Do surface reflectance properties and 3-D mesostructure influence the perception of lightness?

Alexandra C. Schmid
School of Psychology, University of Sydney, NSW, Australia

Barton L. Anderson
School of Psychology, University of Sydney, NSW, Australia

A series of experiments were conducted to assess how the reflectance properties and the complexity of surface “mesostructure” (small-scale 3-D relief) influence perceived lightness. Experiment 1 evaluated the role of surface relief and gloss on perceived lightness. For surfaces with visible mesostructure, lightness constancy was better for targets embedded in glossy than matte surfaces. The results for surfaces that lacked surface relief were qualitatively different than the 3-D surrounds, exhibiting abrupt steps in perceived lightness at points at which the targets transition from being increments to decrements. Experiments 2 and 4 compared the matte and glossy 3-D surrounds to two control displays, which matched either pixel histograms or a phase-scrambled power spectrum, respectively. Although some improved lightness constancy was observed for the 3-D gloss display over the histogram-matched display, this benefit was not observed for phase-scrambled variants of these images with equated power spectrums. These results suggest that the improved lightness constancy observed with 3-D surfaces can be well explained by the distribution of contrast across space and scale, independently of explicit information about surface shading or specularity whereas a surface’s mesostructure interacts with the light field to create shading, shadows, and illuminance flow.

A number of researchers have hypothesized that the visual system uses shading, shadows, and specular reflections to somehow discount, compensate for, or estimate the intensity of the illumination, thereby constraining estimates of surface reflectance (e.g., Boyaci, Doerschner, & Maloney, 2006; Kraft, Maloney, & Brainard, 2002; Maloney, 2002; Snyder, Doerschner, & Maloney, 2005; Yang & Maloney, 2001). Gilchrist and Jacobsen (1984) found that observers could reliably discriminate two 3-D scenes that were painted either uniformly (matte) white or black; the white room looked white, and the black room looked midgray, independent of its brightness. They noted that the contrast of the shadows in the black room was stronger than in the white room, which arose from the difference in the strength of interreflections for the two albedos. Their white surfaces reflected ~90% of the incident illumination whereas their black surfaces reflected ~3%. There-

Introduction

The image our eyes receive is perceptually parsed into different causal sources, such as object shape, illumination, and surface reflectance properties (lightness, color, gloss, and transparency). An extensively studied problem in vision science is how the visual system recovers the albedo or lightness of surfaces, which is defined as the proportion of light a surface reflects diffusely. The computation of lightness is typically regarded as an underconstrained problem because any particular luminance could be generated by an infinite combination of surface albedos, illuminations, and surface pose. The majority of lightness studies have used simple scenes with only a few luminance values arising from smooth, matte, flat surfaces. In such scenes, disentangling the contributions of lightness and illumination is an ill-posed problem. However, most natural surfaces are made of materials that are not characterized by purely diffuse reflectance and often contain medium-scale surface relief (mesostructure). A surface’s microstructure typically generates both diffuse and specular reflections whereas a surface’s mesostructure interacts with the light field to create shading, shadows, and illuminance flow.


doi: 10.1167/14.8.24 Received April 7, 2014; published July 29, 2014
ISSN 1534-7362 © 2014 ARVO

Downloaded From: https://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/933549/ on 10/17/2018
fore, much more light indirectly illuminated surfaces in the white scene compared to the black scene, which provides additional information about surface albedo.

