https://doi.org/10.2352/issn.2694-118X.2020.LIM-44
This work is licensed under the Creative Commons Attribution 4.0 International License.
To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

Surface color under environmental illumination

H. E. Smithson and T. Morimoto, Department of Experimental Psychology, University of Oxford, UK

Abstract

Objects in real three-dimensional environments receive illumination from all directions, characterized in computer graphics by an environmental illumination map. The spectral content of this illumination can vary widely with direction [1], which means that the computational task of recovering surface color under environmental illumination cannot be reduced to correction for a single illuminant. We report the performance of human observers in selecting a target surface color from three distractors, one rendered under the same environmental illumination as the target, and two rendered under a different environmental illumination. Surface colors were selected such that, in the vast majority of trials, observers could identify the environment that contained non-identical surface colors, and color constancy performance was analyzed as the percentage of correct choices between the remaining two surfaces. The target and distractor objects were either matte or glossy and presented either with surrounding context or in a dark void. Mean performance ranged from 70% to 80%. There was a significant improvement in the presence of context, but no difference for matte and glossy stimuli, and no interaction between gloss and context. Analysis of trial-by-trial responses showed a dependence on the statistical properties of previously viewed images. Such analyses provide a means of investigating mechanisms that depend on environmental features, and not only on the properties of the instantaneous proximal image.

Background

One distinguishing feature of objects is their spectrally-selective surface reflectance – the proportion of incident light reflected from the surface as a function of wavelength, $R(\lambda)$. The perceptual correlate of this feature is surface colour, and significant effort has been invested in understanding the perceptual mechanisms of colour constancy [2-4], which describes the extent to which objects retain a constant colour appearance despite changes in the spectrum of the incident illumination, $I(\lambda)$. Much of the early work on colour constancy considered matte surfaces under uniform illumination in a two-dimensional world [5]. In this limited case, the signals available to the viewer for a set of n surfaces, $R_i(\lambda)$, where $i=1,\ldots,n$, are fully characterized by the cone signals from the diffuse reflectances, given by the product $I(\lambda)R_i(\lambda)$. More recent work has additionally considered surfaces that exhibit specular reflection, and has situated these surfaces under complex illumination in a three-dimensional world. Under such conditions, the proximal image at the eye will be dependent on the three-dimensional geometry of the scene, encompassing the relative locations of light sources and surfaces, and the viewpoint of the observer. The signal at each location, j, in the image of a surface, i, is a weighted sum of diffuse and specular reflections, $\alpha_j I_j(\lambda)R_i(\lambda)+\beta_j I_j(\lambda)$, which may additionally vary with spatial variation in the spectral composition of the illuminant reaching the corresponding point on the surface, $I_j(\lambda)$. Such conditions elicit the percept of gloss, and the image of the surface will contain a range of chromaticities (Figure 1).

Images of glossy surfaces under a single illuminant contain chromaticities that are linear mixtures of the chromaticities of the diffuse and specular components (Figure 1a). Images of glossy surfaces under complex illuminations typically have a wider chromatic locus, reflecting the range of spectral distributions in the illuminant (Figure 1b).

In the present paper, we quantify the performance of human observers in a task that requires them to identify surfaces across different illuminants, on the basis of spectrally-selective surface reflectance. We adopt a performance-based measure of colour constancy [6]. Rather than asking observers to match the appearance of two samples across illuminants, we measure their ability to identify the same surface across illuminants, compared to a distractor that differs in spectrally-selective surface reflectance. The complexities of matching colour appearance for surfaces that contain a range of chromaticities has been highlighted before [7]. A performance-based measure avoids this issue, and measures instead a

