whether a sample is bleached. When the effect of bleaching was limited or uniformly distributed across a sample's surface, observers were uncertain about the bleaching magnitude and seemed to apply cognitive strategies to achieve color consistency.

Introduction

Humans have an outstanding ability to recognize materials and visually perceive their properties across a wide variety of views and illuminations. For instance, the apparent color and lightness of surfaces does not change very much if the surface is moved from indoor to outdoor, even though the spectral composition and the intensity of the proximal stimulus does change. These abilities of the visual system are usually referred to as *color constancy* and *lightness constancy*, respectively. At the computational level, achieving such constancy is difficult because the problem is underdetermined. The properties of the light reflected from a surface depends not only on the reflectance of the surface but also on the color and intensity of the illuminant, the interaction between the geometry of the surface and the spatial structure of the illumination, and the point of view. Furthermore, reflectance is often not uniform across surfaces, for example, surfaces could be textured, painted, or stained. Because the combined effect of these factors determine the properties of the light reaching the eye, their individual contributions are confounded at the input level of the visual system.

There are numerous processes that can lead to changes in the chromatic properties of surfaces. If these chromatic changes follow statistical regularities, we may be able to learn their effects and be able to compensate for them. For example, ageing tends to cause bleaching of fabrics, characterized by color fading and increased lightness due to a series of natural factors like humidity or light exposure (e.g., Lead, 1949; Beek & Heertjes, 1966; Moir, 1971). Because bleaching of clothes is something that typical observers have extensively experienced, it is interesting

Citation: Toscani, M., Milojevic, Z., Fleming, R. W., & Gegenfurtner, K. R. (2020). Color consistency in the appearance of bleached fabrics. *Journal of Vision*, 20(4):11, 1–20, <https://doi.org/10.1167/jov.20.4.11>.



to investigate whether human observers have perceptual knowledge about how ageing or bleaching affect appearance. Yoonessi & Zaidi (2010) have shown that observers are capable of identifying different material classes across natural transformations, such as rusting, burning, or drying. They asked observers to identify the material or the type of transformation. The presence of color information improved performance for both material and transformation judgments. Furthermore, material recognition performance was higher when observers saw two states of the material. This supports previous studies showing benefits of color for material recognition (Sharan, Rosenholtz, & Adelson, 2009; Wiebel, Valsecchi, & Gegenfurtner, 2013). It also suggests that observers may be able to relate the colors of the surfaces during these material changes.

Previous research has shown that observers can determine and ‘discount’ the causal history of objects under some geometric shape transformations (Schmidt & Fleming, 2016; Schmidt & Fleming, 2018; Schmidt, Spröte, & Fleming, 2016; Spröte & Fleming, 2016; Spröte, Schmidt, & Fleming, 2016; Fleming & Schmidt, 2019). For instance, Spröte and Fleming (2016) asked human observers to adjust the bending of a computer-simulated three-dimensional shape to match the degree of bending applied to another object. Participants could match the degree of bend, suggesting that they could infer the causal history that transformed the target shape into its observed state; more recently Schmidt, Philips, and Fleming (2019) showed this generalizes to a wider range of geometric transformations. Schmidt and Fleming (2016) presented their participants with pairs of objects (“before” and “after” a geometric transformation) and asked them to identify points that corresponded across the transformation. Their participants responded accurately and consistently for a broad range of transformations, suggesting participants can identify features across transformations that affect shape.

There has been very little research on the perceptions of transformations that affect color. Sawayama and colleagues (Sawayama, Adelson, & Nishida, 2017) found that modifying the color distributions of photographs of natural textures can alter whether they appear to be wet or dry. Specifically, enhancing chromatic saturation while increasing darkness and glossiness tended to make dry scenes look wet. These changes are consistent with the physical effects of wetting, indicating that the visual system estimates wetness based on knowledge about its characteristic optical effects. As wetting an object does not usually make it appear to be made of a different material, these results suggest that the effects of wetness on the reflected light are attributed to the surface’s “causal history.”

Here we used a color matching paradigm to investigate whether people can discount the effect of bleaching on fabrics. We first produced a series of bleached counterparts of a sample of 12 fabrics of different uniform colors and took pictures of them to use as experimental stimuli. Contemporary fashion has encouraged industries to develop a number of techniques for simulating fabric ageing (for a review, see Del Signore, 2011). Bleaching techniques supposedly mimic the effects of ageing on fabrics, resulting in similar effects on the coloration of different fabrics (Del Signore, 2011), that is, fading of color and increase of lightness (e.g., Mondal & Khan, 2014; Sarkar & Elias Khalil, 2014; Kan, Yuen, & Wong, 2011). Thus studying the perception of bleached fabrics may provide insights into more general color-altering transformations.

We present three experiments here. In Experiment 1, we presented photographs of the original and bleached fabrics on a computer screen, and asked participants to adjust the color of the matching disk to (a) match the color of the original fabrics as accurately as possible; (b) match the color of the bleached fabrics as accurately as possible; and (c) match the color of the bleached fabrics as they were before bleaching (i.e., revert the bleaching process). When observers were instructed to produce accurate matches of the bleached fabrics, the reproduced color crucially depended on the effect of bleaching, that is, the matches were characterized by faded color (lower chroma) and increased lightness. When observers were asked to revert the bleaching process, this dependency was reduced, and matches were similar to the ones produced for the original fabrics. These results suggest that people can compensate for the effect of bleaching on the color of fabrics.

In Experiment 2, we asked participants to mark a number of relatively bleached and nonbleached areas on the bleached fabrics photographs. Although observers’ choices were mostly driven by lightness, compensation effects both on lightness and chroma (observed in the matching experiment) are consistent with a strategy in which participants base their judgments on the color of the selected nonbleached areas.

In Experiment 3, we measured the probability of each image to be categorized as bleached, which we found to depend on the variability in lightness and chroma. This corroborates the idea that observers evaluate bleaching based on local differences within the fabrics surfaces. For the images on which the effects of bleaching were small and probably scarcely visible, observers seemed to overcompensate, presumably because they were highly uncertain about the magnitude of the bleaching effect. This suggests that, at least in part, the compensation processes could be based on cognitive inferences related to our knowledge about the bleaching process.



Figure 1. Samples. (A) Example of one fabric sample in its original (left) and bleached (right) version. Photographs of the flat and crumpled (C) fabric samples. Each sample is presented with the left half taken from the original, and the right half from the photograph of its bleached version. For some of the samples the effect of bleaching looked uniform across the surface (e.g., first two samples from the left on the top row), others appeared as speckled (e.g., third and fourth samples). Few samples were scarcely affected by the treatment (e.g., the second and the fifth sample from the left on the bottom rows).

General methods

Stimuli and apparatus

We selected 12 different homogeneous fabric samples exhibiting differences in color and material (Figure 1), which we presented flat or crumpled, to extend the generality of our investigations to different geometries. For each sample, we created a bleached version by chemically treating the fabrics with the Denkmil Hygienereiniger chlorine-based cleaner (dm Drogeriemarkt, Karlsruhe, Germany). The bleaching affected the materials differently. Some were relatively uniformly affected, whereas others ended up with a speckled appearance (Figures 1B and 1C). Most of the bleached samples looked quite different from their original versions, whereas some of them appeared nearly unchanged.

To visualize the samples on a computer screen for color matching, all the samples were photographed with a Nikon D70 SLR (Nikon, Tokyo, Japan). The Nikon's settings were adjusted to reduce automatic digital processing. To photograph the samples under standard illumination conditions, we placed them in a JUST Normlicht LED box (JUST Normlicht, GmbH; Weilheim, Germany), which provided a stable D65 illumination of 216 cd/m² and CIE xy chromaticity coordinates of (0.31271, 0.32902). The chromatic properties of the illumination were measured by placing

a PR650 RS3 PTFE white reflectance standard (Photo Research, Inc., Syracuse, NY) in the LED box and measuring the light reflected from its surface with a Konica Minolta Spectroradiometer CS-2000A (Konica Minolta Holdings Inc., Marunouchi, Tokyo, Japan).

To rule out the possibility that our results were an artefact of (unknown) color processing operations in the Nikon camera, rather than accurately reflecting the effects of bleaching, we obtained hyperspectral images of the fabric samples. To do so, we used a hyperspectral camera consisting of a mirror-based scanning system with a CCD chip (VNIR HS-CL-30-V8E-OEM; Specim, Spectral Imaging, Ltd., Oulu, Finland). Hyperspectral images were taken under the same conditions as the photographs used in the experiments, although they appeared less sharp, probably because of the limitations of the optics of the hyperspectral camera. Our hyperspectral measuring system is described in detail in our previous work (Ennis, Schiller, Toscani, & Gegenfurtner, 2018). Analyses of the hyperspectral images confirmed the results from the photographs. Because the hyperspectral images tended to be less sharp than the ones taken with the Nikon camera, we used the latter ones as experimental stimuli.

For the experiments, the photographs were displayed on an Eizo CG223W 10-bit LCD monitor (Eizo Nanao Corporation, Hakusan, Ishikawa, Japan), connected to a Dell Precision 380 computer (Dell Inc., Round Rock, TX). The monitor was calibrated according to standard methods (e.g., Hansen & Gegenfurtner, 2013), using a Konica Minolta CS2000-A spectrophotometer (Konica-Minolta Inc., Tokyo, Japan). We measured the following monitor primaries: red primary CIE xyY coordinate (x: 0.6796, y: 0.3073, Y: 28.61 cd/m²), green primary CIE xyY coordinate (x: 0.2062, y: 0.6955, Y: 67.9 cd/m²), and blue primary CIE xyY coordinate (x: 0.153, y: 0.051, Y: 6.21 cd/m²). To compute the color of the photographs as they were presented on the screen, we measured the gamma curve of each of the RGB channels of the monitor and used them to linearize the RGB values of the photographs for colorimetric analyses. Then we converted the linear RGBs to CIE L*a*b* color space, which is the cylindrical representation of the CIE L*a*b* color space (CIE, 1978).

