Colour expectations across illumination changes

Hamed Karimipour [*], Christoph Witzel

School of Psychology, Southampton, United Kingdom

ABSTRACT

This study investigates human expectations towards naturalistic colour changes under varying illuminations. Understanding colour expectations is key to both scientific research on colour constancy and applications of colour and lighting in art and industry. We reanalysed data from asymmetric colour matches of a previous study and found that colour adjustments tended to align with illuminant-induced colour shifts predicted by naturalistic, rather than artificial, illuminants and reflectances. We conducted three experiments using hyperspectral images of naturalistic scenes to test if participants judged colour changes based on naturalistic illuminant and reflectance spectra as more plausible than artificial ones, which contradicted their expectations. When we consistently manipulated the illuminant (Experiment 1) and reflectance (Experiment 2) spectra across the whole scene, observers chose the naturalistic renderings significantly above the chance level (>25 %) but barely more often than any of the three artificial ones, collectively (>50 %). However, when we manipulated only one object/area's reflectance (Experiment 3), observers more reliably identified the version in which the object had a naturalistic reflectance like the rest of the scene. Results from Experiments 2–3 and additional analyses suggested that relational colour constancy strongly contributed to observer expectations, and stable cone-excitation ratios are not limited to naturalistic illuminants and reflectances but also occur for our artificial renderings. Our findings indicate that relational colour constancy and prior knowledge about surface colour shifts help to disambiguate surface colour identity under illumination changes, enabling human observers to recognise surface colours reliably in naturalistic conditions. Additionally, relational colour constancy may even be effective in many artificial conditions.

1. Introduction

Imagine you bought something, say a T-shirt, under the artificial illumination of a shop, but when you look at it again under a different light, its colours aren't as expected (cf. Fig. 1). This shows that you have expectations about how the colour of that T-shirt should look under different illuminations. Understanding these colour expectations is important for both scientific research on colour constancy (for review, see Foster, 2011; Hurlbert, 2019; Witzel & Gegenfurtner, 2018) and colour and lighting applications in art and industry (for review, see, e.g., Houser et al., 2016; Masuda & Nascimento, 2013; Nascimento & Masuda, 2014; Pinto et al., 2010; Pinto et al., 2006; Zhai et al., 2016). For example, artwork and products with artificial, manmade pigments may have unreliable, unexpected colours under illumination changes (Berns, 2016; Fairchild & Johnson, 2004; Samadzadegan & Urban, 2013). In addition, illumination through manmade light sources, such as LED lights, may appear unnatural to human observers even if the light

itself has the same colour as natural daylight (Jost-Boissard et al., 2015). The perceived naturalness of the illumination is a fundamental aspect of the light source and depends on how the illumination renders the colours of objects and materials (henceforth *surface colours*) in a scene (David, 2014; Guo & Houser, 2004; Houser et al., 2016; Houser et al., 2013; Hurlbert, 2019; Royer, 2022; Smet & Hanselaer, 2016). For observers to judge the naturalness of surface colours and scene renderings, they need prior knowledge and expectations about how the colours look under natural conditions.

Surface colour changes in an unexpected, unnatural way across illuminations are possible because of the complex interaction between illuminant and reflectance spectra. Human colour vision does not convey spectral information because it relies on the univariance of the cone photoreceptors. For this reason, two lights may have the same sensory colour signal (in terms of cone excitations), i.e., they are *metameric*. Although we cannot see the colour differences of those lights, their absorption and reflection by surfaces will differ. As a result, the

---

same surface may look differently under two illuminants that are metameric and look the same.

As a further complication, there may be two different surface reflectances that are metameric under one light (i.e., the reflected light from these two surfaces produces the same cone excitations) but look different under another light because they are not metameric under that second light. This phenomenon is kown as *metamer mismatching* (Esposito et al., 2022; Logvinenko et al., 2014; Logvinenko et al., 2015; Witzel et al., 2016; Zhang et al., 2016). Here, we will qualify it as *reflectance-metamer mismatching* and call the aforementioned phenomenon with the metameric illuminants *illuminant-metamer mismatching* to distinguish the two slightly different phenomena.

Hence, without spectral information, the shift of the colour signal under illumination changes is uncertain. Based on all theoretically possible metameric spectra, we can calculate the *metamer mismatch volume*. For a given colour signal under a given illumination, the metamer mismatch volume describes all theoretically possible colour signals under another illumination (Godau & Funt, 2012; Logvinenko et al., 2014; Mackiewicz et al., 2019) and provides an estimate of the theoretical uncertainty of colour shifts (Logvinenko et al., 2015; Witzel et al., 2016). Despite this theoretical uncertainty, human observers have colour constancy, which is the ability to recognise colours across changes of illumination in most of their everyday life environments (Foster, 2011; Gegenfurtner et al., 2024; Witzel & Gegenfurtner, 2018). Chromatic adaptation and local contrast impose unambiguous shifts in colour appearance, which support constant colour appearance across illumination changes (see, e.g., Brainard, 1998; Hansen et al., 2007; Ling & Hurlbert, 2008; Murray et al., 2006). Through inferential colour constancy (Witzel & Gegenfurtner, 2018), human observers are also capable of recognising colours across illuminations without adapting colour appearance, such as, for example when comparing photos taken under different illuminations (cf. Fig. 1; for related arguments, see, e.g., Brainard et al. (1997); Radonjic et al. (2015); Arend and Reeves (1986)). Inferential colour constancy is only possible if the colour shift is unambiguous. In the real world, the theoretical uncertainty may be resolved if (1) spectral variation and metamer mismatching are constrained in the typical everyday environment of human observers, and (2) humans know (implicitly or explicitly) what colour changes to expect under those conditions.

Concerning point (1), not all the spectra on the basis of the theoretical metamer mismatch volume physically exist. We will call spectra "*realistic*" when they occur in reality or produce similar colour shifts as real spectra. It has been found that metamers and therefore metamer mismatches are rare in typical everyday-life environments (Akbarinia & Gegenfurtner, 2018; Foster, Amano, et al., 2006). Those everyday-life environments do not only consist of natural objects and materials, such as stone, soil, and organic surfaces (Chiao, Cronin, & Osorio, 2000; Osorio & Bossomaier, 1992). They also contain manmade things, such as houses or streets (Foster, Amano, et al., 2006; Linhares et al., 2008; Nascimento & Foster, 2023). Nevertheless, the colour distributions of these scenes change in a statistically regular fashion across illumination

changes. They feature nearly constant cone-excitation ratios (Foster, Amano, & Nascimento, 2006; Foster & Nascimento, 1994; Foster et al., 1997; Golz & MacLeod, 2002; Nascimento et al., 2004; Nascimento et al., 2002; Nascimento & Foster, 1997; Nascimento & Foster, 2000), can be closely approximated by linear transformation (Karimipour et al., 2023; Philipona & O'Regan, 2006; Witzel et al., 2015), and tend to show a correlation between luminance and redness (Golz, 2008; Golz & MacLeod, 2002). These statistical regularities seem to hold for natural daylight (e.g., Chiao, Osorio, et al., 2000; Romero et al., 1997) as well as for broadband spectra that emulate natural daylight. To accommodate that those reflectance and illuminant spectra are not necessarily natural, we will call spectra and colour changes *naturalistic* when the (colorimetric/sensory) colour changes are the same (or almost the same) as those that human observers typically experience in their everyday lives. In contrast, spectra involving colour changes that are visibly different from naturalistic spectra will be called *artificial*. For example, Fig. 1 illustrates apparent artificial colour shifts, which are caused by artificial dyes (clothes) under narrowband LED lightings.

While metamer mismatching is very constrained in the naturalistic environment (1), evidence that human observers know and use those constraints (2) is ambiguous. On the one hand, studies suggest that colour constancy is specific to naturalistic illuminant spectra. Some found that colour constancy is higher for naturalistic than artificial illuminants (Lucassen & Walraven, 1996), and others have shown that human observers are sensitive to the different kinds of colour statistics across naturalistic illuminations (Foster, Amano, & Nascimento, 2006; Golz, 2008; Golz & MacLeod, 2002; Lucassen et al., 2013; Nascimento et al., 2004; Nascimento & Foster, 1997, 2001; Nascimento & Foster, 2023).

On the other hand, observers do not judge naturalistic daylight illuminants as being most natural (Masuda & Nascimento, 2013; Nascimento & Masuda, 2012). In addition, observers seem to consider perfectly constant ratios to be more natural (Nascimento & Foster, 1997) or more like illuminant changes (Nascimento & Foster, 2001; Nascimento & Foster, 2000; Nascimento & Foster, 2023) than actual illuminant changes, in which cone ratios may vary slightly. These latter studies suggest that observers do not expect naturalistic colour changes but rather rely on cone excitation ratios. Other studies found that constant cone excitation ratios do not well predict colour constancy when measured through asymmetric matches (Witzel et al., 2016) and achromatic adjustments (Weiss et al., 2017). Considering these inconclusive findings, it remains an open question (1) whether human observers have expectations about the naturalistic colour changes they have experienced in most of their everyday lives, and (2) whether those expectations are precise enough to resolve the ambiguity of metamer mismatching and allow for inferential colour constancy in a naturalistic environment.

Rather than addressing these questions with respect to the theoretical uncertainty of metamer mismatching, we focused on realistic variations of colour shifts and limited our examination to naturalistic and realistic artificial spectra. We investigated whether observers expect

Fig. 1. Illustration of Colour Expectations. Photos of the same clothes were taken under four illuminations (a-d) in an LED lightbox with narrowband LED lights, which produce particularly unexpected colour changes across illuminations. Since the surface colours do not change as expected, they are not colour-constant and seem to change across the four illuminations.

naturalistic colour changes across illuminations and whether these expectations allowed them to distinguish naturalistic from realistic artificial colour shifts. In our first approach, we reanalysed the asymmetric matches of Witzel and colleagues (2016) to estimate observer expectations and compare them with naturalistic and artificial colour changes. Then, three experiments specifically tested whether observers could tell naturalistic from artificial colour changes, such as those in Fig. 1. In contrast to the rough illustrations in Fig. 1, we rendered scenes in these experiments through hyperspectral images (Foster & Amano, 2019). This allowed us to manipulate reflectance and illuminant spectra while controlling scene content and illumination colours. In the first experiment, we tested naturalistic and artificial illuminants while keeping the original, naturalistic reflectance spectra of the hyperspectral images. In the second experiment, we varied both reflectance and illuminant spectra. In the last experiment, we tested whether observers have expectations towards single objects in comparison to the rest of the scene.

2. Asymmetric colour matches

Asymmetric matches require observers to estimate a surface colour under a test illumination while simultaneously comparing it with a version under a reference illumination. To do this, observers need to rely on their implicit or explicit expectations about how the illumination change affects the surface colour. Besides expectations, sensory mechanisms, such as local contrast induction, brightness anchoring, and partial adaptation, may contribute to the colour appearance across illumination change (e.g., Hansen et al., 2007; Kraft & Brainard, 1999). These mechanisms should support observer expectations because they serve the same purpose, i.e., colour constancy in the naturalistic environment. So, the central tendency of asymmetric matches indicates the colours observers expect across the illumination changes. We call the colour expected by an observer the *subjective target*. In contrast, *candidate targets* are objective target predictions based on realistic illuminant and reflectance spectra that produce metamer mismatches. If observer expectations are shaped by naturalistic colour shifts, then the subjective target should coincide with the naturalistic rather than an artificial candidate target.

In addition, the variation of the asymmetric matches reflects the uncertainty of observers about their subjective target, independent of whether the subjective target coincides with the naturalistic target or not. We refer to this as *subjective uncertainty*. The subjective uncertainty defines whether expectations are precise enough to distinguish the subjective target from alternative candidate targets. The ability to distinguish candidate targets (metamer mismatches) from subjective targets can be understood as a *mismatch sensitivity* and depends on both the subjective uncertainty and the distance of candidate targets from the subjective target. We call these differences *mismatch differences*. If observer expectations are precise enough to reduce the ambiguity resulting from metamer mismatching, subjective uncertainty must be lower than mismatch differences.