In a more recent study (Sharan, Li, Motoyoshi, Nishida, & Adelson, 2008), observers viewed photographs of matte and glossy uniformly painted surfaces containing complex mesostructure that were equated for mean luminance. When these surfaces were viewed in isolation, observers’ lightness judgments were positively correlated with true surface reflectance although these data exhibited a regression to the mean (white surfaces appeared darker than they were whereas black surfaces appeared lighter). In addition to the 3-D mesostructure present in all of their stimuli, some of their stimuli also contained specular reflections. A number of authors (Boyaci et al., 2006; Kraft et al., 2002; Maloney, 2002; Snyder et al., 2005; Yang & Maloney, 2001) have proposed that the visual system uses cues, such as specular highlights, shadows, and shading, to generate an estimate of the scene illuminant, which they term an equivalent illumination model (EIM). An EIM is a representation of the spatial and spectral distribution of light in a scene (also called the visual light field) and may contain estimates of the direction, intensity, chromaticity, and diffuseness of the light source. The EIM is then used to constrain estimates of the color or lightness of surfaces in the scene that lack explicit illumination cues. For example, Boyaci et al. (2006) had observers judge the lightness of a flat matte surface that varied in orientation in a stereoscopically viewed virtual scene, which either contained or lacked cast shadows, shading, and specular highlights. When each potential cue was presented in isolation, observers’ lightness judgments appeared to take 3-D surface pose into account. Furthermore, observers’ lightness estimates were more reliable when several cues to the illumination were present than when each cue was presented in isolation. Boyaci et al. (2006) concluded that the visual system uses these cues to derive information about the illumination to constrain estimates of surface albedo. Similar arguments were made by Snyder et al. (2005), who assessed observers’ ability to compensate for illumination gradients in binocularly viewed virtual scenes. They found that observers judged lightness more veridically when the scene contained floating, glossy spheres than for similar scenes that contained no specular cues.

One limitation of previous lightness studies is that there is currently no systematic attempt to assess the relative importance of different illumination cues directly. Boyaci et al. (2006) found that including all cues to the illumination in a scene was better than having one cue present, but they did not test all cue combinations to determine their relative contributions. It is also unclear whether the preceding results arose from an estimation of the illumination field, as suggested by EIMs, or whether they were the result of low-level differences in image content. The addition of specular highlights and/or the presence of shadows and shading would have changed the range and distribution of luminance values in the image, which may have influenced lightness judgments. Previous studies have not attempted to tease apart low-level explanations involving luminance and contrast distributions from midlevel explanations involving representations of the light field.

The following experiments were designed to tease apart low-level contextual influences on perceived lightness from midlevel explanations that invoke the estimate of the illuminant. The stimuli in the following experiments are center-surround displays with flat, matte central test patches and various surround types. In Experiment 1, we test the hypothesis that lightness constancy of the test patch will improve when the surround contains complex mesostructure (high surface relief) and specular highlights (gloss) compared to when the surround lacks this information. In a series of control experiments, we test whether any improvements in lightness constancy arise from information about the light field or whether they can be attributed to low-level attributes, such as the distribution of luminances and contrasts across different spatial scales. To anticipate, we find that complex mesostructure and specular highlights elicit more consistent lightness judgments than images lacking this information. However, control experiments reveal that this contextual effect most likely arises from the contrast and luminance distributions across space and scale rather than an explicit estimate of the illuminant.

**Experiment 1a and 1b: Varying surface relief and gloss level of the surround**

In Experiment 1, observers performed lightness judgments on test patches embedded in four surround types: low-relief matte, low-relief glossy, high-relief matte, and high-relief glossy (see Figure 1). Rendering the stimuli produced almost identical images for the low-relief matte and glossy conditions. For completeness, we included both conditions in Experiment 1a. However, the low-relief glossy condition was removed in Experiment 1b after verifying that the stimuli from both low-relief conditions produced indistinguishable images and essentially identical results.
Methods

Observers

Experiment 1 included two populations of observers. Five observers participated in Experiment 1a. Observers AS (an author), PM, and KT were experienced in psychophysical experiments. Observers DC and RS were inexperienced and paid $20 per hour for participation. These observers performed five repeats of each condition (see procedures).

Twenty undergraduate first-year psychology students at the University of Sydney participated in
Experiment 1b. These observers were awarded course credit in exchange for participation and performed one repeat of each condition. All observers except AS from Experiment 1a were naïve to the aims of the study.

Apparatus and stimuli

Stimuli were presented on a LaCie Electron 22 Blue IV monitor running at a refresh rate of 75 Hz and with a resolution of 1280 × 1024 pixels, controlled by a Mac Pro computer running Mac OS X 10. Stimulus presentation and data collection were controlled by a Matlab (R2010a; Mathworks) script using the Psychophysics Toolbox (Brainard, 1997). Stimuli were viewed in a dark room at a viewing distance of approximately 70 cm. The carpet and walls of the room were black so that the only source of light came from the monitor on which the stimuli were displayed.