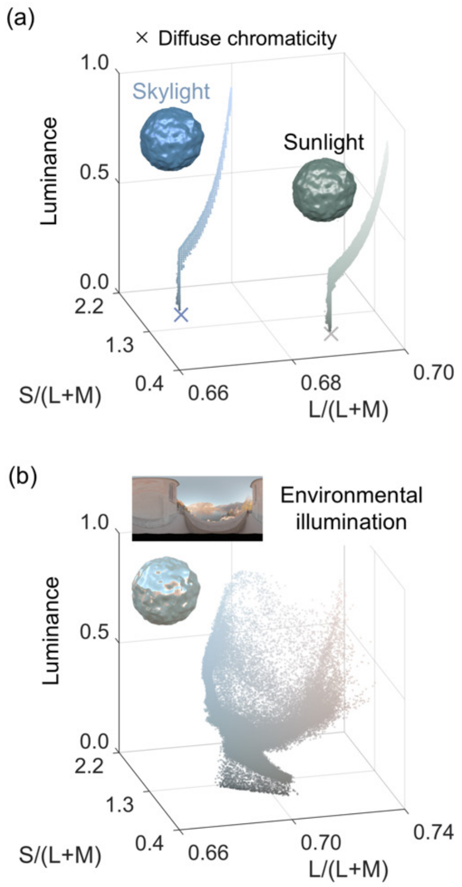

Figure 1. (a) The color distribution of pixels in the image of an object rendered either under sunlight or under skylight. (b) The color distribution of the same object under environmental illumination. Color distributions are plotted with MacLeod-Boynton chromaticity coordinates on the x- and y-axes and luminance on the z-axis. Note that the three objects have the same surface properties.

behaviorally-relevant aspect of colour constancy in terms of the role of surface colour in matching objects across different illumination contexts.

All stimuli in our experiments were computer-generated, rendered images. Lighting was specified using environmental illumination maps from an online database [8]. Such maps specify, at each pixel in unwrapped spherical coordinates, the light incident at a particular point in a scene as a function of the direction of incidence (see Figure 1b). In the main color constancy experiment, in which participants were required to identify surfaces across illuminants, four objects were presented, two of which were rendered under one environmental illumination, and two of which were rendered under a second environmental illumination. Three (distractor) objects shared the same spectral reflectance, and the fourth (target) object had a different spectral reflectance. The color constancy task in this experimental design can be considered in two parts. Firstly, to identify the environmental illumination containing the odd-one-out the observer must discriminate surface reflectance within illuminants, since the environment with a non-matching pair of surfaces must contain the target. Then, the observer must decide which surface of this pair is also presented under the other illuminant, and which surface is unique. The experimental design allows separate assessment of color constancy performance across illuminants and color discrimination performance within illuminants. Importantly this allows the experimenter to ensure that task performance is not limited by chromatic discrimination, and is instead a true test of color constancy. To determine the appropriate levels of spectral variation for this experiment, we first we measured thresholds for correctly selecting a (colored) target object with a non-uniform spectral reflectance from (grey) distractors with uniform spectral reflectance. This provided 'discrimination ellipses' for surface color in the context of complex environmental illumination. A full account of this discrimination experiment is given in Morimoto and Smithson [9], and we report here the key results in relation to the main color constancy experiment.

Methods

All stimuli were computer-graphics generated. The geometry of each scene (the three-dimensional surface of the objects, the viewpoint and an illumination map) was specified using the 3-D modelling software, Blender (Blender Foundation). Multispectral images (31 channels, from 400nm to 700nm with 10nm steps) of the rendered objects were produced with the physically-based renderer, Mitsuba [10]. We used bumpy spherical objects with one of two levels of specularity (matte or gloss). The glossy surface had a specular reflectance of 0.20 across all wavelengths, as defined within Mitsuba when set to render according to the Ward reflectance model [11] in which the image of the surface is a weighted sum of diffuse and specular components. We used two environmental illuminations from the Vogl database [8]: (1) "Distant Evening Sun (Hallstatt)" and (2) "Overcast day at Techgate Donaucity". The environmental illumination maps were originally 1024×512 images with three channels (RGB) but were promoted to multi-spectral images within Mitsuba's rendering process using a method by Smits [12]. This method uses seven spectral basis functions, whose positive weights were determined to best match the RGB values. The multispectral rendered images were converted to LMS cone coordinates based on the Stockman & Sharpe 2° cone fundamentals [13] and finally converted to RGB values for display on a calibrated CRT. We used Rendertoolbox [14] to automate the production of the multispectral images and MATLAB (MathWorks) to convert the images to RGB.