We tested these hypotheses by reanalysing the asymmetric matches from a previous study (Witzel et al., 2016). Results of that study suggested that asymmetric matches cluster around the prediction by a naturalistic target (Figs. 2 and Figures S1 and S2 in Witzel et al., 2016). However, a comparison with alternative, artificial targets was not provided. Here, we defined candidate targets by naturalistic and a range of realistic artificial illuminant and reflectance spectra. We computationally simulated the realistic spectra rather than using physically real ones to experimentally control their properties, especially their metamer mismatching. So, "realistic" means that these virtual spectra were designed to resemble real, physical spectra and could potentially occur in reality unlike the five-transition spectra that define the hull of the metamer mismatch volume. The results of these reanalyses also informed the design of the subsequent experiments.
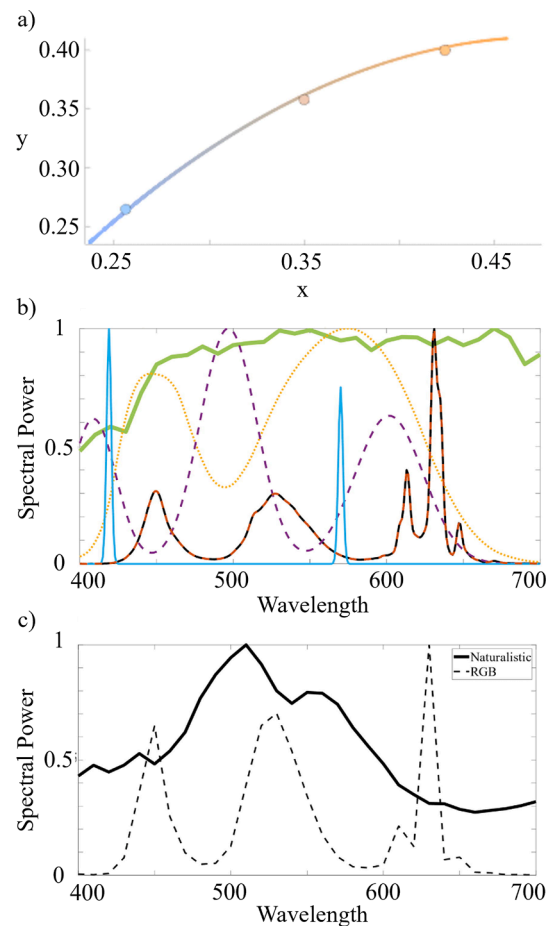
Fig. 2. Illuminants and reflectances. (a) Circles show the CIE1931 chromaticity coordinates for illuminants in Experiments 1 to 3. The curve indicates the daylight locus for reference. (b) metameric yellow illuminants: naturalistic daylight illuminant (green solid line), artificial RGB illuminant (red–black dashed line), Fundamental illuminant (orange dotted line), 3-Gaussian illuminant (purple dashed line) and 2-Narrow illuminant (blue solid line). (c) metameric reflectances: naturalistic broadband (solid line) and artificial RGB reflectances (dashed line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.1. Method

Human Data. Witzel et al. (2016) measured asymmetric matches with 20 observers, for 15 colours (dark, medium, and light grey, 3 types of red, yellow, green, and blue), across two illumination colours (yellowish and bluish), and four photorealistic scenes with grey and coloured tiles. One of the 12 coloured tiles was randomly selected as the target and shown in a random colour. Observers adjusted its chromaticity in the scene under the test illumination (e.g., yellowish) to match the colour of the same tile in the scene under the reference illumination (e.g., bluish), hence producing a surface colour match (or "paper match" in Arend and Reeves (1986)). The measurements were repeated across two sessions to determine intraindividual variation as an estimate of uncertainty. Adjustments were done in CIELUV space. The two illuminants were naturalistic daylight spectra; the yellowish one was computed based on Judd's daylight model (Judd et al., 1964) at a Correlated Colour Temperature of 5000 K (cf. green curve in Fig. 2.b) and the bluish one was based on a black body simulator at a Correlated Colour Temperature of 12,000 K (cf. Fig. 5.a in Witzel et al., 2016). Surface colours were rendered based on reflectances of Munsell chips retrieved from the database of the Joensuu Colour Group (Kohonen et al., 2006; Parkkinen et al., 1989). The background was the same as the

medium-grey surface.

Mismatch Simulations. We simulated realistic metamer mismatching across the two illumination colours (yellowish vs bluish) by rendering surface colours and the grey background based on naturalistic and artificial illuminant and reflectance spectra. Previously, we had observed that Munsell reflectances produce colour shifts very similar to a principal component model of natural reflectances (cf. Fig. 2, Figures S1 and S2, and Figures S10 in Witzel et al., 2016). In addition, here we compared Munsell chips to another approximation of naturalistic reflectances based on nonnegative matrix factorisation (for details, see Figure S1 and S3 in supplementary material). It confirmed that Munsell colours behave similarly to naturalistic reflectances (cf. Figures S1 and S3). Hence, we considered Munsell colours to be naturalistic. We created metameric artificial reflectances through linear combinations of the spectra of monitor R-, G-, and B-primaries. A different set of 15 RGB reflectances (incl. Background) has been calculated to be metameric with the naturalistic ones under each reference illuminant (yellow vs blue).

We created four types of artificial illuminants that were metameric with the two naturalistic ones (see Table S1 for illuminants' chromaticities): Fundamental, RGB, trimodal Gaussian, and bimodal narrowband illuminants (Fig. 2.b). (1) Fundamental illuminants were the Fundamental metamers (Cohen & Kappauf, 1982, 1985) of the naturalistic illuminants; although, these illuminants might not necessarily exist as such, Fundamental metamers were interesting because they are common to all real illuminants after discounting for metameric black (see Table S2 for equation). (2) RGB illuminants were calculated as linear combinations of the aforementioned monitor spectra (see Table S2 for equation). **(3)** Trimodal Gaussian illuminants **(3-Gaussian)** were combinations of three medium band-width Gaussian functions with a standard deviation of 20 nm. The average and amplitude were adapted to create the metamers (by minimising the colorimetric difference) with the constraint that the averages of three Gaussians remained within the intervals 400–420 nm, 490–510 nm, and 590–610 nm, respectively. (4) Bimodal narrowband illuminants (*2-Narrow*) were made of only two very narrow Gaussian functions, similar to the mixture of two narrowband LEDs (e.g., Fig. 2 in Hurlbert, 2019). They were obtained by minimising the colorimetric difference (less than 0.001) while varying average, standard deviation, and amplitude, with the constraint that the standard deviation was within 2.5 and 7.5 nm.
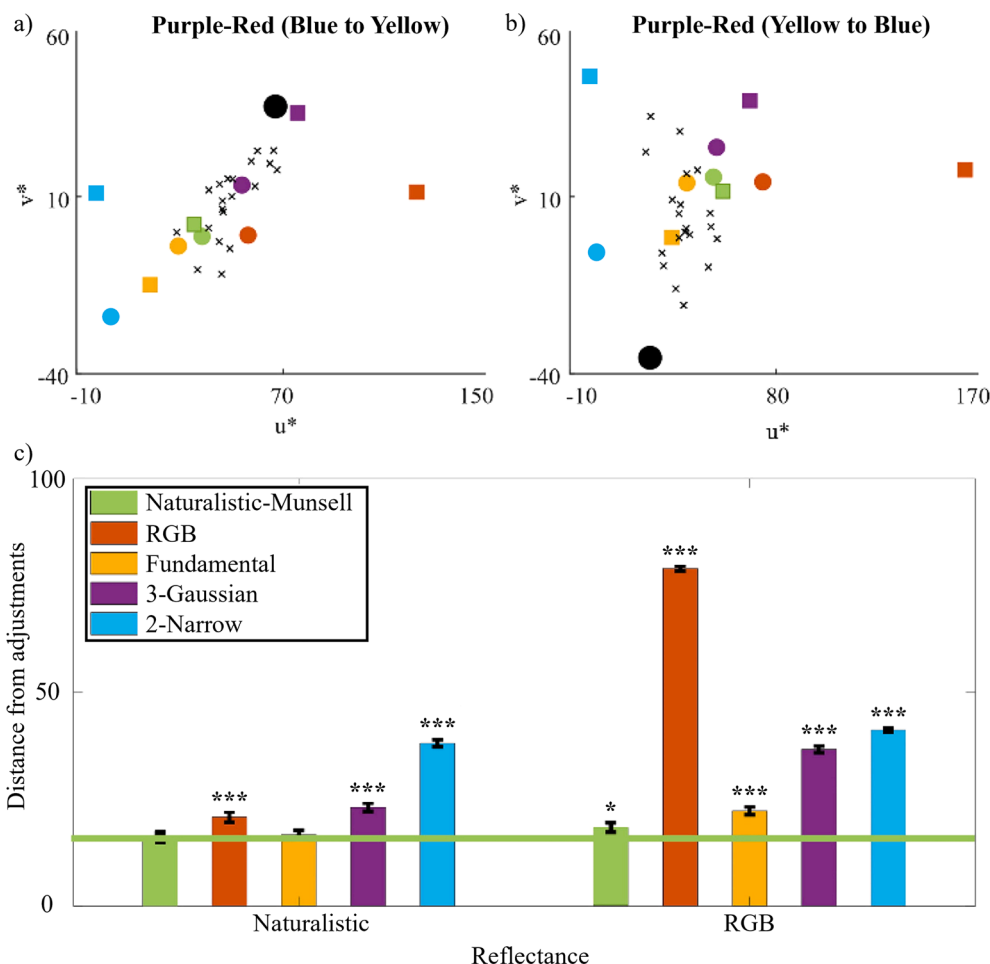
Fig. 3. Expectations in asymmetric matching. Panels a and b compare subjective and candidate targets for a purplish red Munsell surface (7.5 PB5/8) under a blue-to-yellow and a yellow-to-blue illumination changes, respectively. Axes correspond to green–red (u*) and blue-yellow (v*) in CIELUV space. Small black crosses mark subjective targets of each observer. The black disk in each panel indicates the sensory signal under the reference illuminant. The other symbols correspond to candidate targets; their colour indicates the illuminant (see panel c for legend) and their shape the reflectance (circle of Munsell, square for RGB). Panel c illustrates mismatch differences averaged across all 15 surface colours. The x-axis corresponds to candidate targets, the y-axis to Euclidean differences in CIELUV between candidate targets and observer adjustments. Bar colours refer to the illuminants in a and b (cf. legend). The left group corresponds to naturalistic Munsell (circles in a-b), the right to RGB reflectances (squares in a-b). Error bars indicate standard errors across observers. The result for the naturalistic target based on Munsell reflectances (leftmost green bar) is reproduced by the horizontal line to facilitate comparisons. Significant differences from the naturalistic target are marked with * (p < 0.05), ** (p < 0.01) and *** (p < 0.001). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. Results

We compared subjective targets (adjustments by observers) with candidate targets (metamer mismatches) in CIELUV space. These comparisons were done for all 30 illumination shifts, i.e., the 15 surface colours and the two illumination changes from yellow to blue and blue to yellow. Fig. 3.a and .b illustrate these comparisons with two examples; see Figures S1 and S3 for all surface colour changes. Subjective targets were calculated as average adjustments across the two repeated measurements (crosses in Fig. 3.a-b), and candidate targets were computed based on the naturalistic and artificial illuminant and reflectance spectra (chromatic symbols in Fig. 3.a-b).

Subjective Targets. First, we examined whether subjective targets corresponded with naturalistic targets. In Fig. 3.a-b and S1 and S3, individual adjustments cluster between the reference (black circle) and the naturalistic targets (green symbols). The distribution towards the reference indicates an undershoot of adjustments that has been noted earlier (Witzel et al., 2016). Apart from the undershoot, the adjustments seem to be directed towards the naturalistic target, in line with the prediction. We tested whether the subjective targets were closer to naturalistic than to any artificial target. We calculated the Euclidean distance between each observer's average adjustments and each candidate target, for each of the 30 illumination shifts (Figures S2 and S4). For the main analyses, we averaged the distances across the 30 illumination shifts (Fig. 3.c).

For comparison, the distance for the naturalistic target, i.e., the one based on daylight illuminant and Munsell reflectance (green bar on the left of Fig. 3.c), is the same as the Munsell targets in Figure S10 of Witzel et al. (2016). We calculated 9 t-tests across observers comparing each artificial candidate target with the naturalistic target (cf. green line in Fig. 3.c). Except for the naturalistic reflectance under the Fundamental illuminant (Orange bar on the left of Fig. 3.c), all candidate targets differed significantly from the naturalistic target. For artificial RGB reflectances this was true for almost all 30 illumination shifts under artificial illuminants, as shown when doing these analyses separately for each of the 30 colour shifts (Figures S2 and S4). However, results for the naturalistic illuminant and for the naturalistic reflectance under Fundamental, RGB and 3-Gaussian were ambiguous because they depended on the surface colour and the illumination shift.

Subjective Uncertainty. Second, we tested whether observer's subjective uncertainty was lower than mismatch differences. In one approach, we assessed subjective uncertainty intra-individually, as the average difference of the two repeated measurements from the subjective target (i.e., their average). Mismatch differences were calculated as the difference between each candidate target and the individual subjective targets (crosses in Fig. 3.a-b). For each candidate target, a separate t-test across participants assessed whether the mismatch differences were larger than the subjective uncertainty. Since intra-individual variation was low, this was the case for all ten candidate targets, including the naturalistic one (all t(19) > 7.2, all p < 0.001). This implies that individual subjective targets were further away from naturalistic targets than predicted by subjective uncertainty. In a second approach, we calculated subjective uncertainty as the difference of individual subjective targets from the average across individuals and tested (with t-test across participants) whether these differences were lower than the mismatch differences. This was the case for all candidate targets (all t(19) < -3.8, all p =< 0.001) except for the naturalistic reflectance under the naturalistic and Fundamental illuminants, and the RGB reflectance under the naturalistic illuminant (.64 > all t(t(19) > -0.1.5., all p > 0.1) Additional analyses for each of the 30 colour shifts separately suggest that results vary depending on the surface colours and the illumination changes (cf. Figure S2 and S4). Nevertheless, observer expectations were precise enough to allow discarding most candidate targets and hence achieve some level of mismatch sensitivity.

2.3. Discussion

Results demonstrated that, generally, observer adjustments aligned more closely with the naturalistic than the artificial colour shifts, suggesting that observer expectations are in line with their typical everyday-life environment. Results also suggested some level of mismatch sensitivity that allows reducing the uncertainty inherent in metamer mismatching. However, subjective targets did not precisely coincide with naturalistic targets because of systematic undershoots, and the Fundamental target was close to and difficult to distinguish from the naturalistic targets. This was also the case for a few other candidate targets when considering every single of the 30 colour shifts (cf. Figures S2 and S4). These latter observations contradicted the idea that observers' expectations correspond to naturalistic conditions.

The undershoot could be explained by partial adaptation in asymmetric matching and/or by observers partially engaging in colour appearance rather than surface colour matches despite the stimulus design and instructions (Witzel et al., 2016, p. 18). Hence, asymmetric matches may underestimate the alignment of observer expectations with naturalistic colour changes. In addition, colour adjustment procedures such as those in the asymmetric matches, include errors and noise because navigating through colour space requires cognitive efforts and time putting a strain on concentration, motivation, and task handling. For this reason, asymmetric matches may also underestimate the precision of observer expectations and hence their mismatch sensitivity.