Stimuli were computer-rendered center-surround displays (Figure 1). The displays contained flat, matte centers that varied in albedo and surrounds that varied in albedo, amount of surface relief, and gloss level. Surrounds had either low surface relief (flat; Figure 1A and 1B) or high surface relief (rocky; Figure 1C and 1D) and were either matte (Figure 1A and 1C) or glossy (Figure 1B and 1D). The surfaces were modeled in the open-source software Blender (v. 2.6). The rockiness in the surround was generated with the displace modifier, a tool in Blender that displaces vertices in a mesh based on the intensity of a texture. To obtain different effects of rockiness, various textures were used to deform the surface (the inbuilt cloud, marble, and stucci textures as well as textures from images of rocks and rough paper; see Appendix A for details).

Stimuli were rendered using the RADIANCE rendering software (Ward, 1994), which simulates physical interactions between illuminants and surfaces. The surfaces were rendered using the Ward BRDF model, termed “plastic” in RADIANCE. This model has five parameters: diffuse components R, G, and B; specularity ($P_s$, the proportion of light reflected by the specular component, uncolored); and microroughness ($z$, which determines the amount of specular scatter). Centers were embedded in the surrounds and rendered as part of the same scene. Gray shades were assigned to the center and surround regions by adjusting the diffuse reflectance parameters, keeping relative RGB values equal. Matte surrounds were assigned a specularity value of 0 and a roughness value of 0 whereas glossy surrounds were assigned a specularity value of 0.05 (5% of the light reflected from the surface is specular) and a roughness value of 0.01. Surfaces were rendered frontoparallel to the observer with two ambient reflections.

All surfaces were illuminated by the “grove” light probe from the Debevec Light Probe Image Gallery (Debevec, 1998). A gray scale version of this light probe was created so that all surfaces were illuminated by achromatic light. To produce high-quality images, all surfaces were rendered 10 times larger than required and antialiased, resulting in high dynamic range (HDR) images with dimensions of 900 × 900 pixels. The HDR images were tone-mapped to fit the luminance range of the monitor. This was achieved by linearly compressing the diffuse component and nonlinearly compressing the specular component of the images (see Appendix B).

Procedure

Observers judged the lightness of flat target patches embedded in various surrounds via an asymmetric matching task (see Figure 1 for stimuli). In each trial, a target surface (14.88”) was presented on the computer screen. Below the target surface was a smaller surface with an adjustable test patch (5.83”). The surfaces were separated by 4.96” and were presented against a black background. Observers were instructed to change the lightness of the flat center patch on the adjustable surface until it looked like it was the same lightness or painted with the same paint as the flat center patch on the target surface.

In Experiment 1a, target surfaces consisted of a flat, matte, central target patch surrounded by one of four surface types: flat and matte (Figure 1A), flat and glossy (Figure 1B), rocky and matte (Figure 1C), or rocky and glossy (Figure 1D). For each of these conditions, there were six different surround albedos ranging from black to white (see Table A2). This produced 24 surround conditions in total. For these 24 surround conditions, observers judged 13 to 15 test patch albedos. Table A2 shows specific test patch albedos included for each of the surround conditions, and Figure 2 displays examples of these test patch albedos for surround reflectance 19.8%. Eleven of the test patch albedos were standard for all surround conditions. We included a further two to four unique test patch albedos very close in lightness to each surround albedo. Two of these extra values were increments and two were decrements except for black surrounds, which had only two extra increments, and white surrounds, which had only two extra decrements. There were 344 conditions in total. Observers performed five repeats of each condition, resulting in 1,720 trials.

The flat, glossy condition in Experiment 1a was included for completeness. However, due to the specific viewing angle of the camera in relation to the surface and the light source, rendering matte and glossy flat surfaces generated indistinguishable images, thus producing almost identical results. For this reason, the flat glossy condition was removed in Experiment 1b. Additionally, observers performed only one repeat of
Each condition, resulting in 258 trials. All other aspects of Experiments 1a and 1b were the same.

Figure 1E shows the adjustable surface that was used for all conditions in Experiments 1, 2, and 4. The surround was a checkerboard surface with equal amounts of black (3% reflectance) and white (90% reflectance) bordering the center patch. The surface relief and gloss level of the surround were identical to that of the glossy, high-relief test surfaces. Observers were able to adjust the albedo of the center patch by moving a computer mouse left and right. They could choose from 201 prerendered Munsell values ranging from zero to 10 in equal increments on the Munsell scale.