Effect of bleaching

We evaluated the effect of the bleaching procedure on our samples by comparing lightness (L*), chroma (C*), and hue (h*) of the bleached fabrics with the ones of their original versions. To do so, we averaged the lightness, chroma, and hue values for each photograph (excluding pixels outside the textile), and computed differences between the averaged values in the bleached samples and in the original samples. We did this

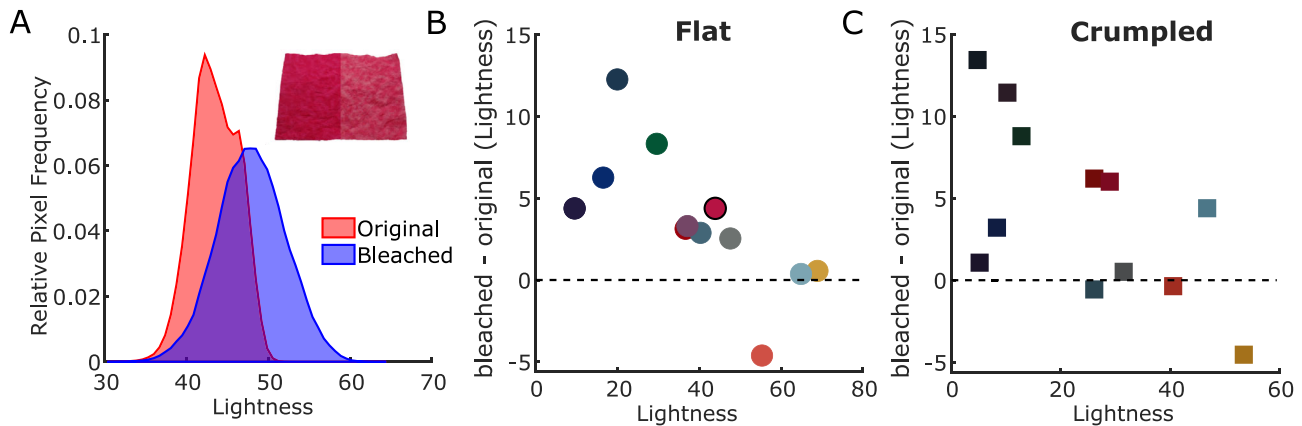


Figure 2. Lightness changes because of the bleaching process. (A) Histogram of the lightness distribution of an example photograph. Lightness on x-axis, relative pixel frequency on y-axis. Lightness distribution of the original version in red, of the bleached version in blue. The example sample is shown on the top right, combining its original (left half) and bleached version (right half). (B–C) Difference between the average lightness of the original samples and their bleached versions, on the y-axis. Average lightness on the x-axis. Each data point indicates one sample (its original and bleached versions); circles represent the flat samples; squares represent the crumpled ones. Data points are represented with the average color of the samples. The black ring indicates the example sample shown in (A), flat samples in (B), crumpled samples in (C).

separately for the flat and crumpled samples. In fact, the illumination of the flat samples is nearly uniform, minimizing the effect of shading. Therefore color differences between the original and the bleached versions are almost only because of the bleaching process. Conversely, the crumpling process creates random differences in shading between the original and bleached samples, which might mask the effects of bleaching. Additionally, we estimated the effect of bleaching for each area of the flat bleached surfaces by geometrically transforming the image of each original sample to maximally overlap with the correspondent bleached version. To do so, we used the MATLAB (version R2017b, <http://www.mathworks.com>) function `imregister()` to determine the affine transformation, which maximizes the overlap between the silhouettes of the images of the bleached and nonbleached samples. This analysis was possible only for the flat samples because the crumpled geometry was different between the bleached and original samples.

The average effect of bleaching on hue was small ($\sim 2^\circ$) and nonsystematic. In fact, a Watson U^2 test (Zar, 1999) failed to show a significant difference in the average hue between the original and the bleached samples ($u = 0.0411$, $p = 0.889$). In addition, we estimated the likelihood of a difference in hue between the bleached and original samples as compared with the likelihood of no hue difference, by means of Bayesian statistics. To do so, we computed the differences between the mean hue of the bleached and original samples (Δh) and transformed them so that each difference Δh was the minimum (in absolute value) between $\Delta h - 2\pi$, Δh , and $\Delta h + 2\pi$. After this transformation, the differences

were rather small and scattered around zero, thus we could approximate them as noncircular and calculate the Bayes factor for a one-sample t -test (following Rouder, Speckman, Sun, Morey, & Iverson, 2009). The likelihood of a difference between hue of the bleached and the original samples was around five times smaller than the one of no hue difference (Bayes factor = 0.22). Hence we focused our analyses on chroma and lightness.

Figure 2 illustrates the changes because of the bleaching process in lightness.

Bleaching tended to shift the lightness distributions toward higher values (example shown in Figure 2A). On average, lightness (Figures 2B and 2C) was higher in the bleached samples than in their original versions, both for the flat ($t(11) = 3$, $p < 0.05$) and the crumpled samples ($t(11) = 2.694$, $p < 0.05$).

Figure 3 illustrates the changes because of the bleaching process in chroma.

For most of the flat samples, bleaching shifted the chroma distributions toward lower values (example shown in Figure 3A). Average chroma (Figure 3B) was significantly lower in the bleached flat samples than in their original versions (chroma: $t(11) = -2.226$, $p < 0.05$). The effect of bleaching on chroma was no longer measurable in the crumpled samples (Figure 3C) ($t(11) = 0.152$, $p = 0.882$), probably masked by the geometric differences between the original and the bleached samples (i.e., differences in shading and occluded regions).

Because bleaching might not have a uniform effect on the surfaces of our samples, we attempted to produce a local estimate of the effect of bleaching on lightness

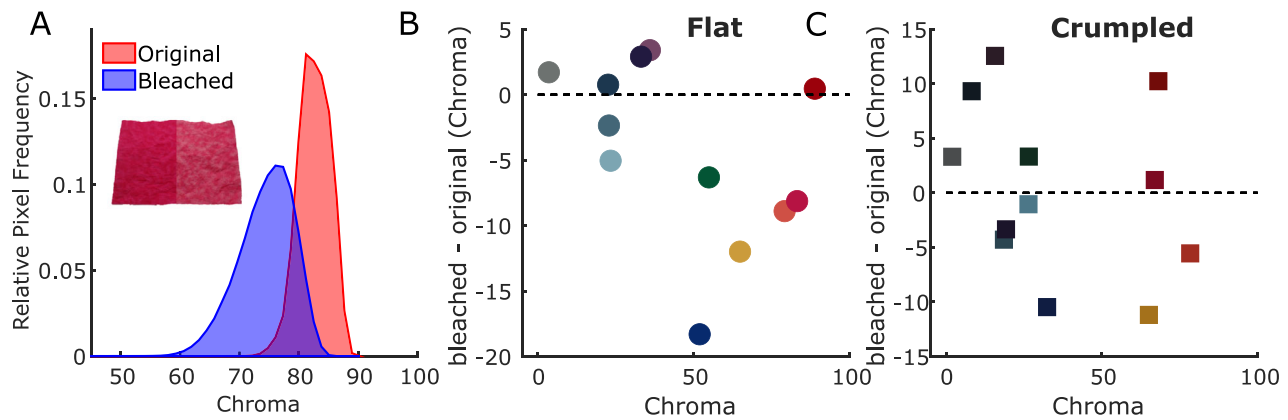


Figure 3. Chroma changes because of the bleaching process. (A) Histogram of the chroma distribution of an example photograph. Chroma on x-axis, relative pixel frequency on y-axis. Chroma distribution of the original version in red, of the bleached version in blue. The example sample is shown on the left, combining its original (left half) and bleached version (right half). (B–C) Difference between the average chroma of the original samples and their bleached versions, on the y-axis. Average chroma on the x-axis. Each data point indicates one sample (its original and bleached versions); circles represent the flat samples; squares represent the crumpled ones. Data points are represented with the average color of the samples. The black ring indicates the example sample shown in (A), flat samples in (B), crumpled samples in (C).

and chroma. To do so, we superimposed the image of each original sample onto the corresponding bleaching sample, and computed the pixel differences between the two images. Because the spatial correspondence between the bleached and nonbleached versions of the fabric samples could not be imposed at the pixel level, the images were low-pass filtered before estimating the effect of bleaching by computing the local differences between the two images.

Figure 4A shows an example of the estimated effect of bleaching on a sample's surface. Bleaching tended to affect large parts of each surface, although its effects were not uniform.

The estimated effects of bleaching for each fabric sample were approximately normally distributed, that is, the correlation between the quantiles of their distributions and the corresponding quantiles of the standard normal distribution were remarkably high (Pearsons' r range: [0.945 1]). Thus the mean is an appropriate estimator of the effect of bleaching on lightness and chroma.

Figure 4B shows the mean effect of bleaching for each sample. On average, bleaching increased lightness for all the samples but one; the mean effect of bleaching was significantly different from zero ($t(11) = 3$, $p < 0.05$). Concerning chroma, seven samples exhibited a negative mean effect of bleaching, whereas the effect was positive but close to zero for the remaining samples. On average, we observed a trend for the mean effect of bleaching on chroma to be significantly smaller than zero ($t(11) = -2.2$, $p = 0.05$).

Overall, the chromatic changes we observed are similar to what bleaching and ageing processes are

supposed to cause in fabrics, that is, fading of color (decreased chroma) and increase in lightness (increased lightness) (Kan et al., 2011; Mondal & Khan, 2014; Sarkar & Elias Khalil, 2014). The analyses of the photographs were based on their colors as they were displayed on the experimental screen. Because the results of our analyses on the photographs of the fabric samples were confirmed by similar analyses on the hyperspectral images of these same samples, we are confident that our participants were presented with realistic renderings of the original and the bleached fabrics, at least as far as the effects of bleaching are concerned.

Experiment 1

Experiment 1 was a color matching task. Observers had to adjust the color of a disk stimulus to match the color of photographs of the fabric samples. Both matching disk and photographs were presented on a computer screen. When matching the photographs of the original samples, they were instructed to match the color of the fabrics as precisely as possible. When matching the color of the bleached fabrics, they were presented with two different instructions: (a) to match the color of the fabrics as accurately as possible, or (b) to match the color of the fabrics as they were before bleaching. The comparison between these two tasks allowed us to investigate whether participants are able to compensate for the color changes caused by the bleaching process. Crucially, different groups of

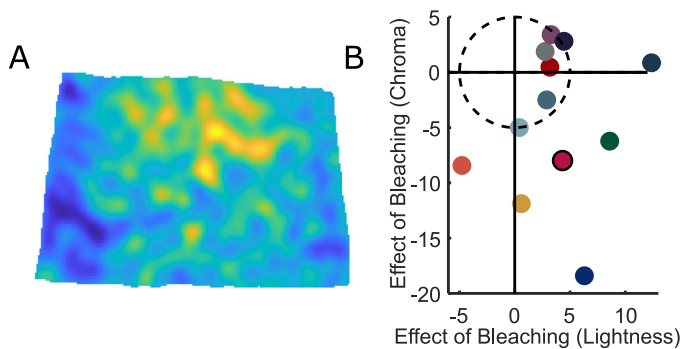


Figure 4. Estimated effect of bleaching. (A) Effect across one example surface, combined for lightness and chroma. The effect of bleaching is represented in false-colors, from blue to yellow. Blue means zero change owing to bleaching; yellow represents the highest change. For each pixel, the effects on lightness and chroma are combined by projecting the corresponding two-dimensional point in the lightness versus chroma space onto the line bisecting the second quadrant (as individuated by: $\text{chroma} = -\text{lightness}$). (B) Mean effect of bleaching for lightness (x-axis) and chroma (y-axis). Each data point represents the mean effect of bleaching on one sample. Different colors correspond to the color of the different samples. The dashed circle surrounds the area within 5 Δ units radius. The effect of bleaching within this area is likely to be scarcely perceivable when the bleached and the original version of a sample are not presented next to each other. The circle with the black ring indicates the fabric shown in (A).

observers matched pictures of the original fabrics and the bleached fabrics under the two instructions. Hence observers who saw the photographs of the bleached fabrics had no direct knowledge of how these particular fabrics would have looked if they were not bleached. Also, observers' impression of the original fabrics could not have been influenced by the experience of their bleached versions in any way.