Three observers were recruited for the discrimination experiment, and seven observers for the identification experiment. The studies received ethical approval from the Medical Sciences Inter-Divisional Research Ethics Committee at the University of Oxford, UK.

Discrimination experiment

We measured discrimination thresholds for eight spectral reflectances compared to a spectrally uniform reference. During each 2-second trial, four objects were presented simultaneously. Three objects (the distractors) were assigned a uniform spectral reflectance while the fourth (the target) was assigned the test reflectance. The observer's task was to "select the object with a different surface color". The four objects were rendered under a particular environmental illuminant (either Environment 1 or Environment 2, in separate testing sessions) and presented in a dark void with no contextual information (see Figure 2a). Each

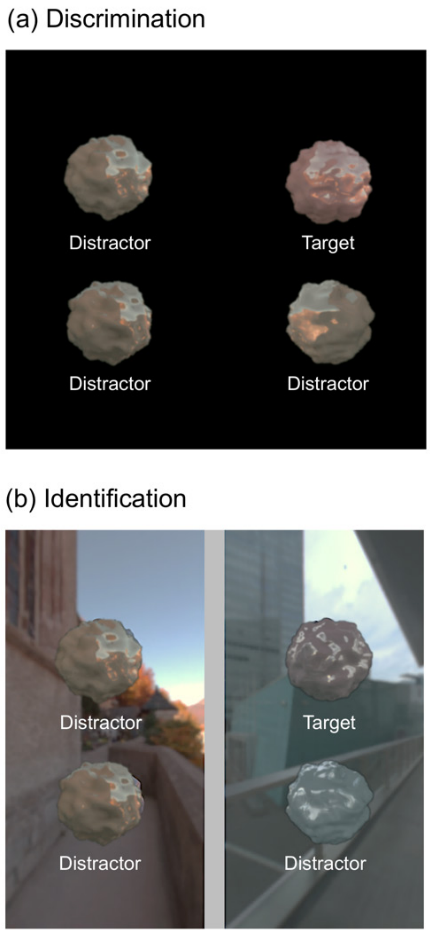

Figure 2. (a) Stimuli for the discrimination task. Four objects were presented simultaneously, and the observer's task was to find the target object (top-right) that had a different surface color from the three distractors. (b) Stimuli for the identification task. Four objects were again presented simultaneously, but the two on the left and the two on the right were placed under different lighting environments. The observer's task was the same: to find the target object (top-right) that had a different surface color from the three distractors.

object was presented from one of 20 different camera angles in order to prevent observers comparing the colors of specific points in images of the objects.

To find a representative set of reflectances to test we selected reflectances from a database of real surfaces such that the chromaticities of the eight samples, when rendered under equal energy white, were approximately evenly distributed around a hue-circle in MacLeod-Boynton chromaticity space [15]. The space was scaled to equate discrimination thresholds for uniform discs along the S/(L+M) and L/(L+M) axes. To obtain continuous control over discrimination difficulty for the reflectance samples, we used weighted mixtures of the spectrally-selective reflectance and a spectrally-uniform reflectance. Chromatic difference between each sample and the reference was therefore parameterised by this weighting, which was used as the continuous variable in the adaptive staircases used to estimate discrimination thresholds. A single testing session comprised eight interleaved staircases (maximum likelihood adaptive staircase, as implemented in Palamedes toolbox [16]), with one staircase corresponding to each spectral reflectance. Each observer completed five repeated sessions for each environmental illuminant.

Identification experiment

We measured percent correct on a reflectance identification task in which observers were required to identify the odd-one-out across illumination conditions, for pairwise combinations of nine reflectance samples. During each 3-second trial, four objects were presented simultaneously, two on the left and two on the right. The left-hand pair were rendered under one environmental illumination and the right-hand pair were rendered under the other environmental illumination. The left and right environments were separated by a vertical gray bar that had the same mean luminance as the whole screen in each trial. For each trial, two reflectances were chosen (without replacement) from the available set of nine. One reflectance was used for three of the objects (the distractors), while the other reflectance was used for the fourth object (the target). The observer's task was to "select the object with a different surface color".