Results and discussion

The results from Experiment 1 are presented in Figure 3, which shows the average data of all five observers from Experiment 1a (left column) and the average data of the 20 observers from Experiment 1b (right column). Three trials (out of 5,160) from Experiment 1b were excluded from analyses due to observers (accidentally) pressing the button to set test patch lightness before they had finished making adjustments. The data reveal that test patch lightness settings are affected by surround type. The different surface relief and gloss level conditions give rise to different patterns in the data. These patterns will be discussed in relation to lightness constancy below.

Low-relief results

One of the most salient differences between conditions is the difference between the low-relief and high-relief data curves in Figure 3 (compare the top two rows to the bottom two rows, respectively). The high-relief data curves are relatively linear whereas the low-relief data curves exhibit a large “step” at which test patch reflectance passes through that of the surround. This sharp lightness change between low-contrast increments and decrements reflects a phenomenon termed crispening by Takasaki (1966), and we retain this terminology here. To emphasize the size of the step for each of the data curves in Figure 3, difference scores were obtained by subtracting the lowest contrast decrement settings from the lowest contrast increment settings. These difference scores are plotted in Figure 4. A lower score indicates a smaller step and therefore less crispening. The darkest (Munsell value 1.95) and lightest (Munsell value 9.5) surround conditions were omitted because they contained only increments or only decrements, respectively.

When the surround is essentially homogeneous (low-relief condition), increments appear much lighter than decrements: For observers in Experiment 1a, the lowest contrast increments appeared, on average, about 2 Munsell values lighter than the lowest contrast decrements (see Figure 4A); for observers in Experiment 1b, the difference was about 1 Munsell value (see Figure 4B). The actual (simulated) increment-decrement Munsell difference was 0.3, indicated by the horizontal dotted lines in Figure 4. Difference scores from the high-relief conditions lie around this line. From these difference scores, it clear that, for both sets of observers, crispening is only induced by the low-relief surrounds. These observations were statistically verified for Experiment 1a by subjecting the difference scores for each surround Munsell condition to a within-subjects two-way ANOVA with two levels of surface-relief (low, high) and two levels of gloss (matte, glossy). F values and p values are displayed in Table A3. There was a main effect of surface relief for all surround Munsell conditions. This confirms the observation that the increment-decrement step is larger when test patches are surrounded by low-relief compared to high-relief surrounds. There is one inconsistency for surround Munsell 8, at which the step seems to be smaller for the low-relief glossy surround (Figure 4A).
Indeed, for this condition, there was a main effect of gloss and an interaction between surface relief and gloss level. Follow-up tests suggested that the low-relief glossy surround was not as effective in inducing crispening as the low-relief matte surround. However, Figure 3 (left column, second panel from the top, orange data points) clearly shows strong crispening for this condition. Closer inspection of the stimuli revealed that, for surround Munsell 8, the lowest contrast decrement was practically indistinguishable from the low-relief glossy surround. Rendering the low-relief surrounds with gloss made the surfaces appear darker than their matte counterparts (because part of the light was reflected in the specular component that was not visible from the camera’s point of view). Darkening the surround slightly caused the lowest contrast decrement to appear close enough to the surround color that it was almost undetectable. Observers were therefore...

Figure 3. Average data for Experiment 1a (left panels) and 1b (right panels). LR stands for low relief; HR stands for high relief. Each colored data curve represents test patch settings for a different surround albedo condition. The legend shows the Munsell values of each surround. For low-relief conditions (top three panels), there is an increment–decrement “step” (crispening) as the test patch reflectance passes through that of the surround. This step is absent in the high-relief conditions (bottom four panels). Comparing the high-relief data, lightness settings are more consistent for the glossy condition (last row) compared to the matte condition (third row).
likely to match this test patch to the surround color, reducing the lowest contrast increment-decrement step.