Methods

Participants

Twenty-seven observers participated in [Experiment 1](#). All participants were naive to the purpose of the experiment. They all had normal or corrected to normal visual acuity, as well as normal color vision according to the Ishihara color plates, 24-Plate Edition ([Ishihara, 2004](#)). Observers were compensated for their participation in the experiments. All observers gave written informed consent in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The experiments were approved by the local ethics committee (approval number LEK 2009-0008).

Experimental instructions

The color matching instruction for the photographs of the original (nonbleached) samples was always to reproduce their color as accurately as possible (match original samples: original condition). Concerning the bleached samples, two different matching instructions were given to different groups of participants. They were asked either to reproduce their color as accurately as possible (appearance condition), or to reproduce their color as it was before bleaching (i.e., to compensate for bleaching: compensation condition). Thus participants were divided in three groups of nine people each, according to the type of instruction.

Stimuli

Photographs of the original and bleached samples were presented on the left side of a computer screen. The left border of each photograph was presented at 60.7° from the center of the screen; its top border was 12.6° from the upper edge of the screen. The photographs of the flat samples measured $22.5^\circ \times 12.6^\circ$, whereas the ones of the crumpled samples had smaller width ($18.7^\circ \times 12.6^\circ$).

Procedure

Observers sat 38 cm from the screen with their chins supported by a chin rest. They produced color matches by adjusting the color of a 13° visual angle disk, which was presented on the right side of the screen (vertically centered, horizontally displaced 17.8° to the left from the center). Adjustments were done in CIE $L^*C^*h^*$ color space. As noted before, this space is the cylindrical representation of the CIE $L^*a^*b^*$, which is designed to be perceptually uniform. Observers could adjust hue, chroma, and lightness. In all of the experiments, the observers provided three matches for each of the 24 images (12 flat and 12 crumpled), resulting in a total of 72 matches. Matching instructions were given at the beginning of each session.

Analyses

The three repetitions for each image were averaged for the analyses, yielding a triplet $L^*C^*h^*$ for each image for each participant. Because bleaching mostly affected lightness and chroma, we focused our analyses on these two dimensions, which we treated separately for most of the analyses. We first compared the average matches with the mean values of each fabric. We did this to investigate potential biases with respect to the central parts of the color distributions of our stimuli, as often reported in the color matching literature for chromatic textures ([Kuriki, 2004](#); [Sunaga & Yamashita, 2007](#)) and natural stimuli ([Giesel & Gegenfurtner, 2010](#);

Toscani, Valsecchi, & Gegenfurtner, 2013a). Then, we assessed the effect of the instruction (appearance vs. compensation) on the matches, to investigate whether observers could compensate for the effect of bleaching. To do so, we represented the effect of bleaching on color matches (perceptual difference) by expressing the matches in the appearance and compensation conditions as a difference from the average matches (across participants) in the original condition. For each sample, we represented the physical effect of bleaching on the lightness and chroma distributions as the mean of the estimated effect of bleaching obtained by superimposing the images of the nonbleached samples on the one of the bleached samples (as described earlier in effect of bleaching). Then we regressed the perceptual difference computed for each sample as a function of the effect of bleaching for the corresponding samples. A regression slope of one would suggest no compensation: a given change in the lightness or chroma distribution owing to bleaching yields a corresponding change in the matches. A slope of zero means perfect compensation: observers match the bleached surfaces as the nonbleached ones, despite the changes owing to bleaching. The average slopes were compared across instruction conditions and for different geometries (flat vs. crumpled samples). Crucially, a difference in the slopes between the appearance and the compensation conditions would imply compensation. This is an intuitive way to test for an effect of instructions on the color matches for the bleached samples, however, it does not combine the effects on lightness and chroma into a single measure and is not sensitive to the potential differences between samples.

To assess compensation (for lightness and chroma together) for the individual samples separately, we defined a compensation index based on the difference between the matches of the bleached (for the appearance or compensation conditions) and the original samples (original condition), relative to the estimated colorimetric effect of bleaching. Specifically, we computed the vector difference between the matches for the bleached and the original samples (BO), in the lightness versus chroma space. We projected this vector onto the vector identified by the average colorimetric effect of bleaching for that sample (CBE). This projection was then normalized by the norm of CBE, subtracted from one and expressed as a percentage. If the color match for a bleached sample were the same as for its original version—that is, despite the effect of bleaching the observer perfectly reproduced the color of the original sample—then the norm of BO would be zero, thus compensation index values 100%. Conversely, if BO were equal to CBE, the compensation index would be zero, that is, the shift of the color match for the bleached sample with respect to the match of its original version is equal to the colorimetric effect of bleaching, indicating no compensation. The

compensation index was computed separately for each sample, each of the appearance and compensation conditions, and each participant.

Results

Matching results: potential color biases

Figure 5 represents the lightness matching results for three example samples (panel A), and averaged across samples (panel B). Average lightness matches are presented for the original samples and their bleached versions for the two instructions (appearance and compensation), together with the lightness distributions computed from the photographs of the samples.

When observers were asked to match the appearance of original and bleached samples (original and appearance conditions), lightness matches were higher than the mean value of the corresponding distributions for all the flat and crumpled samples. To test the statistical significance of this observation, after averaging across samples, we ran a *t*-test against the null hypothesis that the lightness matching results are on average equal to the mean of the lightness distribution of the samples separately for each matching condition and geometry (flat vs. crumpled). With respect to the original samples, lightness matches were on average significantly higher (13% and 17%, respectively for the flat and the crumpled samples) than the mean of the lightness distributions ($t(8) = 5.824$, $t(8) = 11.488$, for the flat and the crumpled samples; $p < 0.0005$). The same is true for the appearance condition ($t(8) = 7.721$, $t(8) = 12.17$, for the flat and the crumpled samples; $p < 0.0005$). This suggests that observers based their matches on the above mean portions of the lightness distributions of the stimuli, consistent with previous findings on color and lightness matching with realistic objects (Giesel & Gegenfurtner, 2010; Toscani et al., 2013a; Toscani, Valsecchi, & Gegenfurtner, 2017; Toscani, Zdravković, & Gegenfurtner, 2016). However, when observers were asked to compensate for bleaching (compensation condition), they produced matches closer to the averages of the physical distributions, as confirmed by a *t*-test comparing the matches in the compensation and in the appearance conditions ($t(16) = 4.166$, $p < 0.05$; $t(16) = 2.1357$, $p < 0.05$; for the flat and the crumpled samples, respectively). This suggests that, to some extent, in the compensation condition observers were able to discount the effect of bleaching on lightness. Notably, for some of the samples the lightness matches are much lower than the mean of their lightness distribution (e.g., Figure 5A, third example from the left).

Figure 6 shows a similar pattern for chroma: observers set above mean values in the original (23% and 54% higher; $t(8) = 7.56$, $p < 0.05$, $t(8) = 12.812$,

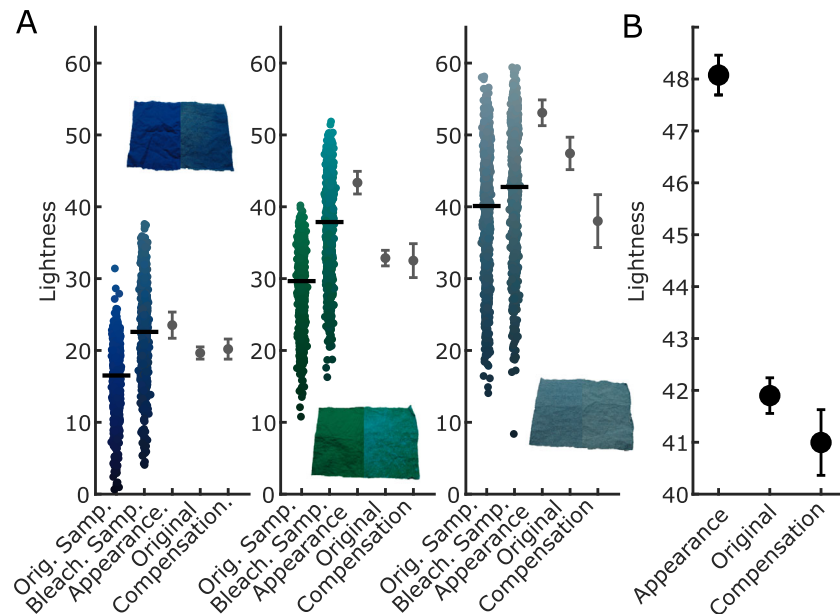


Figure 5. Lightness matching results. (A) Results for three representative samples in their flat versions. Each subpanel represents the average matches and color distributions of the photographs of the bleached and the original samples. On x-axis: lightness distribution of the original (orig. samp.) or bleached (bleach. samp.) sample, matches for the appearance, Original or compensation conditions, from left to right, respectively. Lightness on y-axis. The distributions are represented by colored dots, each representing the corresponding value of each pixel from a random sample of 1000. Individual dots are scattered horizontally by a random jitter to reduce overlapping. Horizontal black lines on the colored dots represent the mean of the distributions. The gray dots represent the mean matches, with the error bars being the standard error of the mean. (B) Average matches pooled across images for the three conditions. Error bars represent the standard error of the mean across participants.

$p < 0.05$; for the flat and the crumpled samples, respectively) and appearance condition (27% and 47% higher; $t(8) = 7.079$, $p < 0.05$, $t(8) = 8.29$, $p < 0.05$; for the flat and the crumpled samples, respectively). In the compensation condition, chroma matches were closer to the mean of the sample distributions, although a t -test failed to show a difference with the matches in the appearance condition. Again, for some of the samples, the chroma matches were much lower than the mean of their lightness distribution (e.g., Figure 5A, third example from the left).