In sum, while the results from asymmetric matching suggest that observers tend to expect naturalistic rather than artificial colour changes, they only provide a rough idea about the actual content and precision of those expectations. We developed an alternative, fake-detection approach to assess observer expectations without the noise from adjustment procedures and conducted the following three experiments to more specifically assess observer expectations.

3. Fake lights

The fake-detection approach avoided noise from task handling through an Alternative-Forced Choice task that does not require observers to navigate through colour space. We presented four alternative scene renderings under illumination changes, and observers pick the one they consider plausible and realistic. Expectations were measured by the observers' abilities to identify the naturalistic colour change. We did not explicitly ask observers to judge naturalness because lay concepts of naturalness are complex and do not necessarily reflect actual properties of the natural environment (Buijs, 2009; Rozin et al., 2012; Yendrikhovskij et al., 1999). For example, lay observers find exaggeratedly saturated colours more natural than realistically saturated ones (Yendrikhovskij et al., 1999). They also strongly associate green with naturalness (Dantec et al., 2022; Rozin et al., 2012). So, an exaggerated green image of a forest might yield higher ratings of naturalness than the original, realistic photo of that forest. The present study is not about such lay concepts of naturalness.

Instead, this study was aimed at the colours observers experience in real life, no matter whether they are aware of the "naturalness" of these colours or not. Lay observers may be colour constant without being aware of their implicit assumptions about object and illumination colour (de Almeida & Nascimento, 2009; Granzier et al., 2009). The impact of such implicit assumptions on colour perception has been strikingly illustrated by #theDress (Witzel, Racey, & O'Regan, 2017). Such implicit assumptions constitute the observer's subjective reality. Contradictions will be perceived as incorrect, as it was the case when people were confronted with alternative perceptions of #theDress. Similarly, observers' expectations towards colour changes in our images may be implicit assumptions about reality rather than explicit judgements of naturalness. The fake-detection approach was designed to probe such implicit expectations. We assumed that images with unexpected colour changes would be perceived as implausible or "fake" even if they are

realistic, i.e., physically possible. So, observers are asked to tell a real photo from fake, manipulated images in this approach.

The first experiment focused on illuminants. It tested if naturalistic illuminants, unlike artificial ones (i.e., "fake lights"), render the colours of scenes in the way observers expect so that colour changes look plausible and realistic. This idea was examined for illumination changes along the daylight locus because such illumination changes are the ones observers are most familiar with.

3.1. Method

Participants. 21 participants (14 female, 20.45 ± 2.56 years) took part in the experiment; however, we excluded one participant from the analysis because they made too many errors in catch trials (5 out of 8). None of the participants had red-green colour vision deficiencies as shown by Ishihara plates. Participants were undergraduate and postgraduate students who either received course credits or £8 for their participation. The experiment was approved by the Ethics Committee of the University of Southampton (ERGO 63190). Informed consent was obtained from each participant.

Apparatus. Stimuli were displayed on a 24.1-inch EIZO ColourEdge CG2420 LED monitor with an NVIDIA GeForce GTX 1650 graphics card (NVIDIA Corporation, Santa Clara, CA), 8-bit-per-channel colour depth, a spatial resolution of 1920×1200 pixels, and a refresh rate of 59.95 Hz. The spectra of the monitor's primaries were measured using a CS2000 spectroradiometer. CIE1931 chromaticity coordinates (x, y) and luminance (Y) for the red, green, and blue primaries, were R = (0.685, 0.312, 89.34 cd/m^2), G = (0.216, 0.721, 248 cd/m^2), and B = (0.150, 0.045, 20.50 cd/m^2), respectively. Gamma corrections were applied. A blackened viewing tunnel and the dark experimental room ensured that observers chromatically adapted to the computer display only. Visual angle of stimuli was controlled through a chin rest at a viewing distance of 50 cm.

Stimuli. We aimed at a broad range of different scenes but were limited by the quality of available hyperspectral images and by keeping their renderings within the monitor gamut. This resulted in 14 hyperspectral images from 6 databases (Brainard, 1997; Foster, Amano, et al., 2006; Foster et al., 2016; Kleynhans et al., 2020; Nascimento et al., 2017; Shi et al., 2018), featuring both paintings and natural scenes (Figure S5.a).

We focused on three illumination colours (yellow, neutral, blue) along the daylight locus (Fig. 2.a), resulting in 6 pairs of illumination colour changes (e.g., yellow to neutral). For each illumination colour, we created one naturalistic broadband illuminant spectrum based on Judd's daylight model (Judd et al., 1964). Using the same approaches as in section 'Asymmetric Colour Matches', we constructed four artificial illuminants (including the Fundamental, RGB, 3-Gaussian and 2-Narrow illuminants) metameric to the naturalistic illuminant of the corresponding illumination colour. The renderings were created with the naturalistic broadband and the metameric artificial illuminant spectra. We adapted an existing toolbox for rendering hyperspectral images (Foster & Amano, 2019). The spectra for each pixel of the hyperspectral images were multiplied with the illuminant spectra and the CIE1931 colour matching functions, which correspond with cone sensitivities (Foster & Amano, 2019). The spectra were sampled at regular intervals of 2 nm across the wavelength range of 400–700 nm. The resulting XYZ values were converted into gamma-corrected RGB values. We fixed the width of each image at 300 pixels and adapted the height accordingly. Colour differences (in CIELAB) between naturalistic and artificial renderings are illustrated in Fig. 4.

Fig. 5 illustrates the stimulus display of the main task with particularly visible example stimuli from Experiment 2.b (see Figure S6 for examples specific to Experiment 1). The upper row showed a single *reference* image of one of the scenes rendered with one of the three illumination colours, e.g., yellow. The reference image was always rendered with a naturalistic illuminant. In the lower row, we presented four *test* images of the same scene, but rendered with another illumination colour, e.g., blue. The four test images differed by the spectral properties of the illuminants. Only one of those metameric illuminants was naturalistic and the other three were artificial. Despite being metameric, the four different illuminant spectra rendered the colours in the scene differently. According to our hypothesis, the rendering with the naturalistic illuminant is in line with the observer's expectations and appears realistic. Thus, we call it *naturalistic target*. In contrast, observers cannot have expectations towards our artificial illuminants (*distractors*). These should appear unreal, or "fake".

Procedure. Observers completed a 4-Alternative Forced Choice (4AFC) task to identify the naturalistic target among the four test renderings (See Fig. 5, and Figure S6 for an example specific to Experiment 1). We told observers that only one of the four images in the bottom row is a real photograph under another illumination colour than the reference, and that we manipulated the other three images. We asked
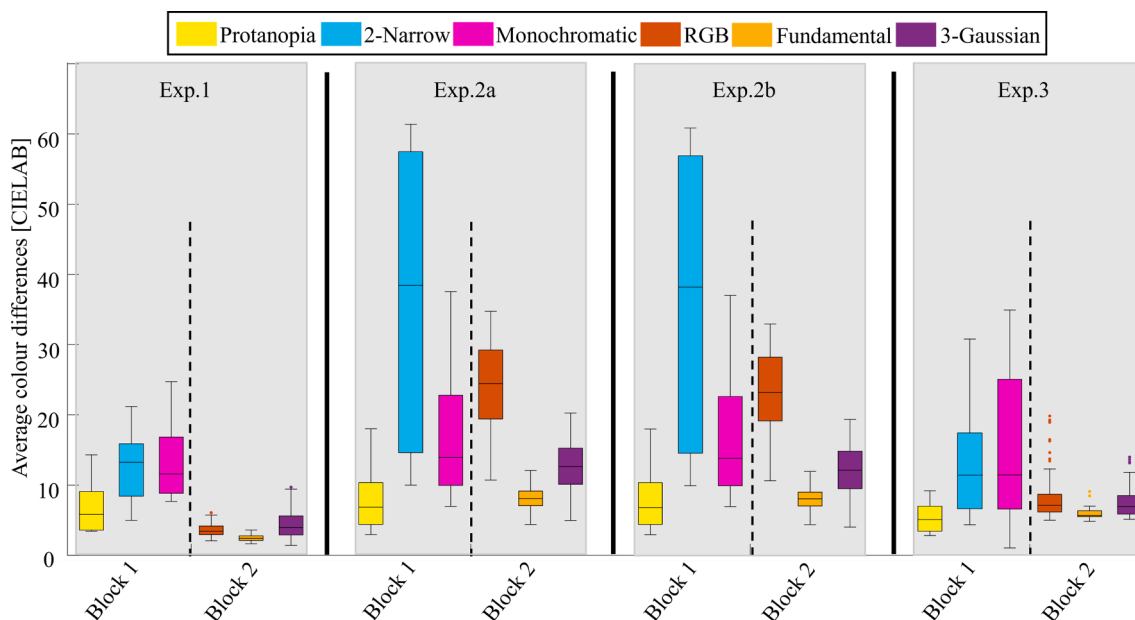

Fig. 4. Colour differences (in CIELAB) between naturalistic and artificial renderings in Blocks 1 and 2 across Experiments 1–3.

Fig. 5. Task & Stimulus Display. Observers had to select the real photograph in the bottom row. This example trial comes from the main condition (block 2) of Experiment 2b because it also manipulated reflectances, and differences between test renderings in the bottom row were more visible than in Experiment 1. We added a red arrow to this illustration to help the reader identify the naturalistic target (bottom left image). See Figure S6 for examples from Experiment 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

observers to identify the real photograph by pressing one of four keys (labelled 1 to 4) that corresponded with the order of the images displayed in the lower row, arranged from left to right. Choices and response times were recorded. The trials and the position of the images in the lower row within a trial and the sequence of trials were randomised. After selection, a slider appeared, along which participants could rate their confidence about the choice between 0 and 100. To measure participant engagement and attention, eight catch trials were randomly distributed across the experiment. Catch trials involved different instructions, asking participants to pick an image identical to a reference image. The other images in these trials were deliberately and noticeably different, making it easy for attentive participants to choose correctly and helping to identify any random or unengaged responses. The experiment consisted of three blocks, that served different purposes:

1. **Task Check.** The success of the fake-detection task relied on the assumption that observers would perceive unexpected images as unrealistic. Thus, it was key that participants correctly understood the main task. We checked this with the first block of our experiment. We created a set of distractors with surface colour changes that were most obviously unnatural. These were created (1) using the 2-narrow illuminant, (2) making them monochromatic (by setting all pixels to the illuminant hue), and (3) simulating protanopia. The 2-narrow illuminant, despite being realistic, was considered obviously unnatural because it completely removes some colours due to large parts of the visual spectrum having zero energy. That 2-Narrow illuminants render surface colours clearly beyond observer expectations is also suggested by our reanalysis above (cf. Fig. 3). Four of the fourteen hyperspectral images (the first four images in Figure S5-a) were used for this first block, three paintings and one naturalistic scene. Participants completed the main task with these renderings. Together, this resulted in 4 (scenes) x 6 (illumination changes) = 24 trials. If participants understood the task, they should be able to choose the naturalistic target over those obviously artificial distractors. These measurements were conducted in the first block to make sure participants understood the task when starting the second, main block.

2. **Main Block.** The second, main block measured whether observer expectations would allow to discern the naturalistic from realistic, artificial colour changes in the fake-detection task. Artificial distractors were rendered with RGB, Fundamental-, and 3-Gaussian illuminants. These produce more subtle colour changes than the 2-

Narrow illuminants in the first block and allow probing the precision of observers' expectations (as shown in Fig. 3.a-b). This block included other images than the first block to avoid that observers learn how the naturalistic targets look from the first block. So, the second block featured the remaining ten hyperspectral images, including five naturalistic scenes and five paintings. Thus, there were 10 (scenes) * 6 (illumination changes) = 60 trials.

3. **Visibility Check.** A precondition of our fake-light approach is that our metameric illuminants produced visible differences in scene renderings. A control measurement in the third block checked that this criterion was met. To do so, we used the same procedure as the fake-detection task, but without change of illumination colour and with different instructions. Thus, the naturalistic target was exactly the same as the reference rendering. Observers were asked to identify the image that is identical to the reference in the upper row. If observers can see the difference between the naturalistic target and the three distractors, they should be able to identify the naturalistic target. Successful discrimination between target and distractors of the second block would imply that the more obvious differences in the first block were also visible. Therefore, only the renderings from the second block were needed to ascertain visibility in both. There were 10 (scenes) * 3 (illumination colours) = 30 trials in random order in the second block.

During these measurements of visibility, participants may learn which of the renderings is the naturalistic reference. This knowledge would likely influence the expectations and choices in the main measurements of the second block. For this reason, these control measurements were conducted in the third block at the end of the experimental session.