For Experiment 1b, t tests using Sidak-corrected alpha values of 0.253 per test\(^4\) were carried out to compare difference scores in the low- and high-relief conditions (see Table A4 for \( t \) values, \( df \), and \( p \) values). For three out of the four surround Munsell conditions, low-relief surrounds induced a larger increment-decrement step than both the high-relief matte and glossy surrounds (Figure 4B). The apparent lack of step for surround Munsell 8 can be explained by the probabilistic nature of detecting very low-contrast test patches (Ekroll & Faul, 2012b). It appears that observers in Experiment 1b were, on average, less likely to detect the low-contrast decrement compared to observers in Experiment 1a.

The above findings raise the question of why there are differences in observers’ ability to detect low-contrast test patches between Experiment 1a and 1b. Additionally, there is a discrepancy in the amount of crispening observed (the size of the step) between Experiment 1a and 1b. Previous research suggests that these inconsistencies can be attributed to individual differences in how the stimuli are perceived. For example, Ekroll and Faul (2009) found large individual differences in the size of the crispening effect for colored center-surround displays.

Another quality about the low-relief data curves is the asymmetry between increment and decrement settings. The averaged data in Figure 3 (top two rows) reveals that increment settings are essentially independent of the surround reflectance. In contradistinction, decrement settings are more affected by the surround reflectance, illustrated by the greater spread in the data points for each test patch. Asymmetries between increment and decrement settings have been found previously in the brightness literature (e.g., Heinemann, 1955). Additionally, for test patches on colored surrounds, color induction from the surround has been found to be much stronger for decrements than increments (Bäuml, 2001; Helson, 1938; Helsen & Michels, 1948).

The most notable aspect of the data is that crispening is completely eliminated when surrounds contain shading and shadow information indicative of surface-relief (bottom two rows in Figure 3). This implies a qualitative difference in how the test patch is perceived when embedded in flat compared to rocky surrounds. In the General discussion, we address a growing view in the literature that lightness perception may contain more than one dimension (Ekroll & Faul, 2013; Logvinenko & Maloney, 2006; Vladusic, 2012, 2013; Vladusic, Lucassen & Cornelissen, 2007) and consider the role of perceptual phenomena, such as transparency, influencing the appearance of test patches embedded in uniform surrounds (Ekroll & Faul, 2013).

High-relief results

The above results suggest a qualitative difference in how the test patch is perceived when embedded in rocky surrounds that contain shading information compared to flat surrounds that do not contain this information. For both matte and glossy high-relief conditions, there is a tendency for test patches on darker surrounds to appear lighter than those same test patches on lighter surrounds. However, settings from the glossy condition seem more compressed (more consistent) than those from the matte condition (compare the bottom and second-bottom rows in Figure 3, respectively). For these high-relief conditions, lightness constancy for a given test patch tends to be better when surrounds are glossy compared to matte.

The difference in lightness constancy (vertical spread) between high-relief matte and glossy data...
points was statistically reliable. For each of the 11 test patch values common to all surrounds (see Table A2), the lightness settings in one surround condition were subtracted from the lightness settings in the adjacent darker surround Munsell condition. These difference scores were averaged for all 11 test patch values and plotted in Figure 5. For this and subsequent experiments, a binomial sign test was used to compute the likelihood of obtaining $k$ or more instances in which the observers’ performance in the glossy condition was more consistent than in the matte condition (11 pairs of data points per subject). The results confirmed that, for high-relief conditions, lightness constancy is significantly better when test patches are surrounded by glossy compared to matte surfaces; $p < 0.001$ for Experiment 1a and $p = 0.006$ for Experiment 1b. This implies that surfaces with gloss information provide the visual system with additional cues that can be used to improve lightness constancy.

The amount of lightness constancy exhibited in the data also depends on the test-patch reflectance. However, this effect of test-patch reflectance on lightness constancy differs between observers in Experiment 1a and 1b. Despite these differences between observers, Figure 3 (third row from the top) and Figure 5 demonstrate a general trend, at least for the high-relief matte condition: Lightness constancy is better for lighter compared to darker test patches. This is illustrated by the negatively sloped matte data curves in Figure 5. This is likely to be caused by lighter test patches substantially increasing the range of luminance values in most of the displays.