Effect of instructions

For both the appearance and compensation conditions, we computed linear regression slopes as a measure of compensation (see Analyses section), separately for lightness and chroma, and for flat and crumpled samples.

Figure 7 illustrates the regression results for lightness. In the appearance condition, the regression slopes for the flat (Figure 7A, blue) and the crumpled (Figure 7B, blue) samples approach the unity line (black dashed line). Thus when observers simply matched the color of the bleached samples, the difference with the lightness matches for the corresponding original samples was on

average largely predicted by the effect of bleaching on the lightness distributions of the fabric samples.

Conversely, when observers were instructed to match the color of the bleached samples as it was before bleaching (compensation condition), the slope of the regression lines (red lines) is reduced (Figures 7A and 7B) for both flat and crumpled samples. This suggests that observers compensated for the effect of bleaching, at least to some extent.

Figure 7C shows the regression slopes, computed separately for each observer and then averaged across observers. Because a regression slope equal to one would mean no compensation, in the illustration the average slopes are subtracted from one, so that values close to one indicate high compensation and close to zero indicate little compensation. Compensation is higher in the compensation condition, suggesting an effect of instructions.

Again, results for chroma were similar to what we found for lightness, as illustrated by Figure 8. In the appearance condition, the regression slopes approach the unity line and are reduced in the compensation condition.

We tested the differences in the regression slopes between conditions with two-way mixed analysis of variance (ANOVA), separately for lightness and chroma, each observation being the slope computed

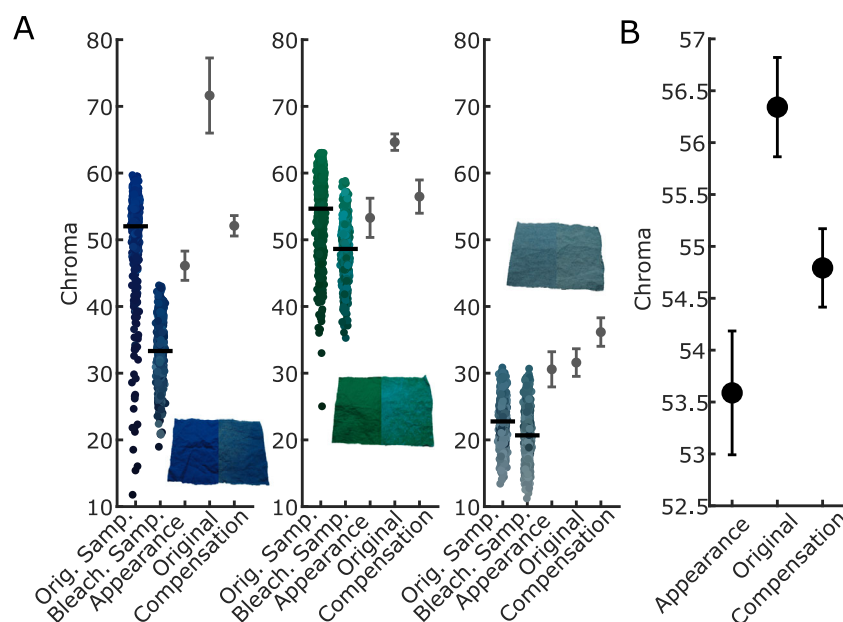


Figure 6. Chroma matching results. (A) Results for three representative samples in their flat versions. Each subpanel represents the average matches and color distributions of the photographs of the bleached and the original samples. On x-axis: chroma distribution of the original (orig. samp.) or bleached (bleach. samp.) sample, matches for the appearance, original or compensation conditions, from left to right, respectively. Chroma on y-axis. The distributions are represented by colored dots, each representing the corresponding value of each pixel from a random sample of 1000. Individual dots are scattered horizontally by a random jitter to reduce overlapping. Horizontal black lines on the colored dots represent the mean of the distributions. The gray dots represent the mean matches, with the error bars being the standard error of the mean. (B) Average matches pooled across images for the three conditions. Error bars represent the standard error of the mean across participants.

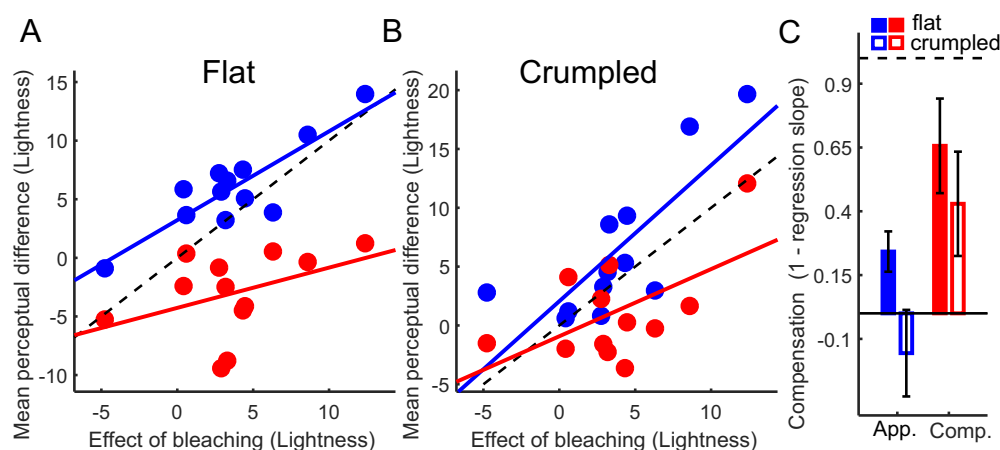


Figure 7. Regression results: lightness. (A–B) Perceptual differences averaged across participants (y-axis) as a function the effect of bleaching (x-axis). Blue circles represent data points from the appearance condition, whereas red circles for the compensation condition. (A) Represents data for the flat samples, (B) for the crumpled samples. Slopes are computed with the perceptual difference averaged across participants. (C) Regression slopes, computed for each participant and averaged across them (y-axis), for the appearance (app. – blue bars) and compensation (comp. – red bars) conditions, for the flat (filled bars) and the crumpled samples (empty bars). Error bars represent the standard error of the mean computed across participants.

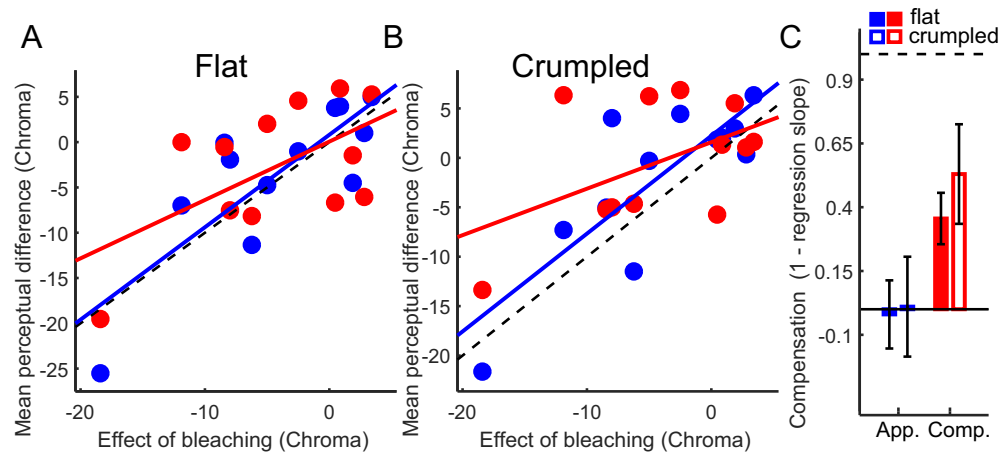


Figure 8. Regression results: chroma. (A–B) Perceptual differences averaged across participants (y-axis) as a function the effect of bleaching (x-axis). Blue circles represent data points from the appearance condition, whereas red circles for the compensation condition. (A) Represents data for the flat samples, (B) for the crumpled samples. Slopes are computed with the perceptual difference averaged across participants. (C) Regression slopes, computed for each participant and averaged across them (y-axis), for the appearance (app. – blue bars) and compensation (comp. – red bars) conditions, for the flat (filled bars) and the crumpled samples (empty bars). Error bars represent the standard error of the mean computed across participants.

A) ANOVAs on slopes computed with the effect of bleaching

Lightness						Chroma				
Source	SS	df	MS	F	p	SS	df	MS	F	p
Instructions	2.248	1	2.248	5.67	0.03*	1.804	1	1.804	4.6	0.048*
Error	6.343	16	0.396			6.274	16	0.392		
Geometry	0.88	1	0.88	8.435	0.01*	0.094	1	0.094	1.213	0.287
Instructions: geometry	0.066	1	0.066	0.636	0.437	0.046	1	0.046	0.599	0.45
Error	1.652	16	0.103			1.242	16	0.078		

B) ANOVAs on slopes computed with the differences between means of distributions

Lightness						Chroma				
Source	SS	df	MS	F	p	SS	df	MS	F	p
Instructions	1.935	1	1.935	7.07	0.017*	2.038	1	2.038	9.287	0.008*
Error	4.379	16	0.274			3.511	16	0.219		
Geometry	0.041	1	0.041	0.317	0.582	2.121	1	2.121	25.209	0.0001*
Instructions: geometry	0.014	1	0.014	0.105	0.75	0.096	1	0.096	1.14	0.302
Error	2.085	16	0.13			1.346	16	0.084		

Table 1. ANOVA table (regression slopes). Two-way mixed ANOVAs with geometry (spread out vs. crumpled) as a within-subjects factor, and instruction (bleached appearance vs. bleached original) as a between-subjects factor. ANOVAs are computed separately for lightness and chroma. Two ANOVAs are computed on the slope as dependent variable, by regressing the perceptual difference of each sample as a function of the effect of bleaching for the corresponding sample. Notes: “Source” indicates factors, errors, and interaction, “SS” their sum of squares, “df” the degrees of freedom, “MS” the mean of the sum of squares, “F”, the F statistics, and “p” the corresponding p value. Tests are significant with $p < 0.05$ (*).

for each participant within the fixed factors: geometry (flat vs. crumpled within subjects) and instruction (appearance vs. compensation between subjects).

ANOVAs (Table 1, panel A) revealed a significant main effect of instruction for both lightness and chroma. Concerning chroma, geometry had no significant effect on the slopes, but slopes were significantly steeper for the crumpled samples when computed based on lightness. The interactions between

instruction and geometry were not significant for either lightness or for chroma.