Objects were either presented in a dark void, with no contextual information, or in the context of the environmental illumination under which they were rendered (see Figure 2b). The four objects were each presented in different orientations to the camera (preventing pixel-by-pixel comparisons) but, within each environment, there was a fixed geometry between the environment and the camera so that the two objects received the same illumination and, when contextual information was present, the same part of the environment map provided the context for the two objects. The visible region from a single viewpoint extended 76.4° × 41.5° in elevation and azimuth angles, respectively, and the region of the environmental illumination that was visible was different from each of 20 possible camera angles.

We chose nine reflectance functions for Environment 1 and nine for Environment 2: a spectrally uniform reflectance and eight spectrally selective reflectances that were constructed using the same method as for the discrimination experiment. Based on the discrimination results, reflectances were set to produce stimuli whose chromaticities were 2.5 times the discrimination threshold from equal energy white. Allowing for the possibility of different thresholds under the two environments, the scaling in each trial was chosen according to the environment under which the target reflectance was presented.

One block continued until the full combination of 9-choose-2 reflectances (${}_9C_2$) was tested. Assigning one reflectance to the target object and the other to distractor objects and vice versa was treated as different, as was placing the target object under the left or right environment. Consequently, one block consisted of 144 trials (${}_9C_2\times2\times2$). In each testing session, we ran four blocks, one for each of four conditions: two specular levels (glossy, and matte) and the presence or absence of surrounding context (context and no-context). Each observer completed four sessions corresponding to 2,304 trials in total.

Results

Discrimination experiment

Average discrimination ellipses for surface colors under Environment 1 and Environment 2 are shown in Figure 3. Thresholds are plotted as the chromaticity that corresponds to the spectrally-selective reflectance on which the staircase converged, rendered under an equal energy white light. Both ellipses are tilted, and the discrimination ellipse in Environment 1 is larger than in Environment 2. Both environments exhibit anisotropic variation in chromaticity, with the axis of maximum variation (red dashed lines) aligning approximately with the black body locus (blue dashed lines). The major axis of the discrimination ellipses is closely aligned to the axis of maximum chromatic variation in the illumination map. The two environments differ in their chromatic statistics, indicated in part in the pixel chromaticities displayed in the backgrounds of the results plots. In particular, the mean chromaticity of Environment 2 is displaced from equal energy white, which is used as the reference for the discrimination task. This separation of mean surface color and mean lighting color may have improved surface color discrimination. More detailed discussion of these data and of the factors influencing discrimination of surface color under complex illumination is published elsewhere [9]. The threshold data from this experiment were used to inform the selection of appropriate spectrally-selective reflectances for the main colour constancy experiment, in which observers were asked to identify surfaces across illuminants.

Identification experiment

The task in the identification experiment can be considered in two parts: first a discrimination task to find the environment that contains two different reflectances, then a colour constancy task to decide which of these reflectances matches the stimuli in the other environment. For the analysis, we included only trials where observers successfully performed the discrimination task (i.e. chose correctly the side that contained the target object). Therefore, chance performance for the colour constancy task corresponds to 50% and not 25%. Importantly, this approach ensures that performance is limited by failures of colour constancy rather than the limits of chromatic discrimination. The success rate for chromatic discrimination was 95.6%, so the excluded trials were only a small proportion of all trials, and there was no systematic difference between trials when the target appeared under Environment 1 or Environment 2.