One last point to note is observers’ settings in relation to ground truth (solid black diagonal line in Figure 3). For the high-relief matte data points, there is a tendency for test patches on lighter surrounds to be more veridical to those on darker surrounds. One reason for this shift might be the high range of luminance values in the adjustable patch’s surround. Matte surrounds with higher reflectance also have a greater range of luminance values (due to light diffuse shading and dark shadows) compared to matte surrounds with lower reflectance (which have dark diffuse shading and dark shadows). This makes the matches to our adjustable patch (which was surrounded by both black and white surfaces) more symmetric when displays are lighter, leading to more veridical matches.

In Experiment 1a and 1b we found that the visual system uses image cues generated by rocky and glossy surfaces when estimating test patch lightness. The presence of these cues leads to better lightness constancy compared to when these cues are absent: Crispening is eliminated when surrounds contain complex mesostructure; for high-relief surfaces, lightness judgments are more consistent when surrounds are glossy compared to matte. Note that the rocky matte data points exhibit similar overall spread to the low-relief displays (i.e., a given test patch embedded in the lightest and darkest surround appears similarly different for the two conditions). In this sense lightness constancy appears to be similar for these two conditions. However, when focusing on data points near the surround value, lightness constancy is revealed to be better in the rocky displays.

One possible explanation of the above results is that the specific luminance patterns generated by naturalistic surfaces may serve as cues to help the visual system decompose luminance values into contributions of lightness and illumination. Advocates of EIMs propose...
that these cues provide the visual system with information about the light field, which is then transferred to other parts of the scene that lack this information, e.g., the flat test patch. However, an alternative explanation is that the improved lightness constancy for glossy conditions could be caused by the greater range and/or variation of luminance values in the image compared to matte conditions (i.e., greater articulation, e.g., see Gilchrist et al., 1999). Experiment 2 was designed to directly test this possibility. Variegated center-surround stimuli were created in a way that eliminated surface structure but retained the range and variation of luminance values in the image.

Experiment 2: Two-dimensional variegated surrounds

Experiment 2 tests whether surround “articulation” (luminance range and/or variation) was responsible for the improved lightness constancy exhibited with the 3-D rocky surrounds in Experiment 1. In Experiment 2, equivalent 2-D flat surfaces with variegated surrounds were created for each rocky surface (Figure 6). These 2-D variegated surfaces had very similar luminance histograms to the 3-D rocky surfaces but lacked information about surface structure, shading, specularity, or illuminance flow. If the specific luminance patterns generated by naturalistic surfaces are important in lightness constancy, then lightness constancy should be worse for the 2-D variegated surfaces than for comparable 3-D surfaces. If surface structure is not important, and it is only the amount of articulation that is crucial, then no difference in lightness constancy between 3-D rocky and 2-D variegated conditions should be observed.

Methods

Observers

Similar to Experiment 1, Experiment 2 had two parts: Observers DC and RS from Experiment 1a participated in Experiment 2a, and 20 first-year psychology students participated in Experiment 2b (none of whom had participated in the first experiment).

Apparatus and stimuli

The task was the same as in Experiment 1 as was stimulus presentation. Two-dimensional variegated surrounds were created in the following way for each gloss level and each surround Munsell condition (see Figure 6). First, the relative frequencies of each luminance value in the 3-D rocky surround were obtained. There were 256 possible luminance values corresponding to each eight-bit pixel value of the monitor. These relative values were used to create a variegated surround consisting of 1,280 squares (36 × 36 minus 16 squares for the central patch). The number of squares of each luminance was weighted according to the relative luminance frequencies from the 3-D rocky surround. This resulted in very similar luminance frequency histograms for corresponding 2-D variegated and 3-D rocky surrounds. In this sense, the corresponding 2-D variegated and 3-D rocky surrounds are equivalent, but variations in shading and gloss in the rendered stimuli appear as part of a random noise.
pattern in the 2-D images. Because the 3-D rocky surround was made up of 799,515 pixels and the 2-D variegated surround contained only 1,290 squares, some luminance values in the 3-D surround did not have enough pixels to make up one square in the 2-D variegated surround. If this occurred, we ensured that the variegated surround contained at least one square with the highest luminance value and at least one square with the lowest luminance value in the 3-D surround. We also ensured the regions immediately surrounding the central patch in the 2-D variegated and 3-D rocky stimuli were matched in average luminance.