Regressions were computed based on our estimate of the effect of bleaching, obtained by transforming and superimposing the bleached and original versions of each sample. To check whether our results depended on the way we estimated the effect of bleaching, we repeated these analyses with the slope computed based on a simpler estimate of the effect of bleaching: the

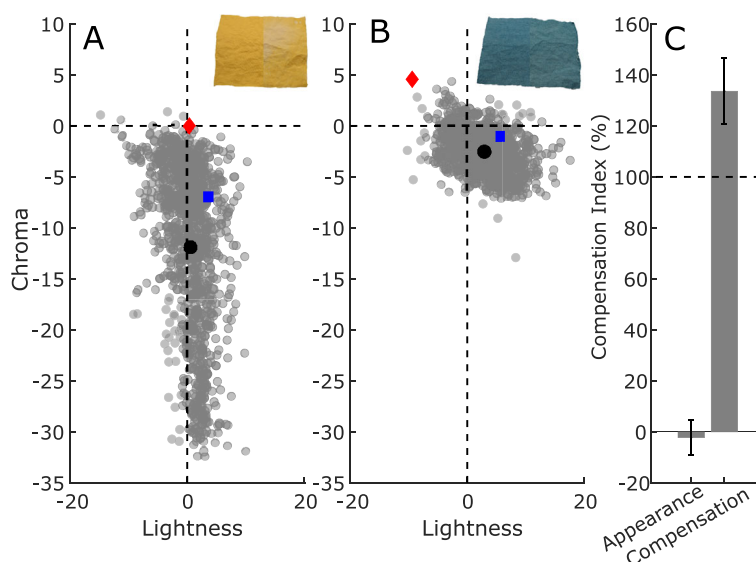


Figure 9. Compensation index. (A–B) Two examples. Lightness on the x-axis; chroma on the y-axis. Gray points represent the physical effect of bleaching distribution and the black circles its mean. Blue squares represent the difference between the matches for the bleached sample and its original version in the appearance condition, red diamonds in the compensation condition. (C) Average compensation index across images, averaged across participants, for the two instruction conditions (appearance and compensation). The error bars represent the standard error of the mean. The dashed line represents perfect (100%) compensation.

difference between the average of the distribution of the image of each bleached fabric and the average of the corresponding nonbleached version. Concerning the effect of instruction, results were equivalent (Table 1, panel B). In contrast, these analyses failed to reveal an effect of geometry on the slopes computed on lightness, but showed that the slopes computed on chroma tended to be steeper for the flat samples than for the crumpled ones.

Compensation for the individual samples

Figures 9A and 9B illustrate the matching results for the bleached samples (appearance and compensation conditions) expressed as a difference from the matches for their original versions, together with the distribution of the effect of bleaching, in the lightness versus chroma space. Two examples are shown. Bleaching increased lightness and decreased chroma, as indicated by the distribution of the effect of bleaching being scattered prevalently in the second quadrant (gray circles). In the appearance condition, the difference between the matches for the bleached samples with the matches for their original versions tends to lie within the distribution to effect of bleaching (Figures 9A and 9B, blue squares). For some samples (e.g., Figure 9A), in the compensation condition, this difference is close to zero (red diamond), indicating that the matches for the bleach samples were very similar to the ones for their original versions, that is, high compensation. However, for other samples (e.g., Figure 9B), this difference between the matches for the bleached and

the original samples is shifted away from the effect of bleaching distribution, in the opposite direction of the mean effect of bleaching, suggesting overcompensation. These samples seem characterized by a limited effect of bleaching (for example, the distance of mean effect of bleaching—black circles—from zero, is much larger in Figures 9A than B).

However, in the compensation condition, the compensation index averaged across samples (see Analyses section) is different from zero but not from 100% (i.e., perfect compensation), as indicated by the confidence interval (CI) of its mean across observers (95% CI [84.41%, 184.83%]). In the appearance condition, the compensation index is not significantly different from zero (95% CI [−29.44%, 24.9%]).

Discussion

These results indicate that when observers were asked to reproduce the original colors of bleached fabrics, they could discount the effects of bleaching both on lightness and chroma to a large extent. The question remains which strategy they were exploiting. Observers could make use of their knowledge about the bleaching process to revert it and retrieve the original colors. Alternatively, they could simply base their judgments on the regions of the bleached samples that they considered to be least affected by bleaching. In this case, to achieve compensation, observers must be able to parse the bleached surfaces into areas that

are more or less strongly affected by the bleaching. We investigated this possibility in a second experiment.

Experiment 2

We aimed to investigate whether observers could individuate more or less bleached areas on the surfaces so that the more bleached areas were characterized by color changes determined by the bleaching process, that is, increased lightness and decreased chroma.

Methods

Participants

Six observers participated in [Experiment 2](#). All participants were naive to the purpose of the experiment. They all had normal or corrected to normal visual acuity, as well as normal color vision according to the Ishihara color plates, 24-Plate Edition ([Ishihara, 2004](#)). Observers were compensated for their participation in the experiments. All observers gave written informed consent in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The experiments were approved by the local ethics committee (LEK 2017-0030).

Stimuli

We used the same photographs as in [Experiment 1](#), presented on the computer screen with the same size and viewing distance.

Procedure

Observers were asked to indicate 10 different locations that appeared particularly strongly bleached, as compared with the whole surface, and 10 appearing particularly spared by the bleaching process. Observers made their selection with the mouse button. During the selection process, a message was displayed horizontally on the top of the screen to inform whether they were to select bleached or nonbleached areas. The bleached areas were always selected first. When the observer made a selection, the selected area was marked with a colored dot (red or green for the bleached or nonbleached areas, respectively; see [Figure 7A](#)).

Analyses

We used a linear classification to assess which color properties of the selected areas could discriminate between bleached and nonbleached selections. For each

selected point we compute the color (in $L^*a^*b^*$ color space) of the pixels in the surrounding area defined by $\approx 0.2^\circ$ of visual angle. We then transformed the $L^*a^*b^*$ coordinates into their polar representation ($L^*C^*h^*$) and averaged across pixels, so that for each selected area we had a single value of L^* (lightness); a^* , b^* , C^* (chroma); and h^* (hue). For each of these color dimensions, the 20 averages (10 selections for the bleached and 10 for the nonbleached areas) for each image were z-transformed, so that they expressed the position of the selected area relative to each particular image. The transformation of the color values of the selected areas for each stimulus was done using the standard deviation and mean of the color distribution of that stimulus. Because of this transformation, predictors from different images were comparable, and thus could be treated together in the analyses as a predictor for the linear classifier. Classification was done for each individual color statistic and for all their combinations. The classifier was trained on the selections from the whole set of stimuli, leaving out the 20 selections from one stimulus at a time. The excluded selections were then tested with the leave-one-out classifier. All classification analyses were done with the linear discriminant analysis routines implemented in the *classify()* function of the statistics toolbox for MATLAB (version R2017b, <http://www.mathworks.com>). Because predictors might be correlated, the classification performance of each single predictor may not reflect its actual individual contribution. Thus the contribution of each predictor was determined by computing the classification performance for all the combinations of predictors, including that predictor and averaging across them, computing performance for all the combinations excluding that predictor and averaging across them, and finally computing the difference between these averages (for details about this procedure, see [Wiebel, Toscani, & Gegenfurtner, 2015](#)). Individual contributions were computed for each participant, and then averaged across participants. One-way repeated-measure ANOVA tested for a difference in the individual contribution between predictors. Additionally, as a direct test of the hypothesis that compensation in the compensation condition was based on local differences within the bleached samples, we used the selections of the nonbleached areas to predict the matches of [Experiment 1](#). In particular, we computed the least bleaching effect as the effect of bleaching for the selected nonbleached areas, as estimated from the superimposition of the images of the original samples on the corresponding bleached samples. This was done for each selection, and then averaged across selections, for each of the flat samples. If the matches in the compensation condition are based on local information (presumably based on the portions that appear least bleached), then the least bleaching effect should be

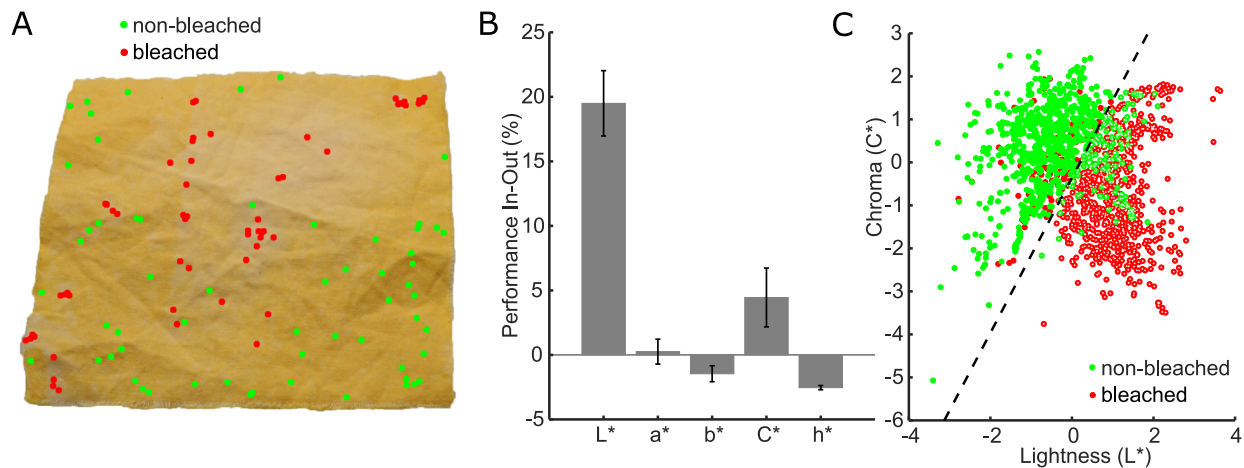


Figure 10. Selection of bleached and nonbleached areas. (A) Selected bleached (red dots) and nonbleached areas (green dots) for all the six participants of [Experiment 2](#). (B) Individual contribution (on y-axis) of each color dimension (x-axis). The bars indicate the mean across participants; the error bars represent the standard error of the mean. (C) Classification based on lightness (x-axis) and chroma (y-axis). Red circles represent the selected bleached areas; green circles the nonbleached ones. The red dashed line represents the classification border produced by the classifier trained on all the selections at once. Empty circles represent selected areas classified as bleached and filled circles as nonbleached.

predictive of the difference between the matches in the compensation condition and the ones in the original condition (matching difference). Thus we computed linear correlations between the least bleaching effect for the 12 flat samples and the corresponding matching difference. Correlations were computed separately for lightness and chroma. Because visual inspection of the data suggested the presence of outliers that could potentially create correlations, we used a robust correlation measure, which is resilient to the effect of outliers ([Pernet, Wilcox, & Rousselet, 2013](#)).