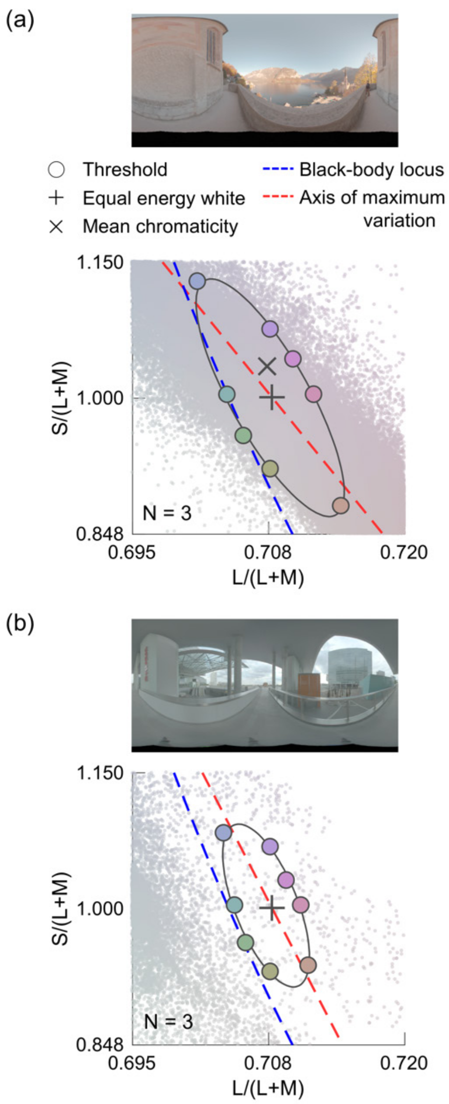

Figure 3. (a) Averaged discrimination ellipse across three observers for objects under Environment 1, "Distant Evening Sun (Hallstatt)". Stimuli are plotted in MacLeod-Boynton space, in which the x- and y-axes represent L/(L+M) and S/(L+M) respectively. The eight coloured circles are chromatic discrimination thresholds, relative to the black plus symbol that denotes the chromaticity of equal energy white from which the threshold was measured. The black cross symbol shows the mean chromaticity across all pixels in the environmental illumination. The black solid line shows the ellipse that best fits the thresholds. The small coloured dots in the background of the plot show the chromatic distribution of the pixels in the environmental illumination (not all pixels are shown within the axis limits). The red dashed line shows the axis that exhibits the maximum variation in chromaticity of the environmental illumination. The blue dashed line shows the black-body locus. (b) Averaged discrimination ellipse across three observers for objects under Environment 2, "Overcast day at Techgate Donaucity". The format of the plot is the same as in (a). Note however that the mean chromaticity of Environment 2 plots beyond the axis limits (L/(L+M) = 0.686, S/(L+M) = 1.19) so this panel has no black cross symbol.

Figure 4a shows percentage correct, averaged across reflectance pairs and across observers, for each of the four conditions (two specular levels (glossy and matte) and the presence or absence of surrounding context (context and no-context). A two-way repeated measures ANOVA revealed a significant main effect of the presence of surrounding context ($F(1,6) = 30.5$, $p = 0.00148$), while the main effect of specular reflection (glossy or matte) was not significant ($F(1,6) = 2.95$, $p = 0.137$), nor was the interaction ($F(1,6) = 4.16$, $p = 0.875$). T-tests show that the effect of context is significant both for matte and glossy objects ($t(6) = 4.981$, $p = .00249$; and $t(6) = 4.748$, $p = .00316$, respectively).

Figure 4b shows performance for individual pairs of target and distractor reflectances. Each cell of the matrix refers to a specific pair of target and distractor reflectances, according to the key presented at the top and left of the matrices in which reflectances are ordered by hue angle when rendered under equal energy white. The lower left and upper right triangle of each matrix shows trials in which the target was presented under Environment 1 or under Environment 2, respectively. Performance is highly symmetric for the two environments.