3.2. Results

If observers clearly expect a naturalistic colour change, they should choose the naturalistic target most of the time, at least more often than any of the three artificial distractors. Fig. 6.a illustrates the main results in terms of image choices (coloured bars) and confidence ratings (grey bars in background). We used a one-sample *t*-test to evaluate whether observers chose the naturalistic target in more than half of the trials; this implies that they are chosen more often than all of the three distractors together. We also used this criterion (50 % of choices) for the third block
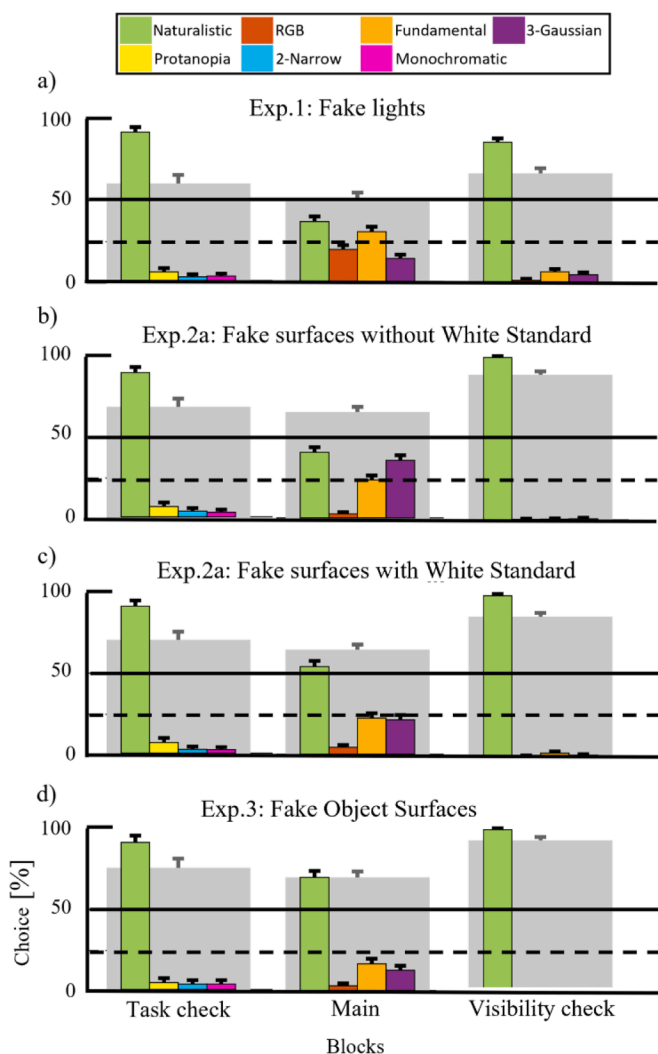
Fig. 6. Main Results. Panels a-d correspond to the different Experiments (Exp.) indicated in the title. The 3 groups of bars in each panel correspond to the 3 blocks of each experiment. The y-axis shows the percentage of choices made by participants. Bars indicate the average frequency for choosing one of the test renderings. The colours of the bars represent the illuminants of those renderings and are specified in the legend at the top. Error bars represent one standard error of the mean. The horizontal solid and dashed line indicate 50% and chance level (25%), respectively. Grey bars in the background illustrate average confidence ratings on the same scale as the percentage along y-axis.

with the discrimination task to facilitate comparisons across blocks. For pairwise comparisons between the frequencies of choosing the naturalistic target and the frequencies of choosing each of the three distractors we Bonferroni-corrected the significance level (α= 0.05/3 = 0.017) when evaluating the three pairwise t-tests.

Task Check. The left group of bars in Fig. 6.a shows the aggregated results with the obviously manipulated distractors in the first block. Observers chose the naturalistic target on average 90 %, which was significantly higher than 50 % (t(19) = 11.25; p < 0.0001). Applying this test to each condition separately showed that the frequency of choosing naturalistic targets was higher than 50 % for each of the four stimuli and each of the three illumination changes (all t(19) > 9.75, p < 0.001). These results were taken as evidence that observers understood the task.

Main Results. The centre group of bars in Fig. 6.a illustrates the results in the second block. Participants chose significantly more often the naturalistic target than the RGB- (t(19) = -4.02, p < 0.001) and 3-Gaussian renderings (t(19) = -8.31, p < 0.001). However, contrary to

our prediction, frequencies of choosing the naturalistic targets (36 %) were below 50 % (t(19) = -7.18; p < 0.001), implying that they were not chosen more often than all the distractors together. Naturalistic targets and Fundamental distractors (30 %) were both chosen above chance level (0.25) (t(19) = 5.34; p < 0.001 and t(19) = 2.92, p = 0.009), and their difference was not significant (t(19) = 1.97, p = 0.06). This suggests that participants confused the Fundamental distractor with the naturalistic target. In addition, choices depended on the illumination changes (cf. Figure S7). Observers chose the RGB distractor (43 %) more often than the naturalistic target (23 %) when the test illumination was blue (t(19) = 2.64, p = 0.016). When the test illumination colour was neutral or yellow, participants chose the naturalistic target or the Fundamental distractor (cf. Figure S7). If observers had expectations from their experience with spectra in their everyday life environment, those expectations should be higher for the five naturalistic scenes than for the five artificial paintings. This was indeed the case (naturalistic: 40 % vs. paintings: 33 %; t(19) = 2.60, p = 0.02).

Visibility Check. Participants selected on average 85 % of the time the correct match (i.e., the naturalistic target) in these control measurements (cf. right group of bars in Fig. 6.a), which was significantly higher than 50 % (t(19) = 11.84; p < 0.001). The minimum performance was 65 % for the painting 'Lake' (cf. the rightmost image in the first row of the second block of Figure S5-a) under the neutral-to-blue illumination change.

Confidence Ratings. The grey bars in the background of Fig. 6.a indicate the average confidence ratings in the three blocks. Two paired t-tests showed that confidence ratings were significantly lower in the main block than in the first and third block (both t(19) < -6.40, both p < 0.001), implying that observers were less certain about their answers in the second block than in the other blocks.

3.3. Discussion

Although, observers show a slim statistical tendency towards picking the naturalistic rendering, this is not consistently the case for all renderings and trials. So, results do not support the idea that observers can reliably tell the naturalistically from artificially illuminated renderings. The results of block 1 show that the missed identifications cannot be due to a failure to understand the task and instructions. Nor are the differences between the renderings too small for observers to discriminate them, as shown by the visibility check in the third block. Hence, observers do not have sufficiently clear expectations about colour changes that would allow them to unambiguously recognise surface colours under naturalistic illumination changes.

4. Fake surfaces

Although visible, the artificial illuminants in the second block of Experiment 1 did not produce strong differences across metameric mismatches compared to naturalistic illuminants (cf. Fig. 4). These colour differences may be too small to be relevant in many everyday life situations, implying that observers might not need to develop such precise expectations about colour changes. However, human observers may also encounter artificial reflectance spectra in everyday life, especially if we consider manufactured goods, such as paints or pieces of clothing. Combining artificial illuminants with artificial reflectance spectra produces stronger metamer mismatches of naturalistic spectra than artificial illuminants alone (cf. Fig. 3.a-b). In this second experiment, we included the interaction of metameric illuminants with metameric reflectance spectra and tested whether observer expectations would be sufficiently precise to tell the naturalistic from the artificial colour changes. When manipulating reflectance spectra, artificial reflectances may interact with the illuminant so that achromatic surfaces under one illuminant become chromatic under another illuminant, hence misleading observers about the actual colour of the illuminant. For this reason, it can make a big difference whether there is a veridical

illuminant cue (e.g., a white-standard with flat reflectance spectrum) in the scene because it will indicate unexpected colour shifts of areas that were achromatic under the reference illuminant. To assess the effect of a veridical illuminant cue, we ran this Experiment without (2.a) and with (2.b) a white-standard in the rendered images. The colour of the illuminant shown by the white-standard is the same across all four test renderings because illuminants are metameric.

4.1. Method

Participants. Two different samples of 20 observers participated in Experiment 2.a (15 female, 19.80±2.66 years) and Experiment 2.b (14 female, 32.3±9.07 years). There were no exclusions from the analysis because every participant had error-free catch trials and normal colour vision. Recruitment, compensation, colour-vision screening, and ethics approval and obtaining participants' consents were the same as in Experiment 1.

Apparatus & Stimuli. The Apparatus was the same as in Experiment 1. The same fourteen scenes were used as in Experiment 1 and all reference renderings and naturalistic targets were also the same. The key difference to Experiment 1 was that the artificial distractors were created using artificial reflectance spectra. For each distractor, the naturalistic reflectances were replaced with RGB reflectance spectra (cf. Asymmetric Colour Matches and Fig. 2.c). We computed the RGB reflectance spectra for each pixel to be metameric with the corresponding pixel of the reference rendering, i.e., the naturalistic reflectance under the naturalistic illuminant (cf. top row of Fig. 5). This implied that we had to produce three different sets of reflectances to be metameric with either the yellow, neutral, or blue naturalistic illuminant. The white-standard in Experiment 2.b was simulated by replacing a uniform area in the lower left corner of the image with a flat white reflectance that showed the colour of the illuminant (cf. example in Fig. 5). However, we reduced the intensity of the white-standard so that it fits within the monitor gamut. The white-standard was consistently circular, taking up on average 0.6 % of the image area with a standard deviation of 0.1 %. The resulting artificial distractors differed much more strongly from the naturalistic target than those of Experiment 1 (cf. Fig. 4). Fig. 7.a illustrates an example with white-standard under the blue illuminant.

Procedure. The procedure remained the same as in Experiment 1, with one exception. The third block that checked visibility of metamer mismatches featured twice as many trials as in Experiment 1. This is because for every naturalistic target, there existed two sets of distractors with different reflectances. Each set was designed to be metameric with the renderings based on one of the two alternative reference illuminants. For example, if the naturalistic target was rendered under the naturalistic neutral illuminant, one set of distractors were rendered to be metameric under the yellow reference illuminant, and the other set was metameric under the blue reference illuminant. Rendering each set of references under neutral illuminants produces slightly different distractors, and we tested discrimination with both versions. So, there were overall 3 (illuminant colours) x 2 (distractor sets) x 10 (scenes in the second block) = 60 trials.

4.2. Results

Fig. 6.b-c illustrates results without and with the white standard, respectively. Like in Experiment 1, the frequency of choosing the naturalistic target was very high in the first (88 % & 89 %) and last block (98 % & 97 %) and significantly above 50 % (all t(19) > 14.63, all p < 0.001). Besides the above one-sample and paired t-tests, independent-sample t-tests were used for comparisons across experiments.

Exp. 2.a (Without White-Standard). The centre group of Fig. 6.b illustrates the main results without white-standard. As in Experiment 1, observers chose the naturalistic target (38 %) less than 50 % (t(19) = -4.98, p < 0.001), hence contradicting our prediction. Nevertheless, they

chose it above chance level (t(19) = 5.47, p < 0.001), and more often than Fundamental (t(19) = 3.78, p = 0.001) and RGB- (t (19) = 11.43, p < 0.0001) distractors. Observers confused the naturalistic target with the 3-Gaussian distractor (34 %), which was also chosen significantly above 25 % (t(19) = 5.64, p < 0.001) and did not differ from the naturalistic target (t(19) = 1.05, p = 0.30). More detailed additional analysis revealed differences depending on the type of illumination changes (see Figure S7). Under the blue test illumination, the 3-Gaussian distractors were chosen more often (77 %) than naturalistic targets (12 %) (t(19) = 8.81, p < 0.001). There was a similar tendency as in Experiment 1 of naturalistic targets being chosen more often for naturalistic scenes (41 %) and paintings (36 %), but the difference missed significance (t(19) = 2.07, p = 0.052). Also like in Experiment 1, confidence ratings were lower in the second than in the first (t(19) = -4.67, p < 0.001) and third block (t(19) = -8.66, p < 0.0001), confirming the difficulty of the second block.

Exp. 2.b (With White-Standard). Main results (second block) with the white standard are shown by the centre bars in Fig. 6.c. Again, observers' choices of the naturalistic target (53 %) were significantly above chance (t (19) = 5.52, p < 0.001). In addition, these choices were higher than in Experiment 2.a without white-standard (t (38) = 2.67, p = 0.01; cf. centre bars in Fig. 6.b-c) and higher than any one of the three distractors (all t(19) > 3.90, all p < 0.001), but not sufficiently different from 50 % (t(19) = 0.61, p = 0.55). As in Experiment 1, observers chose the naturalistic targets more often for naturalistic scenes than for paintings (57 % vs 49 %), in line with our prediction (t(19) = 2.53; p = 0.02). Confidence ratings were again lower in the second than in the two other blocks (t (19) = -4.67, p < 0.001, t (19) = -8.66, p < 0.001, first and third blocks, respectively).

Comparison with Experiment 1. Compared with the main block of Experiment 1, target choices were significantly higher in Experiment 2.b (t (38) = 3.19, p = 0.003), but this tendency was not significant for Experiment 2.a (t(38) = 0.78, p = 0.44). Confidence ratings were higher in Experiment 2.a-b than in Experiment 1 (both t (38) > 3.2, both p = 0.002).

4.3. Discussion

In Experiment 2.a, observers chose the naturalistic target slightly more often than in Experiment 1. The mere presence of a visible cue to the illumination colour in Experiment 2.b further increased the choice of the naturalistic target. In addition, Experiment 2 strengthened the evidence from Experiment 1 that observers have clearer expectations towards naturalistic scenes than towards manmade paintings. Nevertheless, observers did still not choose the naturalistic target more often than all three distractors together. Thus, expectations are not precise enough to reliably recognise the naturalistic target among the artificial distractors even when colour shifts were comparatively strong due to the interaction between artificial illuminants and reflectances (cf. Fig. 4).