Results

[Figure 10B](#) shows the individual contributions of each predictor. Classification is mostly based on lightness differences. One-way repeated-measure ANOVA revealed a difference in performance between predictors, $F(4,20) = 28.577$, $p < 0.05$. Tukey post hoc comparisons indicated that performance for lightness was higher than for each of the other predictors (all p values < 0.05). [Figure 10C](#) illustrates the classification of the selected areas (bleached and nonbleached, red and green dots, respectively) by means of the combination of the two best performing predictors, that is, lightness and chroma, suggesting again that classification is mostly driven by difference in lightness. In fact, the individual contributions of chroma did not result to be significantly different from zero ($t(5) = 1.953$, $p = 0.108$). This could be explained by limited effect of bleaching on chroma.

However, regression results show that the differences between the matches in the compensation and in the original condition (matching difference) are predictable from the results of [Experiment 2](#). The physical effect of bleaching in the areas that were perceived as being least affected by the bleaching (least bleaching effect) was correlated with the matching difference, both for lightness ([Figure 11A](#)) and chroma ([Figure 11B](#)). Specifically, the correlation coefficients computed for each observer were on average ([Figure 11C](#)) higher than zero ($t(5) = 4.822$, $p < 0.05$; $t(5) = 6.46$, $p < 0.05$; for lightness and chroma, respectively).

Taken together, the results suggest that selecting nonbleached areas on the basis of lightness could explain the compensation results also for chroma, presumably because the effects of bleaching on lightness and chroma correlated across the surface of the fabric samples. Thus it is reasonable to suppose that to retrieve the color of the nonbleached versions of the bleached samples, participants in the compensation condition of [Experiment 1](#) based their matches on the areas of the samples, which looked as unaffected by the bleach as possible (i.e., relatively low lightness). To do so, bleached regions must be visible on the surfaces. However, for some surfaces, the effect of bleaching was limited and probably hard to notice (effect of bleaching within the dashed circle in [Figure 4B](#)). Therefore if compensation happens because observers based their matches on the chromatic differences within each image, for these images it should be relatively poor. We tested this idea with further analyses.

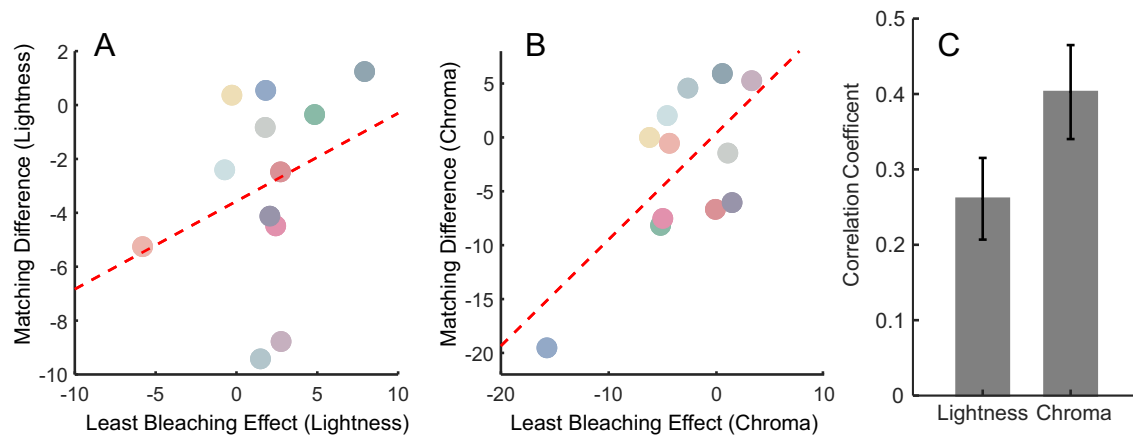


Figure 11. Matching difference as a function of least bleaching effect for lightness (A) and chroma (B). (A–B) Matching difference for each image, averaged across participants, on y-axis. Least bleaching effect for each image, on x-axis. Different colors represent the average colors of the 12 flat samples. The red dashed line is the robust regression line computed for the data points in the figure. (C) Correlation coefficients averaged across observers (y-axis), for correlations computed on lightness and chroma (x-axis). Error bars represent the standard error of the mean.

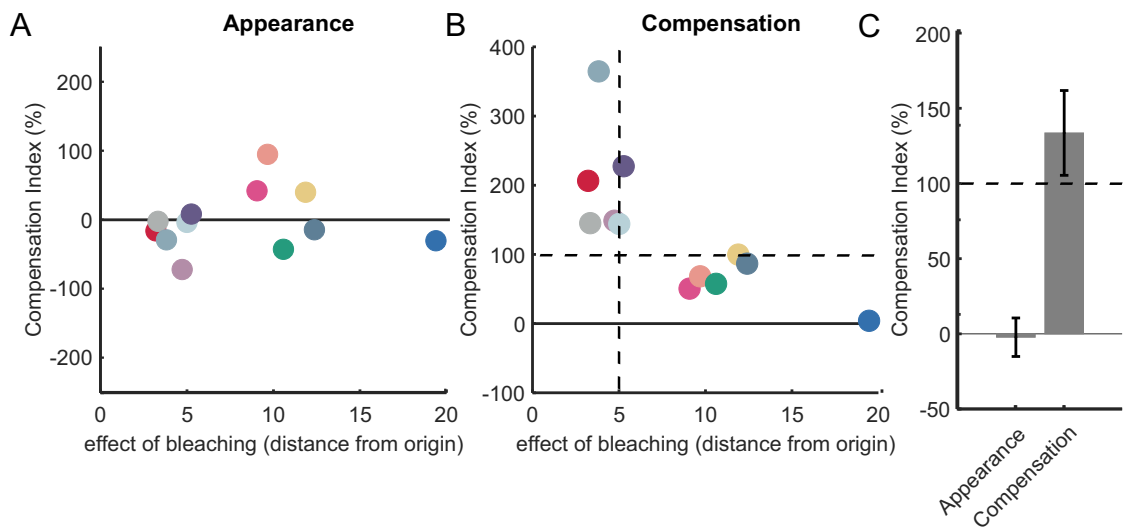


Figure 12. Compensation and effect of bleaching. (A–B) Compensation index (averaged across observers, on y-axis) as a function of the mean effect of bleaching (x-axis). The color of each data point represents the average color of the corresponding sample. Compensation index for the appearance condition in panel (A), for the compensation condition in panel (B). (C) Compensation index averaged over images and observers. Error bars represent the standard error of the mean across observers. The horizontal dashed lines represent perfect compensation. The vertical dashed line represents the bleaching effect of 5 Δ units.

Compensation and effect of bleaching

We tested the hypothesis that compensation was poor for the surfaces on which the effect of bleaching was limited and probably hard to notice because observers could not base their matches on the chromatic differences within each image. For each sample, we computed the mean effect of bleaching and used it to predict the compensation index. Figures 12A and 12B show the compensation index plotted as a function of the mean effect of bleaching. For the appearance

condition (Figure 12A), compensation was generally low and not related to the size of the effect of bleaching. For the compensation condition (Figure 12B), the compensation index seems to depend on the mean effect of bleaching.

Figure 12B shows that in the compensation condition, when the effect of bleaching was rather small (i.e., approximately within 5 Δ units – vertical dashed line; Figure 12B), observers tended to overcompensate, whereas a moderate degree of compensation is associated with samples more affected by bleaching.

For only one sample, compensation is close to zero, although bleaching clearly affected its color. However, after averaging across samples, the mean effect of bleaching across participants was higher than zero (134%; [Figure 12C](#), $t(11) = 4.74$, $p < 0.01$).

We speculate that, when local differences within the stimuli were visible, observers compensated based on local differences within the bleached samples. When bleaching was limited and scarcely visible on the surface, observers changed their compensation strategy and based their color matches on their knowledge about the bleaching process, that is, they set their matches to arbitrarily lower lightness and higher chroma, resulting in overcompensation. This probably represents cognitive inference encouraged by our instructions. In fact, in the compensation condition of [Experiment 1](#), participants were told that they were presented with photographs of bleached images, irrespectively of whether or not these images presented any cue of the bleaching process. To test this possibility, that is, that the images for which participants overcompensated are not perceived as bleached, we ran a third experiment in which observers had to decide whether each sample presented in isolation was in its original or bleached version.

Experiment 3

We presented participants with the full stimulus set and asked them to classify each image as original or bleached. For the bleached samples, we related the probability of being classified as bleached to the estimated average effect of bleaching. Furthermore, we investigated whether classification performance could be explained by image cues, that is, whether we could find an image statistic that is diagnostic for bleaching.

Methods

Participants

Twenty observers participated in [Experiment 3](#); all were naive to the purpose of the experiment. They all had normal or corrected to normal visual acuity, as well as normal color vision according to the Ishihara color plates, 24-Plate Edition ([Ishihara, 2004](#)). Observers were compensated for their participation in the experiments. All observers gave written informed consent in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The experiments were approved by the local ethics committee (LEK 2017-0030).

Stimuli

We used the same photographs as in [Experiment 1](#), presented in the center of the computer screen with the same size and viewing distance as in the previous experiments. In addition, near to the bottom left corner of the screen we presented the word “original,” and “bleached” on the opposite side (bottom right).

Procedure

Observers were told that some of the images were artificially bleached and asked to indicate whether each sample was one of those. Observers made their choice by a mouse click on the “original” or on the “bleached” text, in the left or right bottom corner of the screen, respectively. The images were shown in random order, and only one choice was made per image.

Results

[Figure 13](#) shows observers’ performance in telling apart bleached and original samples. For each image, we computed the probability across participants to be judged as bleached and to be properly classified. On average, observers could tell apart bleached and original samples better than chance ([Figure 13A](#), $t(19) = 7.86$, 13.37, 13.6, 7.43; p values < 0.001 ; for the original, flat, and crumpled, and the bleached, flat, and crumpled samples, respectively).

The probability of a bleached image to be judged as bleached seems related to the average effect of bleaching ([Figure 13B](#)). The samples with an estimated average effect of bleaching smaller than 5 Δ units (vertical dashed line in [Figure 13B](#), dashed circle in [Figure 4](#)) are not judged as bleached better than chance level, and one of them is consistently judged as original, indicating that for such images bleaching was scarcely visible. [Figure 13C](#) shows that overcompensation was found for images for which the effect of bleaching was not clearly visible, confirming our speculation based on Δ units.

Our estimate of the effect of bleaching is based on the original images as ground truth, the question remains which aspects of the bleached images alone make this effect visible.