The overall improvement with contextual information is evident, but there are also systematic patterns of performance for specific reflectance pairs. When two reflectances have similar hue angle (represented by cells close to the negative diagonal), the trials are more difficult. The structure of the matrix also indicates a tendency to fail when both target and distractor fall within the blue-lilac-pink range, or within the green-yellow-orange range, but to succeed when target and distractor are drawn from different ranges, which is consistent with the smaller range of effective (threshold-scaled) hue angles for these two clusters of reflectances (shown in Figure 3). Finally, the bottommost row and rightmost column of the matrix show data for trials in which the target or distractor has uniform reflectance and is compared to one of the spectrally-selective samples. If performance in these trials followed the within-illuminant discrimination performance, each combination would elicit the same performance. Variation in performance between these combinations indicates a further interaction with the chromatic properties of the environments that are being compared.

Discussion

The chromatic statistics associated with an image of an object are markedly different when the object is viewed under a single illuminant or under environmental illumination (Figure 1). The present experiment quantifies performance-based color constancy across different environmental illuminations. Specifically, we tested the ability to identify an object with a target spectral reflectance amongst distractors when objects were presented under different environments. Under conditions where performance was not limited by chromatic discrimination within illuminants, we found mean performance on the color constancy task ranged between 70 and 80%. Performance was significantly poorer when object images were presented in a dark void compared to when they were presented in the context of the environmental illumination under which they were rendered. The specularity (glossiness) of the surface, despite changing the chromatic statistics of the image of the object and potentially providing direct information about the illumination [17], had no effect on performance.

Observers do not have direct access to surface color independent of illumination. Robust comparison of surface color across illuminants therefore requires compensation for the lighting context, which we generically term a 'color conversion', and we discuss possible color conversions below.