What made the strong colour shifts in Experiment 2 look so plausible was the fact that all shifts seemed to be consistent across all colours in a scene. Fig. 7.a illustrates a condition with particularly strong colour shifts. Note that the RGB rendering (i.e., RGB reflectance under RGB illuminant) looks much pinker than both the other test renderings and the white-standard (bottom left). Nevertheless, it looks realistic if we assume a pinkish illumination (cf. Figure S1.a in Karimipour et al., 2023). Following the idea of relational colour constancy, we examined whether the apparent consistence of colour shifts corresponds to stable cone-excitation ratios across illumination changes (Foster & Nascimento, 1994; Foster et al., 1997; Nascimento et al., 2004; Nascimento et al., 2002; Nascimento & Foster, 1997). We quantified the deviations in cone excitation ratios for each reference/test image pair, following previous approaches (Nascimento & Foster, 1997; Nascimento & Foster, 2000; Nascimento & Foster, 2023). For the specific equations used, please refer to Table S3. The highest cone-excitation ratio deviations
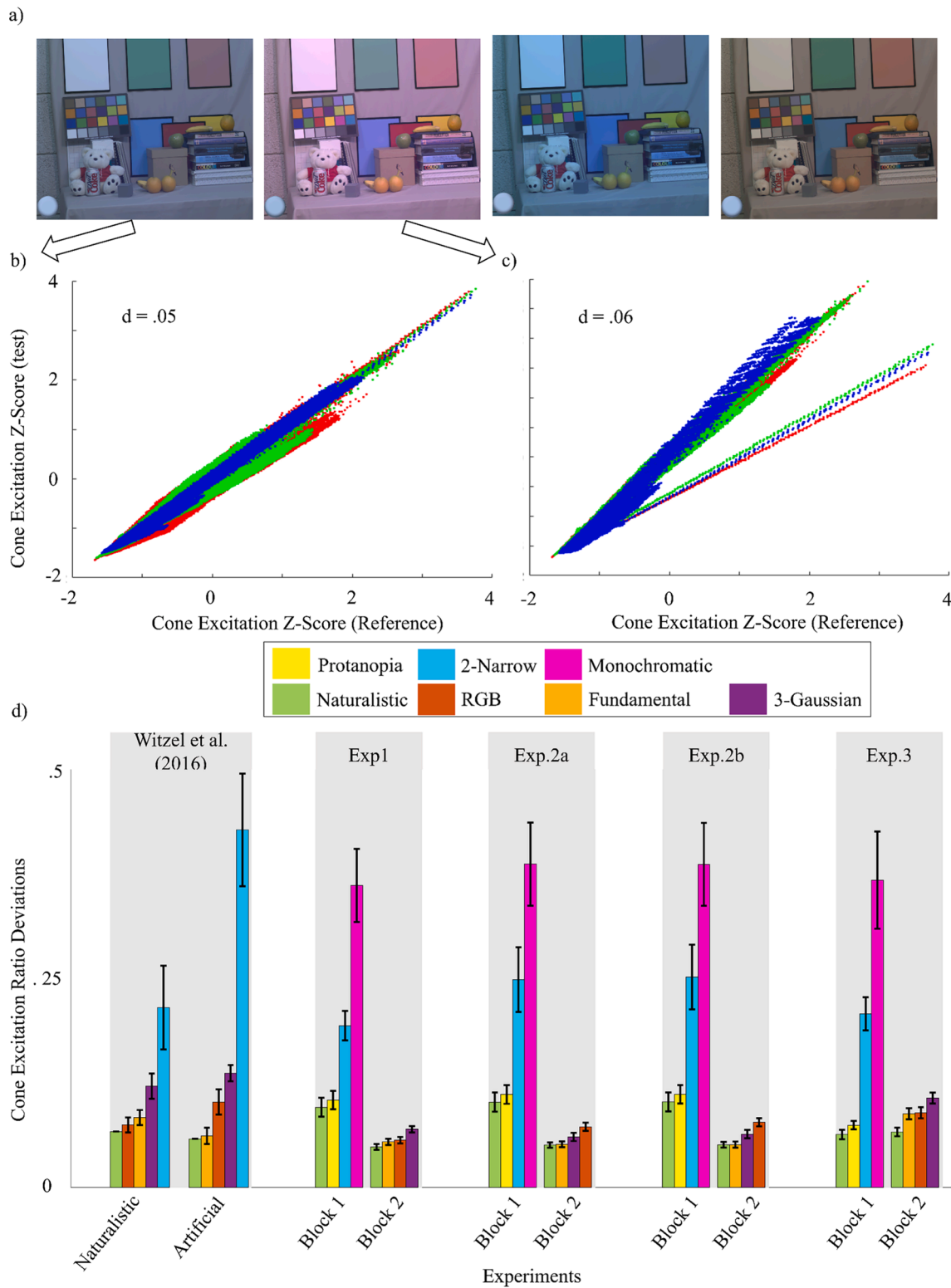
Fig. 7. Stability of Cone-Excitation Ratios across Illumination Changes. a) An example of four renderings under blue illumination from a trial of Experiment 2.b. From left to right: naturalistic target, RGB, Fundamental and 3-Gaussian distractors. Panel c and d illustrate the stability of cone-excitation ratios by showing the alignment between cone-excitations for the reference rendering along the x-axis and each of the first two test renderings based on naturalist and RGB spectra (cf. white arrows). Cone-excitations were z-scored to allow representing L-, M-, and S-cones on the same scale. Red, green, and blue dots correspond to L-, M-, and S-cones. The deviation of the cone-excitation ratios between the reference and test (d) is given in the upper left corner as an index of the stability of excitations ratios. d) Bars show the deviations of cone excitations ratios between the reference and test renderings averaged across stimuli. Those for the asymmetric colour matches from Witzel et al. (2016) were averaged across all 30 combinations of test colours and illumination changes (left group of bars); those for the first and second blocks of the experiments were averaged across all combinations of scenes and illumination changes. Bar colours correspond to types of illuminants as in the legend on top. Error bars indicate standard errors of mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

averaged across scenes and illumination colour changes occurred for the monochromatic version of the naturalistic target followed by the 2-Narrow rendering in the first block of Experiment 2.b (see corresponding pink and blue bars in Fig. 7.d; deviation = 0.35 and 0.25, respectively). The low deviations (ranging from 0.05 to 0.08) in the second block mean that all our combinations of illuminants and reflectances produced almost constant cone-excitation ratios, no matter whether they were naturalistic or artificial.

The high stability of cone-excitation ratios may partially explain why the frequencies of target choices were lower than expected (i.e., not above 50 %) in Experiment 2. According to relational colour constancy, colour shifts look plausible and realistic when cone-excitation ratios are stable (Nascimento & Foster, 1997, 2001; Nascimento & Foster, 2023). This idea implies that, contrary to our hypothesis, observers do not have specific expectations as to how illuminations shift each surface colour to specific points in colour space; instead, they would consider any naturalistic and artificial colour shift as plausible if it satisfies relational colour constancy. This idea is only partly supported by the results from block 1 (in all Experiments). On the one hand, the overall lower stability of cone-excitations is in line with the high target-choice frequency in block 1 (cf. the corresponding bars of the first blocks across experiments in Fig. 7.d). On the other hand, cone-excitation ratios are similarly stable for protanopic (yellow bars in Fig. 7.d) and naturalistic renderings (green bars), but observers seem not to confuse them (Fig. 6.b, left).

A partial contribution of relational colour constancy is further supported by the higher target-choice frequency in Experiment 2.b than in Experiment 2.a. In the naturalistic target (first image in Fig. 7.a) the surface colours change in the same direction as the illuminant. This can be seen from the alignment of all cone-excitations in Fig. 7.b. In contrast, the surface colour shift with the RGB rendering deviated from the illumination change: While the white-standard, like the illuminant, changed from yellow to blue, the RGB rendering did not change towards blue, but towards pink (cf. second image in Fig. 7.a). This discrepancy can also be seen in Fig. 7.c where the cone-excitation ratio of surface colours (broader distribution on top) does not align with the cone-excitation ratio of the white-standard (thinner lines below). This discrepancy between white-standard and other surface colours contradicts relational colour constancy and provides an important cue to the naturalism of the rendering.

The higher target-choice frequency in Experiment 2.b suggests that observers realised this discrepancy and used it as a proxy to decide whether a rendering is implausible. What is not clear yet, is whether this effect is specific to the illuminant cue provided by the white-standard, or whether it holds for any deviation from relational colour constancy. We conducted a third experiment to clarify this question.

5. Fake object surfaces

In this last experiment, we undermined constant cone-excitation ratios by manipulating the reflectance spectra of only a single object while keeping naturalistic reflectance spectra for the rest of the scene. Such viewing conditions are common in everyday life. For example, a manmade object dyed with peculiar artificial pigments may look different under the artificial illumination in the shop than under the natural daylight outside the shop. Due to the artificial pigments, the object changes in an unforeseen way while the rest of the scene, having more or less naturalistic reflectances, appears in foreseeable colours. In this situation, the observer can compare the actual colour of the object with the colour they expected based on the surrounding colours. So, the naturalistic object can be identified by inspecting the relationship between the colour of that object and its environment under the different illuminants. In the most obvious case, a colour of another object that was similar or the same colour as the target object under one illumination, may look different from the target object under an artificial illuminant. This is the case with the two halves of the Metacow (Fairchild & Johnson, 2004) or the failed retouch of Picasso's *The Tragedy* (Berns, 2016).

To make sure that participants did not oversee the relevant object, we told them to look at the target object when completing the fake-detection task. We hypothesised that the comparison between object and context allows observers to detect deviations from relational colour constancy and thus help distinguishing naturalistic from artificial colour changes. We hence anticipated a higher frequency of choosing the naturalistic target in the current experiment than in the previous ones.

5.1. Method

Participants. Twenty-three participants were involved in Experiment 3 (14 female, 24.58 ± 6.32 years). All participants demonstrated normal colour vision, and no errors were observed in the catch trials. Procedures regarding recruitment, compensation, colour vision screening, ethical approval and obtaining participants' consents adhered to the protocols established in Experiment 1.

Apparatus & Stimuli. The Apparatus was the same as in Experiment 1. Contrary to the comprehensive replacement of naturalistic spectra with artificial RGB reflectance spectra in Experiment 2, we only changed the reflectance spectra of a designated object or area within each scene in Experiment 3 (See the white areas in Figure S5.c). The original, naturalistic reflectance spectra of that object or area were replaced with artificial RGB reflectance spectra. The remaining parts of the scene were left with their original naturalistic reflectance. Most of the hyperspectral images were the same as in Experiments 1 & 2. Unfortunately, all objects and coloured areas had fuzzy edges in six of the hyperspectral images (1 image in the first block and 5 images in the second block). So, the separation of any such objects and areas from their background would create spurious artifacts. We wanted to make sure that observers judge plausibility purely based on illuminant-specific colour shifts, not on failures of object delineations. So, we replaced those six images with six new images with definable areas (three faces, two scenes with fruits and vegetables and one landscape scene). We focused on images that showed naturalistic rather than artificial, manmade content. According to the previous results this should slightly increase observer expectations. The details of these new replacement images, including their visual content and the specific objects or areas they encompass, are documented in the rightmost image of Block 1 and the last row of Block 2 in Figure S5.b of the Supplementary Material. There was no white standard in the renderings.

Procedure. The procedure was the same as in Experiments 1–2, except for the one changes of scenes in block 1 and five changes of scenes in block 2 (Figure S5.b). The first block now included the previous naturalistic scene, an image of the newly added faces, and only one of the previous (Experiment 1–2) and one newly added painting. The ten scenes in the second block featured four of the previous naturalistic scenes, three new ones including naturalistic landscape and fruits and vegetables, one of the previous paintings, and two newly added faces. The numbers of trials were the same as in Experiment 2, namely 24, 60, and 60 trials in blocks 1–3, respectively.

5.2. Results

Fig. 6.d illustrates the results of Experiment 3. As in Experiment 1 and 2, the frequency of choosing the naturalistic target in the first block (93 %) and last block (98 %) was very high and significantly above 50 % ($t(22) = 46.3$, $p < 0.001$, $t(22) = 84.57$, $p < 0.001$).

Main Results. In the main, second block, participants chose the naturalistic target with 72 % clearly above 50 % ($t(22) = 11.21$, $p < 0.001$). This was not due to the five newly added scenes, as the frequencies (69 %) remained significantly above the 50 % for only the other five (old) scenes ($t(22) = 9.95$, $p < 0.001$). The naturalistic target was chosen at least 50 % of the time for almost all (~90 %) combinations of the scenes illuminant changes(Figure S.9). Nevertheless, the frequency of choosing the naturalistic target was lower in the second block than in the first block ($t(22) = -9.61$, $p < 0.001$). Confidence ratings in

this second block were also lower than in the first (t (22) = -3.83, p < 0.001) and third blocks (t (22) = -9.04, p < 0.001). So, the second block was still more difficult than the other blocks, in which the target was obvious.

Comparison with Experiments 1 + 2. Consistent with our hypothesis, independent-sample t-tests showed that participants chose the naturalistic target in the second block of Experiment 3 more often than in the second block of Experiment 1 (t (41) = 11.82, p < 0.001), Experiment 2.a (t(41) = 11.07, p < 0.001) and Experiment 2.b (t(41) = 3.68, p < 0.001). Confidence ratings in the main block of Experiment 3 were also significantly higher than in Experiment 1 (t(41) = 3.98, p < 0.001). This tendency was not significant in comparison with Experiment 2 (t (41) = 0.65, p = 0.52; t (42) = 0.44, p = 0.66). These results were the same when only including the five hyperspectral images that were common to all three experiments.

5.3. Discussion

In sum, observers chose naturalistic targets more often in Experiment 3, where only a target object/area was manipulated, than in Experiment 1 and 2, where the whole scene was manipulated. Experiment 3 was the only experiment in which observers chose the naturalistic target significantly more often than all the distractors together (>50 %). According to our 50 %-criterion for reliability, these results indicate that observers could reliably recognise the naturalistic target. This was the case even though the colour differences between the artificial distractors and the naturalistic rendering were much smaller than of those in Experiment 2a-b (cf. Fig. 4). The higher target selection in Experiment 3 suggests that observes compare the target object/area with the surround to judge the plausibility of the colour of the target area, in line with relational colour constancy. Two caveats are important to consider when interpreting and generalising these results.