There were very slight luminance variations within the test patches embedded in rocky surrounds caused by interreflections. Test patch values for the 2-D variegated stimuli were created by averaging these luminance values.

**Procedure**

The layout of trials and instructions were the same as in Experiment 1. There were two surround types:

- Variegated surrounds that were equivalent to the matte surrounds in Experiment 1 (“matte equivalent”)
- Variegated surrounds that were equivalent to the glossy surrounds in Experiment 1 (“glossy equivalent”). The surround Munsell conditions and test patch conditions were the same as in Experiment 1, leading to 172 trials in total.

**Results and discussion**

The average results of Experiment 2 are presented in Figure 7. Two trials from Experiment 2b were excluded from analyses due to observers making premature lightness judgments (see Experiment 1 results). The data reveal that test patch lightness settings are affected by surround type. To more clearly compare lightness settings, Figure 8 (left panels) plots the average difference scores between adjacent Munsell conditions (the same method was used as in Experiment 1). A smaller difference score indicates more consistent settings and therefore better lightness constancy. We
also plotted the standard deviation of test patch settings for different surround Munsell conditions (Figure 8, right panels) to demonstrate that the effects reported are not a result of the specific method chosen to quantify lightness constancy.

Binomial sign tests were carried out on the average difference scores and standard deviation scores.5 The results of Experiment 2a suggest that when surrounds are variegated, observers exhibit better lightness constancy when there is greater articulation (luminance variation/range) in the surround (glossy equivalent condition) compared to when there is less articulation (matte equivalent condition), \( p < 0.001 \) (Figure 8, top panels; compare blue filled circles to blue open squares). However, this result was not replicated in Experiment 2b, \( p = 0.5 \), indicating that greater surround articulation may not influence lightness constancy for all observers (Figure 8, bottom panels). The results from both sets of observers confirm that lightness constancy is better when test patches are embedded in 3-D rocky compared to 2-D variegated surrounds, at least for the gloss conditions, \( p < 0.001 \) (Experiment 2a, Figure 8, top panels), \( p = 0.006 \) (Experiment 2b, Figure 8, bottom panels; compare open blue squares to open red squares). This suggests that 3-D surfaces containing shading and specular reflections provide additional cues over and above luminance range and variation that the visual system uses to improve lightness constancy. The results were not as clear for the matte condition. Observers from Experiment 2a exhibited better lightness constancy when test patches were embedded in 3-D rocky compared to 2-D variegated surrounds, \( p < 0.001 \) (Figure 8, top panel). However, the same was not true for observers in Experiment 2b, \( p = 0.50 \) (Figure 8, bottom panel; compare filled blue circles to filled red circles).
The results of Experiment 2 suggest that the amount of surround articulation in the high-relief conditions cannot account for the results in Experiment 1. The specific luminance patterns generated by shading and shadow contrast may be important for lightness constancy by providing the visual system with some information about surface albedo (i.e., lighter surfaces reflect more light to “fill in” shadows). However, the real advantage of surfaces over variegated surrounds appears to be derived from the specular highlights. This fits with an illumination-estimation account of the data. Light reflected from the diffuse component of a surface depends on both the (direct and indirect) light sources and the surface reflectance, resulting in variations in shadow and shading information between surfaces. Alternatively, the light reflected from the specular component depends primarily on the illumination intensity, causing specular highlights to remain constant in their appearance between surfaces. The visual system may be able to use this constant cue across conditions to estimate the direction and intensity of the illumination from specular highlights, better constraining lightness estimates.

Although an illumination-estimation account fits nicely with the data from Experiment 2, there are two other possible explanations for the difference in lightness constancy found between the 3-D glossy and 2-D glossy equivalent conditions. In Experiments 1 and 2, the adjustable patch was surrounded by a high-relief glossy surface with black and white checks (Figure 1E). We chose the specific pattern and material of the surround to make the test patch appear as surface-like as possible. Surface properties, such as lightness, have little meaning for flat 2-D stimuli, and observers often cannot easily differentiate between lightness and brightness (Blakeslee, Reetz, & McCourt, 2008). Although the pattern and material of the surround made the adjustable patch appear surface-like, this resulted in observers’ matches being more symmetric in the high-relief glossy trials compared to other trials, especially the low-relief condition in Experiment 1 and the variegated condition in Experiment 2. This suggests that the better constancy exhibited in the glossy trials may have been due to the similarity in the match and test surrounds. Experiment 3 addresses this possibility by replicating Experiments 1 and 2 but replacing the surround of the high-relief glossy adjustable patch with a 2-D matte surround.