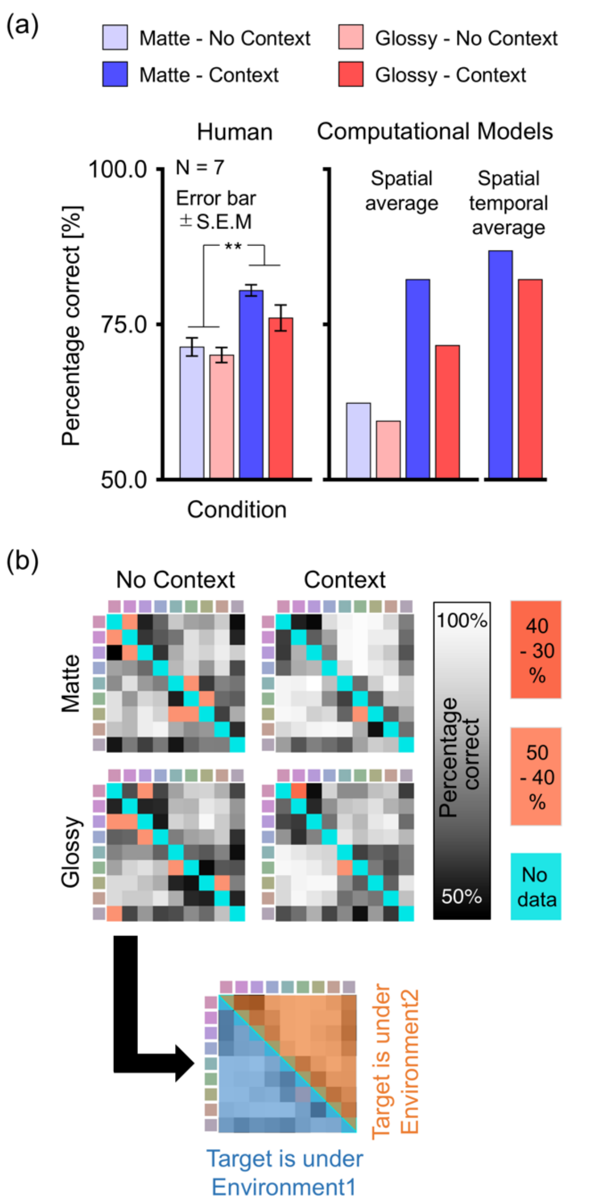

Figure 4. (a) Left panel: Averaged percentage correct values across seven human observers. Error bars indicate ±S.E.M across observers. Asterisks indicate a significant difference between "context" and "no context" for both matte and shiny objects ($p < 0.05$). Center panel: Percentage correct values for a computational model that performs a white-point correction based on the luminance-weighted mean chromaticity of the surrounding context. Note that for the no context condition, no correction was performed. Right panel: Performance of a model that corrects the influence of illumination based on a spatial-temporal luminance-weighted mean chromaticity over trials. (b) Performance for individual pairs of target and distractor reflectances, averaged across observers. Each cell in the matrix represents a particular pair of target and distractor reflectances (color-coded along the upper and left edges of the matrix), and the shading indicates the average percentage correct as defined by the colour bar. The orange cells show percentage correct values in the range 40% to 50% and from 30% to 40%. Since target and distractor were never assigned the same reflectance, cyan cells on the diagonal indicate that there is no data. As shown in the subsidiary figure (shown by the black arrow), the lower left and upper right triangle of each matrix shows trials in which the target object was presented under Environment 1 or under Environment 2, respectively.

The results suggest that it is unlikely that observers are using specular highlights to provide direct samples of the illumination and drive the color conversion. This inference follows specifically from the lack of performance improvement for glossy objects and the disadvantage for the no-context condition, even though the glossy objects contain the same specular highlights in the no-context condition as they did in the context condition.

The simplest model for selecting the target from distractors would be to base the judgement directly on the average chromaticity of the object, with no correction for the illumination. This is predicted to lead to poor constancy, and indeed an observer implementing this process (using the luminance-weighted mean chromaticity of each object) would achieve only 62.3 and 59.4 percent correct for matte and glossy objects respectively (relative to chance performance of 50 percent correct). Real observers not only do better than this, but they also show improved performance when objects are presented with surrounding context. A model that considers the spatial average of the surrounding context and uses this for white point correction of average object color [18] captures some general features of the real observer performance, achieving 82.2 and 71.6 percent correct for matte and glossy objects respectively.

Data from the present experiment additionally encode the observer's dependence on trial-to-trial variation in viewing conditions, introduced primarily by the selection of different camera angles for each environment in each trial. A model that bases white-point correction on an average over past trials achieves 86.8 and 82.2 percent correct for matte and glossy objects respectively. This dependence on the past history of trials replicates earlier findings with two-dimensional matte stimuli under uniform illumination that indicate that temporal context is important in achieving color constancy [19] and that illuminant correction is not instantaneous [20]. However, modelling average performance is not a sensitive test of underlying mechanisms, and trial-by-trial analyses indicate that none of the white-point correction algorithms discussed here provide a very good account of observers' choices between target and distractor reflectances. Indeed, using d-prime to summarize the models' successes and failures in predicting human choices shows that the null model, with no correction for the illuminant, achieves a d-prime of about 0.5 and that none of the white-point correction algorithms do better than a d-prime of about 1.0.

Surface color perception under environmental illumination is a primary example of a visual task that requires extraction of stable signals from variable inputs. The summary models discussed here serve to highlight consistent features of performance, and statistical regularities in chromatic statistics that could contribute to the overall pattern of effects. But, as the d-prime analyses indicate, none of the existing models provide a good account of the trial-to-trial variability in observer performance. It is possible that this variability is the key to further understanding the mechanisms underlying observers' selections of the target object, and surface color perception more generally.

The perception of surface color under environmental illumination is a rather different computational task compared to surface colour perception under uniform illumination. Numerous studies have suggested that observers are indeed sensitive to the three-dimensional structure of the scene and lighting. For example, surface color judgements are sensitive to the way in which mutual illumination depends on the relative orientations of two surfaces [21], and to the way in which directional anisotropies in light-fields change the effective illumination of a tillable surface [22]. Moreover, phase-scrambling manipulations that disrupt the interpretation of the three-dimensional structure of scene, whilst maintaining the chromatic statistics of the proximal image, (e.g. [23]; [24]) show sensitivity to scene geometry. A complete description of surface color perception will need to explain patterns of behavior that depend on these environmental features, and not only on the statistics of the instantaneous proximal image.