Variation of scenes across experiments. Unfortunately, we could not keep the same scenes across experiments without risking delineation artefacts. The higher target-choices in Experiment 3 could be explained by the differences between the stimulus sets. Experiment 3 features comparatively many natural objects and scenes that have memory colours, which are typical colours of objects and materials that observers learned from their life-long experience with their visual environment (for review, see Witzel & Gegenfurtner, 2018). Images with memory colours might involve stronger observer expectations. The memory colours of fruits and vegetables are so strong that they can affect colour appearance (for review, see Witzel & Gegenfurtner, 2018), and the memory colours of faces can produce paradoxical effects on colour appearance (Hasantash et al., 2019). Additional checks did not yield a significant differences between landscapes and faces (t(22) = -1.33, p = 0.20); however, the naturalistic target was chosen significantly more often in landscape scenes (80 %) compared to fruits and vegetables (53 %) (t(19) = 9.60, p < 0.001). This lower target-choice frequency with fruits and vegetables contradicts the idea that these fruits and vegetables could have caused the higher target choices in Experiment 3 compared to the other experiments. Yet, we cannot fully exclude an impact of stimulus differences.

Relational Colour Constancy. While relational colour constancy seems to generally contribute to observer expectations, it is difficult to estimate how much cone-excitation ratios affect observer responses in our experiments. The right side of Fig. 7.d illustrates the deviations of cone excitation ratios across illuminations changes in Experiment 3. The colour of the target area was manipulated to deviate from cone-excitation ratios in the artificial distractors. Nevertheless, cone-excitations were stable (deviation = 0.07-0.10), implying that this manipulation did not have a great effect on the overall stability of cone-excitation ratios. Observers are sensitive to small deviations from cone-excitation ratios (cf. Nascimento & Foster, 2023). Focussing observers' attention on the target area further helped detecting the deviation from the homogenous colour change in Experiment 3. Thus, it seems plausible

that relational colour constancy contributed to the higher target-choices in Experiment 3. However, it would be preferable to directly relate target-choices to deviations from cone-excitation ratios across experiments.

6. Factors beyond expectations

Overall, these experiments show that observers have some knowledge and expectations based on prior experience about how colours shift across illuminations. This is not only supported by the small but systematic tendency of choosing the naturalistic target, but also by the difference between naturalistic scenes and manmade paintings. However, these expectations seem not sufficient to explain reliable colour constancy, which would imply that observers have clear expectations about where exactly each colour shifts in colour space when the illumination changes. Experiment 2–3 suggest that the comparison with the surround additionally increases the identification of the naturalistic target as the most plausible colour change. Nevertheless, it remains open what exactly drives observers' decisions about the plausibility and realism of colour changes. We conducted correlational analyses to explore candidate determinants of the responses in the second block across all experiments and within each experiment.

6.1. Candidate determinants

We calculated the frequency of choosing the naturalistic target for each of the 60 stimulus conditions in each of the experimental measurements (Experiments 1, 2.a, 2.b, and 3). Candidate determinants were quantified for each condition. As observers could only choose between the 4 test images, the frequency of choosing the target depended on the difference of the target from the distractors, rather than the variation of the absolute magnitude of a determinant. For determinants where this matters, we calculated the relative predictor, which was the extent to which a determinant was present in the naturalistic target divided by the sum across all 4 tests. We considered the following factors as potential determinants:

Deviations from Cone-Excitation Ratios (*global and local CER*). Results from Experiments 2 and 3 suggested an effect of deviations from cone-excitation ratios. In one approach, we calculated deviations from constant cone-excitation ratios following an established algorithm (Nascimento & Foster, 1997; Nascimento & Foster, 2000; Nascimento & Foster, 2023). Since this is done across all pixels of each test image, we call this approach global CER (cf. Table S3). In a second approach, we considered that it might be more important how specific areas or objects in Experiments 2.b and 3 deviated from constant cone-excitation ratios. To capture such local deviations (local CER) in Experiment 2b and 3, we calculated the cone-excitation ratios as the cone-excitations of the surround divided by the local area (white-standard in Experiment 2.b, object in Experiment 3). According to relational constancy, target choices should be higher when targets deviate less from cone-excitation ratios than distractors, implying a negative correlation between relative predictor and target choices. Because the quantitative index of local CER is not defined for Experiments 1 and 2.a, we created a binary variable for tests across experiments, which is one for experiments with local areas (Experiment 2.b and 3), and zero for those without (Experiment 1 and 2.a). This binary local CER will capture the higher target choice frequencies in Experiments 2.b and 3 as a positive correlation.

Redness-Luminance Correlation (*Redness*). Alternatively, correlations between luminance and redness may indicate realistic illumination changes (cf. Introduction). We computed (see Table S3 for equation) the redness-luminance correlation for each stimulus following Golz and MacLeod (2002). If those correlations matter, the relative predictor should yield a positive correlation because target choices are expected to be higher when the target has a higher redness-luminance correlation than the distractors.

Illumination Change (*CCT*). Some of our results suggested that

observers chose naturalistic targets less often when the test illumination was blue (cf. Results for Experiment 1 and 2a). To examine the effect of the illumination change, we calculated (signed) differences in Correlated Colour Temperature (CCT) between reference and test illumination colours. Positive differences imply that the reference changed towards a more bluish (higher CCT) test, while negative differences correspond to a change towards yellow. The above results would suggest a negative correlation between CCT and target choices implying that observers choose targets less often under blue illuminations.

Surface Colour Shift (*Shift*). Apart from the illumination change, the illumination-induced shifts of colours in a scene vary depending on the spectral properties of metameric illuminants and reflectances (cf. Fig. 8.a), and the chromatic composition of the scene (cf. Figure S6). We considered that observers would pick the naturalistic target more often if it was more similar to the reference than the distractors. We calculated the average across pixels of the (three-dimensional) Euclidean distances (ΔE_{Lab}) between the reference and each of the four tests. If shifts matter, the relative predictor is expected to correlate negatively with target selection frequencies because observers should find naturalistic targets more plausible when they are more similar to the reference than the distractors are.

Visibility. Observers seem to easily discriminate the naturalistic target from the distractors in the third block of each experiment. Nevertheless, we double-checked whether variations of visibility might have affected plausibility judgements. So, we included the performance of block 3 as a candidate predictor. If visibility matters, we expected a positive correlation with block 2 because observers would confuse target and distractors if they cannot well distinguish them.

Target-Distractor Differences (*Hue, Saturation, and Lightness*). While the surface colour shift is about the difference between reference and tests, we considered that responses might also be influenced by the difference between naturalistic target and artificial distractors. Participants might be more prone to confuse a distractor with the target when the distractor is more similar in terms of hue, saturation, and lightness. We calculated the average absolute difference between target and each of the three distractors in terms of hue (azimuth), chroma (distance from achromatic axis) and L*. For each condition, we calculated relative differences (divided by sum) and took the minimum across the three distractors. We expected a positive correlation because observers would be more likely to confuse the target with one of the distractors if the difference between them is small.

Scene Type. Some of our results suggested that participants chose the naturalistic target most often when images showed naturalistic scenes. We included this candidate factor as a binary variable that is 1 for all naturalistic scenes (including faces and fruits) and 0 for artificial paintings. If this scene type matters, a positive correlation was expected.

Out-Of-Gamut Values (*OOG*). Despite scaling illuminant spectra and handpicking the scenes, it was unavoidable that 12 % of the distractor renderings (74 renderings) for the second blocks across the three experiments contained 1–14 % of pixels that were slightly out of gamut. This seemed acceptable because gamut clipping seemed invisible upon inspection. It was not meaningful to calculate a relative predictor because most naturalistic targets had zero OOGs (only 5 % contained less than 1 % of OOG pixels). For this reason, we calculated the predictor as the average number of out-of-gamut values across all pixels of distractors for each condition and tested whether they might have affected our results. If OOGs influenced our results, the relative predictor would positively correlate with target choices, implying that observers chose naturalistic targets when distractors had high levels of OOGs.

6.2. Results

We calculated a multiple regression with those 9 candidate
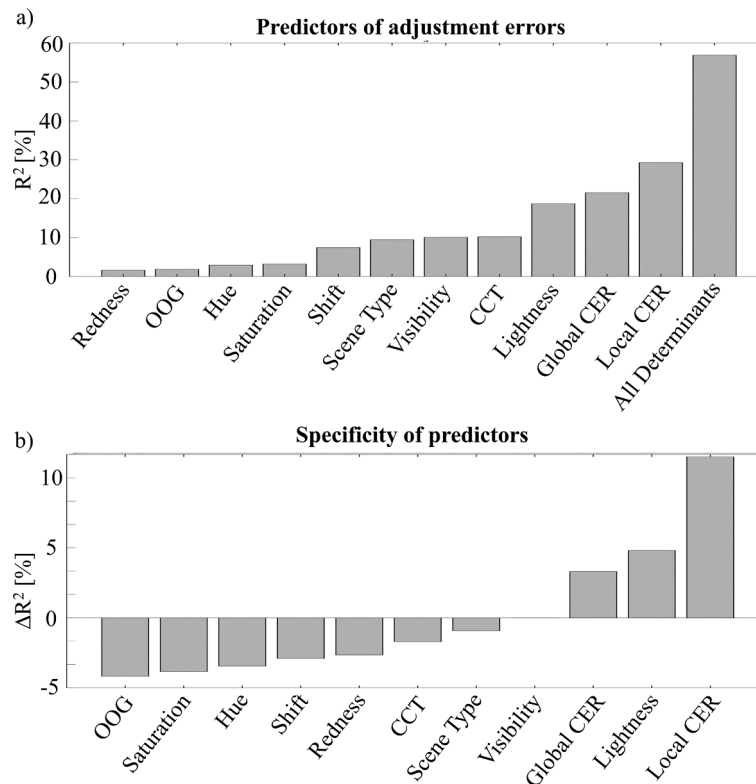
Fig. 8. Contribution of each predictor. (a) Variation in naturalistic target selections described by each predictor. The x-axis includes predictors in order of the amount of variance they explain. The y-axis shows the percentage of variance explained in a single correlation by various candidate predictors. The variation explained by a multiple regression with all determinants as predictors is shown in the right-most bar. (b) The results of the dominant analysis. The y-axis displays the average difference in R^2 in percentage between all multiple regressions with and without each predictor. The format is otherwise the same as in Panel a.

determinants as predictors of naturalistic target selection frequencies and separate correlations for each candidate. We then computed a dominance analysis that has proven useful in previous studies (Weiss et al., 2017; Witzel & Dewis, 2022). For each predictor, this analysis contrasts the explained variance of all possible models including that predictor and all models excluding that predictor. Although this approach does not clarify complex interdependencies, it gives an idea of which predictors are essential, i.e., cannot be replaced by others without reducing the explained variance. Positive values indicate essential predictors. Fig. 8 illustrates the main results. We also did these analyses for each experimental measure separately to assess the consistency of determinants (cf. Figure S8).

Across all Experiments. Fig. 8 illustrates the main results. When calculated across all 240 conditions of the four experimental measurements, the multiple regression with all predictors explained 52 % of the variance (F(11, 228) = 22.54, p < 0.001) (cf. last bar in Fig. 8.a). The binary "local CER" was positively correlated with the observer responses (r(238) = 0.54, p < 0.001) reflecting the higher target choices in Experiments 2.b and 3. This correlation was the highest among the 11 pairwise correlations, explaining more than half (29 %) of the variance achieved by the full multiple regression. It was also identified as the most essential among the 11 predictors in the dominance analysis (Fig. 8.b). The"global CER" was negatively correlated with observer responses (r(238) = -0.46, p < 0.001), in line with the idea that observers chose the naturalistic target more often, the less it deviates from constant cone-excitation ratios (relative to distractors). It explained 22 % of the response variation and was the second most relevant predictor according to the multiple regression (Fig. 8.a) and the third most essential predictor in the dominance analysis (Fig. 8.b). Similarity to the naturalistic target in terms of lightness (L*) explained 19 % of variance (r(238) = 0.43, p < 0.001) and was the third most important predictor according to the multiple regression (Fig. 8.a) and the second essential predictor according to dominance analysis (Fig. 8.b). All other predictors seemed to be dispensable (Fig. 8.b). A multiple regression combining the essential predictors ("local CER", "global CER", and "lightness") explained 47 % of the variation in target selection frequencies (F(3, 236) = 69, p < 0.001). Additional analyses showed that including three-dimensional Euclidean distances in CIELAB as a single predictor produced worse results than hue, saturation, and lightness as separate predictors.

For each Experiment. Figure S8 illustrates analyses for each experimental measure. Experiments differed strongly in the variance explained by the remaining 11 predictors: While all predictors only explained 40 % variance for Experiment 3 (F(11,48) = 2.94, p = 0.005; cf. Figure S8.g), they reached 82 % for Experiment 2.a (F (11,48) = 22.20, p < 0.001; Figure S8.c)."Global CER" was the most important predictor only for Experiment 3 (19 %, r(58) = -0.44, p < 0.001; cf. Figure S8.g-h). In Experiment 2.a-b the illumination change (68 %, r (58) = -0.82, p < 0.001; 22 %, r(58) = -0.47, p < 0.001, respectively), saturation shift (48 %, and r (58) = .69, p < 0.001; 22 % and r(58) = -0.47, p < 0.001, respectively) and Lightness similarity to the naturalistic target (55 %, and r(58) = 0.74, p < 0.001; 16 %, and r(58) = 0.40, p = 0.001; Figure S8.c-f) were most important. In Experiment 1, it was the redness correlation (22 %, r(58) = 0.47, p < 0.001; Figure S8.a-b), which went in the predicted positive direction. Our quantitative "Local CER" was only important in Experiment 3 (11 %, r(58) = 0.33, P = 0.01) but not in the others (cf. Figure S8). In all analyses, out-of-gamut values (OOG) were among the least important predictors, suggesting that the results were not affected by unavoidable limitations of hyperspectral rendering.