Lightness and chroma variability as perceptual signature of bleaching

We argued that, when bleaching was visible, compensation was based on the spatial chromatic variations within each image. The bleaching process indeed left stains on many of our samples. We hypothesized that observers could see the effect of bleaching by spotting those stains (and using the color

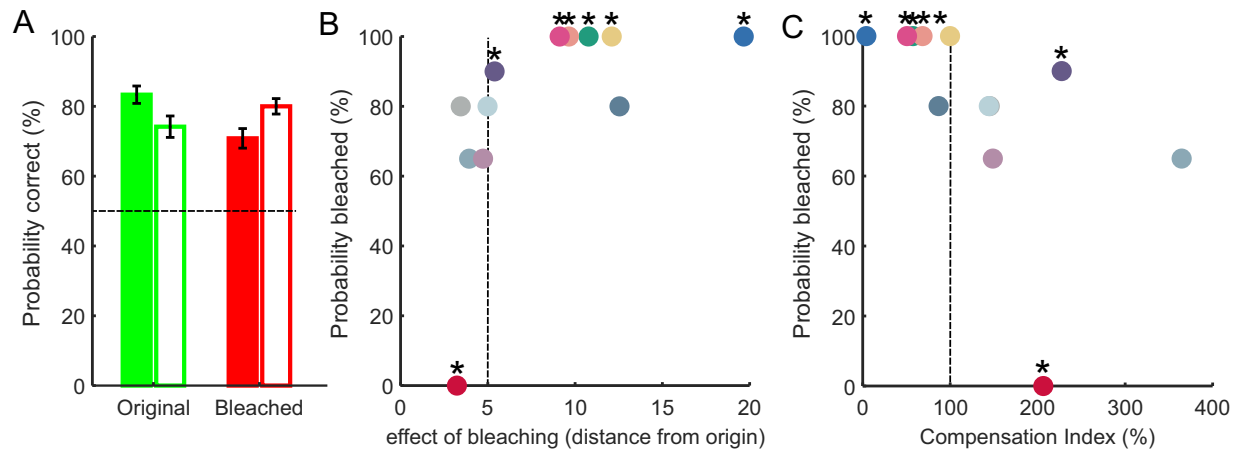


Figure 13. Observers' classification performance. (A) Probability of correct classification, averaged across images and across observers on the y-axis, expressed in percentage. Green bars indicate results for the original images, red for the bleached ones; filled bars for the flat samples, open bars for the crumpled samples. Error bars are the standard errors of the mean across participants. The dashed black line represents the nominal chance level (50%). (B) Probability of being classified as bleached as a function of the average effect of bleaching. (C) Probability of being classified as bleached as a function of the compensation index (y-axis of Figure 12B). Each data point represents a flat bleached sample. On the y-axis, the probability of bleached response, computed as proportion of participants classifying the sample as bleached. The colors of the data points represent the average color of the corresponding samples. Samples judged as bleached significantly more or less than chance are marked with *. Significance was determined with a Bonferroni corrected z-test on proportions (12 comparisons).

of the rest of the surface for compensation). If so, the presence of stains should predict the probability that a given sample is perceived as bleached. A simple measure of this is the standard deviation of lightness and chroma, as we know that bleaching mostly affects these two dimensions. However, the standard deviation not only captures the variability owing to bleaching, but also shading, as well as the texture of the fabrics.

To isolate the variability owing to bleaching, we decided to band-pass filter our images before computing the standard deviation, thus removing the low frequency content, which is presumably owing to shading, and the high frequency content, owing to the fabrics' texture. The stains do not seem to share the same spatial properties, as they cover a large uniform portion of some samples (e.g., yellow and blue samples in Figure 1B), or appear as relatively small spots distributed across the samples' surface (e.g., purple sample in Figure 1B). Thus we used a range of spatial frequency bands to band-pass filter each image and compute the lightness and chroma standard deviations, searching for the band that maximally relates to observers' choices.

Specifically, we first converted each RGB image into a lightness and a chroma image, then for each of them we computed the Fast Fourier Transform, using the `fft2()` MATLAB function (version R2017b, <http://www.mathworks.com>). To filter the images in the spectral domain, we applied a set of ring filters whose radius spanned 30 frequency bins (centers: [0.09 3.99] cycles/degree). Finally, for each filtered image,

we computed the standard deviation, for a total of $30 \text{ (frequency bins)} \times 2 \text{ (lightness and chroma)} = 60$ predictors, for each of the 12 flat bleached samples and the 12 flat original samples. We focused on the flat samples because for those images the variability owing to shading was likely limited to low spatial frequency bands. These predictors were used by a linear classifier trained to predict observers' bleached versus nonbleached responses. We used lightness and chroma separately and combined for each spatial frequency independently.

Classification performance was computed with a linear classifier, which was trained on all the responses but the ones from the one-left-out observer. Performance was computed on the responses of the left-out observer, and this procedure was iterated for all the participants, then performance was averaged across participants.

Peak performance (78% accuracy) is achieved by the joint use of lightness and chroma at approximately 0.1 cycle/degree.

Peak performance is close to the interobservers consistency (~82%), computed as the proportion of times that an observer's classification of a sample agreed with the dominant classification response from all the other observers for that sample, averaged across observers. This suggests high performance of the linear classifier with no overfitting.

Figure 14 shows the standard deviations (lightness vs. chroma) classification plane. Observers judged as original those samples with relatively low lightness

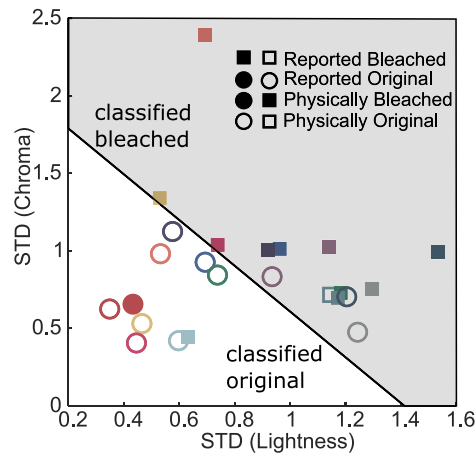


Figure 14. Classification results in the lightness versus chroma plane at the best performing spatial frequency band. Lightness standard deviation on the x-axis; chroma standard deviation on the y-axis. Circles represent the samples classified as original by the majority of the observers; squares represent the samples classified as bleached. Filled symbols represent the bleached samples; empty symbols represent their original versions. The color of each symbol is the mean color of the corresponding fabric. The black line is the classification line.

and chroma standard deviations. Both the dimensions seem to contribute to the classification performance. For instance, the bleached sample that was reported as original by the majority of the observers was characterized by relatively low lightness and chroma standard deviations (left-most reddish data point in Figure 14). Conversely, the original sample that was reported as bleached by the majority of the observers was characterized by relatively high lightness and chroma standard deviations (right-most empty square), although the probability of judging it as bleached was not significantly different from chance ($p = 0.5034$). There seems to be an exception to this pattern: a relatively uniform bleached sample was correctly reported by the majority of observers as bleached (gray square below the classification line). For this sample, mean lightness was relatively high (1.4 standard deviation units higher than mean) and chroma relatively low (1 standard deviation unit lower than average), consistent with the effect of bleaching. However, classification performance based on mean lightness and chroma was poor (56%), suggesting that the absolute values do not provide evidence for bleaching. Furthermore, the probability of this particular sample to be perceived as bleached was not significantly different from chance (Figure 13B), therefore there is no compelling evidence that absolute lightness or chroma are informative about the bleaching magnitude, as they probably depend on the pigment used to dye the individual fabrics. Rather, spatial variations in

lightness and chroma at particular spatial frequencies are stronger cues to bleaching.

Discussion

We investigated whether observers can reproduce the color that bleached fabrics were before they were bleached. To do so, we used experimental stimuli that underwent an artificial bleaching process resembling the effects of ageing and wear on fabrics. These changes mainly involved an increase in lightness and decrease of chroma, consistently with the fading of color and increase in lightness reported in the literature about effects of ageing on textiles (e.g., Beek & Heertjes, 1966; Lead, 1949; Moir, 1971). We used a color matching task in which we asked participants to reproduce the colors of photographs of bleached fabrics as they looked before bleaching, that is, to compensate for the effects of bleaching. These matches were compared with the ones obtained in a baseline condition in which a different group of observers was asked to match the color of the same bleached fabrics as accurately as possible. Observers could compensate for the effect of bleaching on both lightness and chroma.

It has been suggested that humans can to some extent determine and discount the causal history of objects (Leyton, 1989; Schmidt & Fleming, 2016; Schmidt et al., 2016; Spröte & Fleming, 2016; Spröte, Schmidt, & Fleming, 2016; Yoonessi & Zaidi, 2010). Thus it is possible that we can to some extent revert the bleaching process, having internalized through experience the way textiles fade with bleaching or ageing. An alternative parsimonious hypothesis is that observers based their matches on regions that were least affected by the bleaching process. Crucially, this hypothesis does not rely on any internal representation of the bleaching process. It only assumes that observers are able to correctly identify areas of the bleached samples in which the effects of bleaching on color (i.e., increased lightness and decreased chroma) were scarcely pronounced. Because we used homogeneous surfaces, the participants simply had to assume that any observed variations in color were owing to bleaching. Indeed, chromatic variability was probably used as perceptual image cue for bleaching, as suggested by results from Experiment 3.

In future work, it would be interesting to measure the ability of participants to compensate for bleaching on materials with heterogeneous pigmentation, as this would require observers to distinguish between different causes of spatial color variations (patterning vs. bleaching).

The fact that people can infer the original color of bleached surfaces indicates that they can identify the chromatic changes because of the bleaching process and

to some extent discount them. This is consistent with the finding that people use chromatic cues (i.e., color saturation) to perceive a surface as wet (Sawayama et al., 2017). Because the changes are interpreted as a change in the state of the material, rather than a change in the material itself, it suggests that the visual system can somehow separate the saturation increase that is owing to wetting from the intrinsic color of the surface.

In a second experiment, we asked a different group of participants to identify more and less strongly bleached areas on the photographs of our samples. Their choices were mostly based on lightness differences. However, the differences between the matches in the compensation and original conditions can be predicted based on the estimated effect of bleaching on the selected nonbleached areas. This suggests observers may have achieved compensation by identifying and matching the nonbleached areas. Consistent with this hypothesis, when the effect of bleaching was rather small, observers tended to overcompensate, whereas a moderate degree of compensation is exhibited by samples more affected by bleaching. We speculate that when local differences within the samples were perceivable, observers exploited them to retrieve the colors of the original fabrics. By contrast, when the effects of bleaching were spatially uniform, observers used their perceptual knowledge about bleaching and ageing processes, and arbitrarily increased lightness and decreased chroma by a large amount, resulting in overcompensation. Overcompensation might also be explained by the fact that observers had learned the effects of bleaching from the samples on which these effects were visible, that is, their knowledge might come from the stimuli we used rather than from their past experience. Future research may investigate this possibility by controlling the presentation order of the different stimuli.