References

[1] T. Morimoto, S. Kishigami, J. M. M. Linhares, S. M. C Nascimento, and H. E. Smithson, “Hyperspectral environmental illumination maps: characterizing directional spectral variation in natural environments”, Opt. Exp., 27, 22 (2019).

[2] H. E. Smithson, “Sensory, computational and cognitive components of human colour constancy”, Philos. Trans. R. Soc. Lond. B. Biol. Sci., 360, 1458 (2005).

[3] D. H. Foster, “Color constancy”, Vis. Res., 51, 7 (2011).

[4] A. Hurlbert, “Colour constancy”, Cur. Biol., 17, 21 (2007).

[5] L. T. Maloney, “Physics-based approaches to modeling surface color perception” in Eds. K. R. Gegenfurtner and L. T. Sharpe “Color vision: from genes to perception” (Cambridge University Press, 1999) pg. 387–422.

[6] R. Robilotto, and Q. Zaidi, “Limits of lightness identification for real objects under natural viewing conditions”. J. Vis., 4, 9 (2004).

[7] B. Xiao and D. H. Brainard, “Surface gloss and color perception of 3D objects”, Vis. Neurosci. 25, 3 (2008).

[8] B. Vogl, “Light probes”, http://dativ.at/lightprobes/.

[9] T. Morimoto, and H. E. Smithson, “Discrimination of spectral reflectance under environmental illumination”, J. Opt. Soc. Am. A,, 35, 4 (2018).

[10] W. Jakob, “Mitsuba renderer”, https://www.mitsuba-renderer.org.

[11] G. J. Ward, “Measuring and modeling anisotropic reflection,” ACM SIGGRAPH Comput. Graph. pg. 265–272 (1992).

[12] B. Smits, “An RGB to spectrum conversion for reflectances”, J. Graph. Tools 4, 4, (2000).

[13] A. Stockman and L. T. Sharpe, “The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype,” Vis. Res. 40, 13 (2000).

[14] B. S. Heasly, N. P. Cottaris, D. P. Lichtman, B. Xiao, and D. H. Brainard, “RenderToolbox3: MATLAB tools that facilitate physically based stimulus rendering for vision research,” J. Vis., 14, 2 (2014).

[15] D. I. A. MacLeod and R. M. Boynton, “Chromaticity diagram showing cone excitation by stimuli of equal luminance,” J. Opt. Soc. Am., 69, 8 (1979).

[16] F. A. A. Kingdom and N. Prins, “Palamedes: MATLAB routines for analyzing psychophysical data,” http://www.palamedestoolbox.org.

[17] J. N. Yang, and L. T. Maloney, “Illuminant cues in surface color perception: tests of three candidate cues”, Vis. Res., 41, 20 (2001).

[18] G. D. Finlayson, S. D. Hordley, and P. M. Hubel, “Color by correlation: A simple, unifying framework for color constancy”, IEEE Transaction on Pattern Analysis and Machine Intelligence, 23, 11 (2001).

[19] H. E. Smithson, and Q. Zaidi, “Colour constancy in context: Roles for local adaptation and levels of reference”, J. Vis., 4, 3 (2004).

[20] R. J. Lee, K. D. Dawson and H. E. Smithson, “Slow updating of the achromatic point after a change in illumination”, J. Vis., 12, 19 (2012).

[21] M. G. Bloj, D. Kersten, and A. Hurlbert, “Perception of three-dimensional shape influences colour perception through mutual illumination”, Nature, 402 (1999).

[22] H. Boyaci, K. Doerschner, and L. T. Maloney, “Perceived surface color in binocularly viewed scenes with two light sources differing in chromaticity”, J. Vis., 4, 1 (2004).

[23] R. J. Lee, and H. E. Smithson, “Low levels of specularity support operational color constancy, particularly when surface and illumination geometry can be inferred”, J. Opt. Soc. Am. A, 33, 3 (2016).

[24] T. Morimoto, R. J. Lee, and H. E. Smithson, “When the brightest is not the best: illuminant estimation based on highlight geometry”, J. Vis., 17, 10 (2017).