6.3. Discussion

The correlational analyses suggest that deviations from cone-excitation ratios (global and local CER) and lightness similarity (Lightness) account for differences across Experiments (Fig. 8). However, the

source of response variation within each Experiment is more complicated. The global and local CER did not contribute much to the variation within each experiment, except for Experiment 3 (Figure S8). Lightness was important for Experiment 2.a, but much less for the others. The "Lightness" factor suggested that observers tended to choose the naturalistic target less often when one of the distractors was similar to the naturalistic target in terms of lightness. The finding that observers' found more naturalistic-looking distractors more plausible than less naturalistic-looking ones seems in line with our hypothesis. However, the role of cone-excitation ratios and their link with expectations requires further considerations.

Cone-Excitation Ratios. Deviations from cone-excitations reflected the difference of the overall colour shift from the illumination shift revealed by the white-standard in Experiment 2.b (cf. Fig. 7.c), and by the manipulation of single areas in Experiment 3. This explains why global and local deviations from cone-excitation ratios (CER) matter in Experiment 3 and across experiments. It is also understandable that global CERs matter less in Experiment 1 and 2.a because all conditions are very close to constant cone ratios, leaving little variance to explain the variation of observer responses (cf. Fig. 7.d). It is unexpected that global and local CER do not matter that much for responses in Experiment 2.b. The lack of a correlation is in line with our previous observation that relational colour constancy cannot explain all results.

Nevertheless, these additional analyses highlight that stable cone-excitation ratios partially contribute to observer judgements. This conclusion is also supported by previous observations that observers tend to find image changes with fully constant cone-excitation ratios most natural or plausible (Nascimento & Foster, 2000; Nascimento & Foster, 2023). We observed that cone-excitation ratios were not only stable for naturalistic colour changes, but also for different kinds of artificial colour changes (Fig. 7.d). This includes colour changes involving artificial reflectances, but not always. Mixing different types of reflectances like in Experiment 2.b (cf. Fig. 7.c) and 3 can produce deviations from cone-excitations (cf. Fig. 7.d). In contrast, it seems that cone-excitation ratios remain constant when reflectance spectra are of a similar type, in the sense that they can be decomposed into the same or similar basis functions, as in Experiment 2.a, which may happen, for example, with acrylic paints (García-Beltrán et al., 1998). Naturalistic reflectances can also be decomposed into a few basis functions (Maloney & Wandell, 1986), which could explain their stable cone-excitation ratios. Although, surface colour shifts differ strongly between different sets of such reflectances, such as our naturalistic and artificial reflectances, the colour shifts within each set were homogeneous, in the sense that cone-excitation ratios were very stable (See block 2 across Experiments 1 and 2a in Fig. 7.d). This likely explains why observers sometimes confused naturalistic and artificial renderings in Experiments 1 and 2.

Mechanisms. It is not clear what exactly causes the effect of stable cone-excitation ratios on naturalistic-target choices. It would be in line with our hypothesis that observers expect homogeneous colour shifts in line with stable cone-excitation ratios based on their everyday-life experience. However, sensory colour-constancy mechanisms may also explain those effects as temporal chromatic adaptation and local contrast induction shift colour appearance in such a homogenous way (Hansen et al., 2007; Kraft & Brainard, 1999; for review, see, e.g., Smithson, 2005; Witzel and Gegenfurtner, 2018; Foster, 2011). For example, a von-Kries adaptation transform simulates chromatic adaptation by correcting scene renderings based on the cone-excitation ratios of the illuminant (Foster et al., 1997). The simultaneous comparison in our experiments undermined the effect of adaptation across time on those comparisons, and the polychromatic content of the complex scenes did not provide surrounds that reflected the illumination colour. Nevertheless, the higher frequencies of naturalistic target selections in Experiments 2.b and 3 could be explained by local contrast induction between areas and their surround. There could also be partial adaptation when observers move their eyes between target and distractors. The lack of the full effects of those sensory mechanisms in our displays could even

have contributed to observers not always being able to identify the naturalistic target in our experiments. In addition, the sensory mechanisms and observer expectations are not mutually exclusive; on the contrary, the sensory mechanisms may determine what colour shifts observers expect when the illumination changes. Thus, the observed effects of stable cone-excitation ratios may be the results of prior expectations, sensory mechanisms, or a combination of both. It is clear, however, that the consistency of colour shifts is important when observers judge illumination changes.

The Real World. The ability of human observers to detect artificial colour changes in two-dimensional images might not be as pronounced as in real-world situations. Real-world colour constancy is supported by factors that are absent in the images we used here (for review, see Foster, 2011; Witzel & Gegenfurtner, 2018), such as three-dimensional inter-reflections and shading (Bloj et al., 1999), depth segmentation (Werner, 2006), polychromatic illuminants (Ennis & Doerschner, 2019; Granzier et al., 2014; Lee & Smithson, 2016; Witzel, O'Regan, & Hansmann-Roth, 2017; Yang & Maloney, 2001), movements and specularities (Ennis & Doerschner, 2019; Granzier et al., 2014; Lee & Smithson, 2016; Wedge-Roberts et al., 2020; Witzel, O'Regan, & Hansmann-Roth, 2017; Yang & Maloney, 2001). In addition, real-world colour constancy involves chromatic adaptation, which is one of the most important contributors to colour constancy (Hansen et al., 2007; Kraft & Brainard, 1999; for review, see, e.g., Smithson, 2005; Witzel and Gegenfurtner, 2018; Foster, 2011). Observers might perform better in detecting or perceiving artificial lighting in the real world, where all these factors are present. Thus, our measurements with two-dimensional images may be understood as a lower boundary estimation of observer expectations. If this is so, it would further support our conclusion that human observers have expectations about surface colour shifts. An important part of these expectations is that homogenous colour shifts are in line with relational colour constancy; but from the reanalyses of the asymmetric matches (Witzel et al., 2016) and from Experiments 1–2, we take that observers also have a rough idea about the location in colour space to which surface colours typically shift in naturalistic environments.

7. Conclusion

We investigated whether observers have expectations about where in colour space a surface colour is shifted when the illumination changes. The current findings showed that observers have some expectations towards the direction of colour changes, but such expectations are not very certain and precise. In addition, relational colour constancy turned out to be important for observer expectations, too. In contrast to expectations about specific surface colour shifts, relational colour constancy only requires that those colour shifts are homogenous, i.e., produce approximately constant cone-excitation ratios. We observed that cone-excitation ratios are approximately constant, not only across naturalistic, but also across diverse artificial illuminants. This observation indicates that observers can achieve high levels of relational colour constancy under all kinds of naturalistic and artificial lighting conditions. Hence, our findings suggest that a combination of prior knowledge about surface colour shifts, and relational colour constancy help to disambiguate surface colour identity under illumination changes and allow human observers to recognise surface colours with high reliability in naturalistic conditions. In addition, relational colour constancy may even be effective in many artificial conditions.

CRediT authorship contribution statement

Hamed Karimipour: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Christoph Witzel:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data have been shared in Zenodo

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.visres.2024.108451.

References

Akbarinia, A., & Gegenfurtner, K. R. (2018). Color metamerism and the structure of illuminant space. *Journal of the Optical Society of America A, 35*(4), B231–B238. https://doi.org/10.1364/JOSAA.35.00B231

Arend, L., & Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America a-Optics Image Science and Vision, 3*(10), 1743–1751. https://doi.org/10.1364/josaa.3.001743

Berns, R. S. (2016). *Color science and the visual arts : A guide for conservators, curators, and the curious.* The Getty Conservation Institute Los Angeles.

Bloj, M. G., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature, 402*(6764), 877–879. https://doi.org/10.1038/47245

Brainard, D. (1997). *Hyperspectral Image Data.* http://color.psych.upenn.edu//hyperspectral/.

Brainard, D. H. (1998). Color constancy in the nearly natural image. 2 Achromatic loci. *Journal of the Optical Society of America a-Optics Image Science and Vision, 15*(2), 307–325. https://doi.org/10.1364/Josaa.15.000307

Brainard, D. H., Brunt, W. A., & Speigle, J. M. (1997). Color constancy in the nearly natural image.1. Asymmetric matches. *Journal of the Optical Society of America a-Optics Image Science and Vision, 14*(9), 2091–2110. https://doi.org/10.1364/josaa.14.002091

Buijs, A. E. (2009). Lay people's images of nature: Comprehensive frameworks of values, beliefs, and value orientations. *Society & Natural Resources, 22*(5), 417–432. https://doi.org/10.1080/08941920801901335

Chiao, C. C., Cronin, T. W., & Osorio, D. (2000). Color signals in natural scenes: Characteristics of reflectance spectra and effects of natural illuminants. *Journal of the Optical Society of America a-Optics Image Science and Vision, 17*(2), 218–224. https://doi.org/10.1364/Josaa.17.000218

Chiao, C. C., Osorio, D., Vorobyev, M., & Cronin, T. W. (2000). Characterization of natural illuminants in forests and the use of digital video data to reconstruct illuminant spectra. *Journal of the Optical Society of America a-Optics Image Science and Vision, 17*(10), 1713–1721. https://doi.org/10.1364/Josaa.17.001713

Cohen, J. B., & Kappauf, W. E. (1982). Metameric color stimuli, fundamental metamers, and Wyszecki's metameric blacks. *American Journal of Psychology, 95*(4), 537–564. https://doi.org/10.2307/1422186

Cohen, J. B., & Kappauf, W. E. (1985). Color mixture and fundamental metamers - theory, algebra, geometry application. *American Journal of Psychology, 98*(2), 171–259. https://doi.org/10.2307/1422442

Dantec, M., Allain, H., Yvert, N., Sigrist, J., Bensafi, M., & Lafraire, J. (2022). Olfactory attributes and colors associated with naturalness. *Food Quality and Preference, 98*, Article 104495. https://doi.org/10.1016/j.foodqual.2021.104495

David, A. (2014). Color fidelity of light sources evaluated over large sets of reflectance samples. *LEUKOS, 10*(2), 59–75. https://doi.org/10.1080/15502724.2013.844654

de Almeida, V. M., & Nascimento, S. M. (2009). Perception of illuminant colour changes across real scenes. *Perception, 38*(8), 1109–1117. https://doi.org/10.1068/p6277

Ennis, R., & Doerschner, K. (2019). Disentangling simultaneous changes of surface and illumination. *Vision Research, 158*, 173–188. https://doi.org/10.1016/j.visres.2019.02.004

Esposito, T., Royer, M. P., & Houser, K. W. (2022). Measures of an illuminant-induced metameric mismatch: Theory, comparative analysis, and implications for application. *Optics Express, 30*(9), 14686–14708. https://doi.org/10.1364/Oe.451626

Fairchild, M. D., & Johnson, G. M. (2004, November 9, 2004). METACOW: A Public-Domain, High-Resolution, Fully-Digital, Noise-Free, Metameric, Extended-Dynamic-Range, Spectral Test Target for Imaging System Analysis and Simulation The Twelfth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications (CIC 2004), Scottsdale, Arizona, USA. https://repository.rit.edu/other/107.

Foster, D. H. (2011). Color constancy. *Vision Research*. https://doi.org/10.1016/j.visres.2010.09.006

Foster, D. H., & Amano, K. (2019). Hyperspectral imaging in color vision research: Tutorial. *Journal of the Optical Society of America a-Optics Image Science and Vision, 36*(4), 606–627. https://doi.org/10.1364/Josaa.36.000606

Foster, D. H., Amano, K., Nascimento, S. M., & Foster, M. J. (2006). Frequency of metamerism in natural scenes. *Journal of the Optical Society of America. A, Optics, image science, and vision, 23*(10), 2359–2372. https://doi.org/10.1364/josaa.23.002359

Foster, D. H., Amano, K., & Nascimento, S. M. C. (2006). Color constancy in natural scenes explained by global image statistics. *Visual Neuroscience, 23*(3–4), 341–349. https://doi.org/10.1017/S0952523806233455

Foster, D. H., Amano, K., & Nascimento, S. M. C. (2016). Time-lapse ratios of cone excitations in natural scenes. *Vision Research, 120*, 45–60. https://doi.org/10.1016/j.visres.2015.03.012

Foster, D. H., & Nascimento, S. M. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings of the Royal Society B: Biological Sciences, 257*(1349), 115–121. https://doi.org/10.1098/rspb.1994.0103

Foster, D. H., Nascimento, S. M. C., Craven, B. J., Linnell, K. J., Cornelissen, F. W., & Brenner, E. (1997). Four issues concerning colour constancy and relational colour constancy [Research Support, Non-U.S. Gov't]. *Vision Research, 37*(10), 1341-1345. http://www.ncbi.nlm.nih.gov/pubmed/9205725.