One possible explanation for this behavior is that memory tends to be selective for more dominant and impressive characteristics of percepts, and thus perceptual judgments based on memory tend to exaggerate these characteristics (Newhall, Burnham, & Clark, 1957). For instance, when matching the remembered color of objects, observers tend to reproduce colors that are more saturated than in simultaneous color matching (Bloj, Weiß, & Gegenfurtner, 2016; Hanawalt & Post, 1942; Newhall et al., 1957; Uchikawa, 1983).

Our results suggest that bleached regions are identified based on lightness differences. To do so, the visual must somehow identify that the variations are owing to reflectance (i.e., stained areas owing to bleaching) rather than illumination (e.g., shadows). Although edge blurriness and textural continuity across a luminance edge generally promote the illumination interpretation (e.g., Agostini & Galmonte, 2002; Lotto & Purves, 2001), humans tend to perceive a blurred

circular spot on a texture as a stain rather than a circular casted shadow (Sawayama & Kimura, 2015). This suggests that the specific spatial pattern of the bleaching may also be important.

When our observers were asked to match the color of the original and bleached samples as accurately as possible, they produced matches higher in lightness and chroma than the mean of the samples' chromatic distributions. This is similar to what was found for artificial colored patches (Kuriki, 2004; Sunaga & Yamashita, 2007) and for natural three-dimensional objects (Giesel & Gegenfurtner, 2010; Toscani et al., 2013a; Toscani et al., 2016). It was proposed that observers base their lightness matches on the brightest areas of shaded surfaces (Toscani et al., 2013a; Toscani & Valsecchi, 2019) because the luminance of these areas is particularly diagnostic for the surface's reflectance. Thus the visual system would apply a simple and useful heuristic to perceive surface reflectance despite variations in illumination and surface geometry. However, when observers were asked to compensate for the effects of bleaching, lightness and chroma matches were closer to the mean values, showing that this heuristic does not apply to all situations; shading and bleaching have different effects on color, and thus require different strategies to be compensated. In fact, to compensate for bleaching, observers should base their matches on darker regions of the surfaces. Other results also show that the visual system uses the most diagnostic regions for judgments, rather than simply selecting the brightest image regions. For example, when judging the lightness of a texture pattern, observers based their matches either on the lightest or the darkest areas, depending on contrast relative to the background (Toscani, Valsecchi, & Gegenfurtner, 2013b). Furthermore, although color matching experiments had shown that observers tend to base their matches on the most saturated parts of the targets' color distributions, when participants are asked to classify a large sample of photographs of leaves they based their judgments on the average chromaticity of the leaves' color distributions (Milojevic, Ennis, Toscani, & Gegenfurtner, 2018). Our results provide additional evidence that color matching is based on different parts of the targets' color distributions according to the task demands.

Conclusions

Overall, we have shown that observers can compensate for bleaching when they were asked to reproduce the appearance of chromatic stimuli. We speculate that compensation behavior is based both on local differences with the images of the bleached

samples and knowledge about the effects of the bleaching process.

Keywords: causal history, color appearance, material perception

Acknowledgments

The authors thank Anna Metzger and Matteo Valsecchi for helpful discussions and comments. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 222641018–SFB/TRR 135 projects A8, C1, and C2.

Commercial relationships: none.

Corresponding author: Matteo Toscani.

Email: matteo.toscani@psychol.uni-giessen.de.

Address: Department of Psychology, Giessen University, Giessen, Germany.

*MT and ZM are joint first authors.

References

- Agostini, T., & Galmonte, A. (2002). A new effect of luminance gradient on achromatic simultaneous contrast. *Psychonomic Bulletin and Review*, 9(2), 264–269.
- Beek, H. V., & Heertjes, P. (1966). Fading by light of organic dyes on textiles and other materials. *Studies in Conservation*, 11(3), 123–132.
- Bloj, M., Weiß, D., & Gegenfurtner, K. R. (2016). Bias effects of short-and long-term color memory for unique objects. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 33(4), 492–500.
- CIE. (1978). *Recommendations on uniform color spaces, color-difference equations, psychometric color terms*. Paris: CIE.
- Del Signore, A. (2011). Colour measurement in the various methods for the artificial ageing of jeans. *Journal Of Commodity Science, Technology And Quality*, 50(III), 243–259.
- Ennis, R., Schiller, F., Toscani, M., & Gegenfurtner, K. R. (2018). Hyperspectral database of fruits and vegetables. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 35(4), B256–B266.
- Fleming, R. W., & Schmidt, F. (2019). Getting “fumbered”: Classifying objects by what has been done to them. *Journal of Vision*, 19(4):15, 1–12, <https://doi.org/10.1167/19.4.15>.
- Giesel, M., & Gegenfurtner, K. R. (2010). Color appearance of real objects varying in material, hue, and shape. *Journal of Vision*, 10(9):10, 1–21, doi:[10.1167/10.9.10](https://doi.org/10.1167/10.9.10).
- Hanawalt, N. G., & Post, B. E. (1942). Memory trace for color. *Journal of Experimental Psychology*, 30(3), 216.
- Hansen, T., & Gegenfurtner, K. R. (2013). Higher order color mechanisms: Evidence from noise-masking experiments in cone contrast space. *Journal of Vision*, 13(1):26, 1–21, doi:[10.1167/13/1/26](https://doi.org/10.1167/13/1/26).
- Ishihara, S. (2004). *Ishihara's tests for colour deficiency*. Tokyo, Japan: Kanehara Trading Inc.
- Kan, C., Yuen, C., & Wong, W. (2011). Optimizing color fading effect of cotton denim fabric by enzyme treatment. *Journal of Applied Polymer Science*, 120(6), 3596–3603.
- Kuriki, I. (2004). Testing the possibility of average-color perception from multi-colored patterns. *Optical Review*, 11(4), 249–257.
- Lead, W. (1949). Artificial light-fastness tests on coloured textiles. *Coloration Technology*, 65(12), 723–731.
- Leyton, M. (1989). Inferring causal history forms shape. *Cognitive Science*, 13(3), 357–387.
- Lotto, R. B., & Purves, D. (2001). An empirical explanation of the Chubb illusion. *Journal of Cognitive Neuroscience*, 13(5), 547–555.
- Mahy, M., Eycken, L., & Oosterlinck, A. (1994). Evaluation of uniform color spaces developed after the adoption of CIELAB and CIELUV. *Color Research and Application*, 19(2), 105–121.
- Moir, J. (1971). Light fastness assessments—An extension of recipe prediction. *Coloration Technology*, 87(12), 442–450.
- Mondal, M. I. H., & Khan, M. M. R. (2014). Characterization and process optimization of indigo dyed cotton denim garments by enzymatic wash. *Fashion and Textiles*, 1(1), 19.
- Newhall, S., Burnham, R., & Clark, J. R. (1957). Comparison of successive with simultaneous color matching. *Journal of the Optical Society of America*, 47(1), 43–56.
- Pernet, C. R., Wilcox, R. R., & Rousselet, G. A. (2013). Robust correlation analyses: False positive and power validation using a new open source Matlab toolbox. *Frontiers in Psychology*, 3, 606.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16(2), 225–237.
- Sarkar, J., & Elias Khalil, M. (2014). Effect of enzyme washing combined with pumice stone on the

- physical, mechanical and color properties of denim garments. *International Journal of Research in Advent Technology*, 2(9), 65–68.
- Sawayama, M., Adelson, E. H., & Nishida, S. Y. (2017). Visual wetness perception based on image color statistics. *Journal of Vision*, 17(5):7, 1–24, doi:10.1167/17.5.7.
- Sawayama, M., & Kimura, E. (2015). Stain on texture: Perception of a dark spot having a blurred edge on textured backgrounds. *Vision Research*, 109, 209–220.
- Schmidt, F., & Fleming, R. W. (2016). Visual perception of complex shape-transforming processes. *Cognitive Psychology*, 90, 48–70.
- Schmidt, F., & Fleming, R. W. (2018). Identifying shape transformations from photographs of real objects. *PLoS One*, 13(8), e0202115.
- Schmidt, F., Phillips, F., & Fleming, R. W. (2019). Visual perception of shape-transforming processes: ‘Shape Scission’. *Cognition*, 189, 167–180.
- Schmidt, F., Spröte, P., & Fleming, R. W. (2016). Perception of shape and space across rigid transformations. *Vision Research*, 126, 318–329.
- Sharan, L., Rosenholtz, R., & Adelson, E. (2009). Material perception: What can you see in a brief glance? *Journal of Vision*, 9(8):784, 784, <http://www.journalofvision.org/content/9/8/784>, doi:10.1167/9.8.784. [Abstract].
- Spröte, P., & Fleming, R. W. (2016). Bent out of shape: The visual inference of non-rigid shape transformations applied to objects. *Vision Research*, 126, 330–346.
- Spröte, P., Schmidt, F., & Fleming, R. W. (2016). Visual perception of shape altered by inferred causal history. *Scientific Reports*, 6, 36245.
- Sunaga, S., & Yamashita, Y. (2007). Global color impressions of multicolored textured patterns with equal unique hue elements. *Color Research and Application*, 32(4), 267–277.
- Toscani, M., & Valsecchi, M. (2019). Lightness discrimination depends more on bright rather than shaded regions of three-dimensional objects. *i-Perception*, 10(6), 2041669519884335.
- Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2013a). Optimal sampling of visual information for lightness judgments. *Proceedings of the National Academy of Sciences*, 110(27), 11163–11168.
- Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2013b). Selection of visual information for lightness judgements by eye movements. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 368(1628), 20130056.
- Toscani, M., Valsecchi, M., & Gegenfurtner, K. R. (2017). Lightness perception for matte and glossy complex shapes. *Vision Research*, 131, 82–95.
- Toscani, M., Zdravković, S., & Gegenfurtner, K. R. (2016). Lightness perception for surfaces moving through different illumination levels. *Journal of Vision*, 16(15):21, 1–18, doi:10.1167/16.15.21.
- Uchikawa, K. (1983). Purity discrimination: Successive vs simultaneous comparison method. *Vision Research*, 23(1), 53–58.
- Wiesel, C. B., Toscani, M., & Gegenfurtner, K. R. (2015). Statistical correlates of perceived gloss in natural images. *Vision Research*, 115, 175–187.
- Wiesel, C. B., Valsecchi, M., & Gegenfurtner, K. R. (2013). The speed and accuracy of material recognition in natural images. *Attention, Perception, and Psychophysics*, 75(5), 954–966.
- Yoonessi, A., & Zaidi, Q. (2010). The role of color in recognizing material changes. *Ophthalmic and Physiological Optics*, 30(5), 626–631.
- Zar, J. (1999). *Biostatistical analysis* 4th ed. Upper Saddle River, NJ: Prentice-Hall.