García-Beltrán, A., Nieves, J. L., Hernández-Andrés, J., & Romero, J. (1998). Linear bases for spectral reflectance functions of acrylic paints. *Color Research & Application, 23*(1), 39–45. https://doi.org/10.1002/(SICI)1520-6378(199802)23:1%3C39::AID-COL6%3E3.0.CO;2-4

Gegenfurtner, K. R., Weiss, D., & Bloj, M. (2024). Color constancy in real-world settings. *Journal of Vision, 24*(2), 12. https://doi.org/10.1167/jov.24.2.12

Godau, C., & Funt, B. (2012). The Logvinenko object color atlas in practice. *Color Research & Application, 37*(2), 117–125. https://doi.org/10.1002/col.20680

Golz, J. (2008). The role of chromatic scene statistics in color constancy: Spatial integration. *Journal of Vision, 8*(13). https://doi.org/10.1167/8.13.6

Golz, J., & MacLeod, D. I. (2002). Influence of scene statistics on colour constancy. *Nature, 415*(6872), 637–640. https://doi.org/10.1038/415637a

Granzier, J. J. M., Brenner, E., & Smeets, J. B. (2009). Can illumination estimates provide the basis for color constancy? *Journal of Vision, 9*(3), 18. https://doi.org/10.1167/9.3.18

Granzier, J. J. M., Vergne, R., & Gegenfurtner, K. R. (2014). The effects of surface gloss and roughness on color constancy for real 3-D objects. *Journal of Vision, 14*(2). https://doi.org/10.1167/14.2.16

Guo, X., & Houser, K. (2004). A review of colour rendering indices and their application to commercial light sources. *Lighting Research & Technology, 36*(3), 183–197. https://doi.org/10.1191/1365782804li112oa

Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision, 7*(4), 1–15. https://doi.org/10.1167/7.4.2

Hasantash, M., Lafer-Sousa, R., Afraz, A., & Conway, B. R. (2019). Paradoxical impact of memory on color appearance of faces. *Nature Communications, 10*(1), 3010. https://doi.org/10.1038/s41467-019-10073-8

Houser, K., Mossman, M., Smet, K., & Whitehead, L. (2016). Tutorial: Color rendering and its applications in lighting. *LEUKOS, 12*(1–2), 7–26. https://doi.org/10.1080/15502724.2014.989802

Houser, K. W., Wei, M., David, A., Krames, M. R., & Shen, X. S. (2013). Review of measures for light-source color rendition and considerations for a two-measure system for characterizing color rendition. *Optics Express, 21*(8), 10393–10411. https://doi.org/10.1364/OE.21.010393

Hurlbert, A. (2019). Challenges to color constancy in a contemporary light. *Current Opinion in Behavioral Sciences, 30*, 186–193. https://doi.org/10.1016/j.cobeha.2019.10.004

Jost-Boissard, S., Avouac, P., & Fontoynont, M. (2015). Assessing the colour quality of LED sources: Naturalness, attractiveness, colourfulness and colour difference. *Lighting Research & Technology, 47*(7), 769–794. https://doi.org/10.1177/1477153514555882

Judd, D. B., MacAdam, D. L., Wyszecki, G., Budde, H. W., Condit, H. R., Henderson, S. T., & Simonds, J. L. (1964). Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America, 54*(8), 1031–1040. https://doi.org/10.1364/JOSA.54.001031

Karimipour, H., O'Regan, J. K., & Witzel, C. (2023). Sensory representation of surface reflectances: Assessments with hyperspectral images. *Journal of the Optical Society of America A, 40*(3), A183–A189. https://doi.org/10.1364/JOSAA.477276

Kleynhans, T., Messinger, D. W., & Delaney, J. K. (2020). Towards automatic classification of diffuse reflectance image cubes from paintings collected with hyperspectral cameras. *Microchemical Journal, 157*. https://doi.org/10.1016/j.microc.2020.104934

Kohonen, O., Parkkinen, J., & Jääskeläinen, T. (2006). Databases for spectral color science. *Color Research and Application, 31*(5), 381–390. https://doi.org/10.1002/col.20244

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences USA, 96*(1), 307–312. https://doi.org/10.1073/pnas.96.1.307

Lee, R. J., & Smithson, H. E. (2016). Low levels of specularity support operational color constancy, particularly when surface and illumination geometry can be inferred. *Journal of the Optical Society of America. A, Optics, image science, and vision, 33*(3), A306–A318. https://doi.org/10.1364/JOSAA.33.00A306

Ling, Y. Z., & Hurlbert, A. (2008). Role of color memory in successive color constancy. *Journal of the Optical Society of America a-Optics Image Science and Vision, 25*(6), 1215–1226. https://doi.org/10.1364/Josaa.25.001215

Linhares, J. M. M., Pinto, P. D., & Nascimento, S. M. C. (2008). The number of discernible colors in natural scenes. *Journal of the Optical Society of America a-Optics Image Science and Vision, 25*(12), 2918–2924. https://doi.org/10.1364/Josaa.25.002918

Logvinenko, A. D., Funt, B., & Godau, C. (2014). Metamer mismatching. *IEEE Transactions on Image Processing, 23*(1), 34–43. https://doi.org/10.1109/TIP.2013.2283148

Logvinenko, A. D., Funt, B., Mirzaei, H., & Tokunaga, R. (2015). Rethinking colour constancy. *PLoS One1, 10*(9), e0135029.

Lucassen, M. P., Gevers, T., Gijsenij, A., & Dekker, N. (2013). Effects of chromatic image statistics on illumination induced color differences. *Journal of the Optical Society of America A, 30*(9), 1871–1884. https://doi.org/10.1364/JOSAA.30.001871

Lucassen, M. P., & Walraven, J. (1996). Color constancy under natural and artificial illumination. *Vision Research, 36*(17), 2699–2711. https://doi.org/10.1016/0042-6989(95)00346-0

Mackiewicz, M., Rivertz, H. J., & Finlayson, G. (2019). Spherical sampling methods for the calculation of metamer mismatch volumes. *Journal of the Optical Society of America a-Optics Image Science and Vision, 36*(1), 96–104. https://doi.org/10.1364/Josaa.36.000096

Maloney, L. T., & Wandell, B. A. (1986). Color constancy - a method for recovering surface spectral reflectance. *Journal of the Optical Society of America a-Optics Image Science and Vision, 3*(1), 29–33. https://doi.org/10.1364/Josaa.3.000029

Masuda, O., & Nascimento, S. M. C. (2013). Best lighting for naturalness and preference. *Journal of Vision, 13*(7), 4. https://doi.org/10.1167/13.7.4

Murray, I. J., Daugirdiene, A., Vaitkevicius, H., Kulikowski, J. J., & Stanikunas, R. (2006). Almost complete colour constancy achieved with full-field adaptation. *Vision Research, 46*(19), 3067–3078. https://doi.org/10.1016/j.visres.2006.03.011

Nascimento, S. M., de Almeida, V. M., Fiadeiro, P. T., & Foster, D. H. (2004). Minimum-variance cone-excitation ratios and the limits of relational color constancy. *Visual Neuroscience, 21*(3), 337–340. https://doi.org/10.1017/s095252380421327x

Nascimento, S. M., Ferreira, F. P., & Foster, D. H. (2002). Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America. A, Optics, image science, and vision, 19*(8), 1484–1490. https://doi.org/10.1364/josaa.19.001484

Nascimento, S. M., & Foster, D. H. (1997). Detecting natural changes of cone-excitation ratios in simple and complex coloured images. *Proceedings of the Royal Society B: Biological Sciences, 264*(1386), 1395–1402. https://doi.org/10.1098/rspb.1997.0194

Nascimento, S. M., & Foster, D. H. (2001). Detecting changes of spatial cone-excitation ratios in dichoptic viewing. *Vision Research, 41*(20), 2601–2606. https://doi.org/10.1016/S0042-6989(01)00142-0

Nascimento, S. M. C., & Foster, D. H. (2000). Relational color constancy in achromatic and isoluminant images. *Journal of the Optical Society of America a-Optics Image Science and Vision, 17*(2), 225–231. https://doi.org/10.1364/Josaa.17.000225

Nascimento, S. M. C., & Foster, D. H. (2023). Misidentifying illuminant changes in natural scenes due to failures in relational colour constancy. *Proceedings of the Royal Society B: Biological Sciences, 290*(2011), 20231676. https://doi.org/10.1098/rspb.2023.1676

Nascimento, S. M. C., Linhares, J. M. M., Montagner, C., Joao, C. A. R., Amano, K., Alfaro, C., & Bailao, A. (2017). The colors of paintings and viewers' preferences. *Vision Research, 130*, 76–84. https://doi.org/10.1016/j.visres.2016.11.006

Nascimento, S. M. C., & Masuda, O. (2012). Psychophysical optimization of lighting spectra for naturalness, preference, and chromatic diversity. *Journal of the Optical Society of America A, 29*(2), A144–A151. https://doi.org/10.1364/JOSAA.29.00A144

Nascimento, S. M. C., & Masuda, O. (2014). Best lighting for visual appreciation of artistic paintings–experiments with real paintings and real illumination. *Journal of the Optical Society of America A Opt Image Sci Vis, 31*(4), A214–A219. https://doi.org/10.1364/JOSAA.31.00A214

Osorio, D., & Bossomaier, T. R. J. (1992). Human cone-pigment spectral sensitivities and the reflectances of natural surfaces. *Biological Cybernetics, 67*(3), 217–222. https://doi.org/10.1007/Bf00204394

Parkkinen, J. P. S., Hallikainen, J., & Jaaskelainen, T. (1989). Characteristic spectra of munsell colors. *Journal of the Optical Society of America a-Optics Image Science and Vision, 6*(2), 318–322. https://doi.org/10.1364/JOSAA.6.000318

Philipona, D. L., & O'Regan, J. K. (2006). Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Visual Neuroscience, 23*(3–4), 331–339. https://doi.org/10.1017/S0952523806233182

Pinto, P. D., Felgueiras, P. E., Linhares, J. M., & Nascimento, S. M. C. (2010). Chromatic effects of metamers of D65 on art paintings. *Ophthalmic & Physiological Optics, 30*(5), 632–637. https://doi.org/10.1111/j.1475-1313.2010.00726.x

Pinto, P. D., Linhares, J. M., Carvalhal, J. A., & Nascimento, S. M. (2006). Psychophysical estimation of the best illumination for appreciation of Renaissance paintings. *Visual Neuroscience, 23*(3–4), 669–674. https://doi.org/10.1017/S0952523806233340

Radonjic, A., Cottaris, N. P., & Brainard, D. H. (2015). Color constancy supports cross-illumination color selection. *Journal of Vision, 15*(6). https://doi.org/10.1167/15.6.13

Romero, J., GarciaBeltran, A., & HernandezAndres, J. (1997). Linear bases for representation of natural and artificial illuminants. *Journal of the Optical Society of America a-Optics Image Science and Vision, 14*(5), 1007–1014. https://doi.org/10.1364/Josaa.14.001007

Royer, M. (2022). Editorial: What can we learn from the lighting in homes? *Lighting Research & Technology, 54*(5), 411. https://doi.org/10.1177/14771535221115944

Rozin, P., Fischler, C., & Shields-Argelès, C. (2012). European and American perspectives on the meaning of natural. *Appetite, 59*(2), 448–455. https://doi.org/10.1016/j.appet.2012.06.001

Samadzadegan, S., & Urban, P. (2013). Spatially Resolved Joint Spectral Gamut Mapping and Separation. *Color and Imaging Conference, 2013*(1), 2–7. https://doi.org/10.2352/CIC.2013.21.1.art00002

Shi, L., Babaei, V., Kim, C., Foshey, M., Hu, Y., Sitthi-Amorn, P., Rusinkiewicz, S., & Matusik, W. (2018). Deep multispectral painting reproduction via multi-layer, custom-ink printing. *ACM Transactions on Graphics, 37*(6). https://doi.org/10.1145/3272127.3275057

Smet, K., & Hanselaer, P. (2016). Memory and preferred colours and the colour rendition of white light sources. *Lighting Research & Technology, 48*(4), 393–411. https://doi.org/10.1177/1477153514568584

Smithson, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1458), 1329–1346. https://doi.org/10.1098/rstb.2005.1633

Wedge-Roberts, R., Aston, S., Beierholm, U., Kentridge, R., Hurlbert, A., Nardini, M., & Olkkonen, M. (2020). Specular highlights improve color constancy when other cues are weakened. *Journal of Vision, 20*(12). https://doi.org/10.1167/jov.20.12.4

Weiss, D., Witzel, C., & Gegenfurtner, K. (2017). Determinants of Colour Constancy and the Blue Bias. *i-Perception, 8*(6). https://doi.org/10.1177/2041669517739635

Werner, A. (2006). The influence of depth segmentation on colour constancy. *Perception, 35*(9), 1171–1184. https://doi.org/10.1068/p5476

Witzel, C., Cinotti, F., & O'Regan, J. K. (2015). What determines the relationship between color naming, unique hues, and sensory singularities: Illuminations, surfaces, or photoreceptors? *Journal of Vision, 15*(8), 19. https://doi.org/10.1167/15.8.19

Witzel, C., & Dewis, H. (2022). Why bananas look yellow: The dominant hue of object colours. *Vision Research, 200*, Article 108078. https://doi.org/10.1016/j.visres.2022.108078

Witzel, C., & Gegenfurtner, K. R. (2018). Color Perception: Objects, Constancy, and Categories. *Annual Review of Vision Science, 4*(1), 475–499. https://doi.org/10.1146/annurev-vision-091517-034231

Witzel, C., O'Regan, J. K., & Hansmann-Roth, S. (2017). The dress and individual differences in the perception of surface properties. *Vision Research, 141*(Supplement C), 76–94. https://doi.org/10.1016/j.visres.2017.07.015

Witzel, C., Racey, C., & O'Regan, J. K. (2017). The most reasonable explanation of "the dress": Implicit assumptions about illumination. *Journal of Vision, 17*(2), 1. https://doi.org/10.1167/17.2.1

Witzel, C., van Alphen, C., Godau, C., & O'Regan, J. K. (2016). Uncertainty of sensory signal explains variation of color constancy. *Journal of Vision, 16*(15), 8. https://doi.org/10.1167/16.15.8

Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vision Research, 41*(20), 2581–2600. https://doi.org/10.1016/S0042-6989(01)00143-2

Yendrikhovskij, S. N., Blommaert, F. J. J., & de Ridder, H. (1999). Color reproduction and the naturalness constraint. *Color Research and Application, 24*(1), 52–67. https://doi.org/10.1002/(Sici)1520-6378(199902)24:1%3C52::Aid-Col10%3E3.0.Co;2-4

Zhai, Q., Luo, M. R., & Liu, X. (2016). The impact of LED lighting parameters on viewing fine art paintings. *Lighting Research & Technology, 48*(6), 711–725. https://doi.org/10.1177/1477153515578468

Zhang, X., Funt, B., & Mirzaei, H. (2016). Metamer mismatching in practice versus theory. *Journal of the Optical Society of America A, 33*(3), A238–A247. https://doi.org/10.1364/JOSAA.33.00A238