Another alternative explanation of the pattern of data from Experiments 1 and 2 appeals to low-level mechanisms. Although the 2-D and 3-D surrounds are matched in terms of their luminance histograms, they differ in their contrast and luminance distributions across space and scale. For example, the brightest luminance values in the 3-D rocky surround (the specular highlights) are distributed across the entire surround (see Figure 1D). However, Figure 6 illustrates that the brightest luminance values from the 2-D glossy equivalent variegated surround are contained in a single square. Furthermore, the specular highlights in the 3-D rocky surround occur near points at which the diffuse shading gradient is brightest (Marlow, Kim, & Anderson, 2011). However, for the 2-D variegated surround, the contrast of adjacent squares is random. Experiment 4 directly tests the possibility that the specific contrast and luminance distributions in the 3-D rocky surrounds played a key role in affecting lightness perception in Experiments 1 and 2.

### Experiment 3: Flat, matte adjustable surface

Experiment 3 addresses whether the specific surround of the adjustable patch caused the pattern of results in Experiments 1 and 2 due to more symmetrical matching between the high-relief glossy test and adjustable displays.

#### Methods

**Observers**

Forty first-year psychology students participated in Experiment 3. All observers were naive and had not participated in any of the previous experiments. Twenty observers were assigned to the low-relief and the matte and glossy high-relief conditions of Experiment 1b. The other 20 observers were assigned to the matte equivalent and glossy equivalent variegated conditions of Experiment 2.

**Apparatus, stimuli, and procedure**

The task and conditions were identical to Experiments 1b and 2 except that the adjustable patch was rendered with a 2-D, matte checkerboard surround (Figure 1F).

#### Results and discussion

The results of Experiment 3 are presented in Figure 9. Two trials were excluded from analyses due to observers making premature lightness judgments (see Experiment 1 results). The data reveal that test patch lightness settings are affected by each surround in the same way as in Experiments 1 and 2. Figure 9 illustrates that crispening occurs in the low-relief conditions (top
right panel) but not in the high-relief conditions (middle right and bottom right panels). As in Experiment 1, increment–decrement difference scores were obtained to emphasize the size of the step for each of the low- and high-relief conditions (see Figure 10). Recall that a lower score indicates a smaller step and therefore less crispening.

$t$ tests using Sidak-corrected alpha values of .0253 per test were carried out to compare difference scores in the low- and high-relief conditions (see Table A5 for $t$ values, $df$, and $p$ values). The results replicate those of Experiment 1b except that the comparisons for surround Munsell value 5 were not significant (see Table A5). Nonlinearities diagnostic of crispening are present for the low-relief conditions (Figure 9, top right panel). Therefore, as in previous experiments, this slight discrepancy in results can be attributed to individual differences in observers’ ability to detect very low contrast test patches.

The results from the high-relief and variegated conditions are also replicated when the low-relief adjustable patch is used, which can be seen by the
average difference and standard deviation scores in Figure 11 (calculated the same way as in Experiments 1 and 2). Binomial sign tests on these scores revealed that for the 3-D rocky condition, lightness constancy is better when test patches are embedded in glossy compared to matte surrounds, \( p = 0.006 \). However, as in Experiment 2b, this is not the case for 2-D variegated conditions. Importantly, for the glossy condition, lightness constancy is better when test patches are embedded in 3-D rocky compared to 2-D variegated surrounds, \( p < 0.001 \). This is not the case for the matte condition, \( p = 0.5 \). These results suggest that the effects in Experiments 1 and 2 were not caused by the symmetry between the test and adjustable surfaces.

### Experiment 4: Phase-scrambled surrounds

Experiments 1 and 2 demonstrate that the luminance patterns generated by specular highlights improve lightness constancy for flat test patches embedded in rocky surrounds. Experiment 3 demonstrates that this improved lightness constancy is not simply the consequence of the test and match patches having similar surrounds. Experiment 4 aims to determine whether this improvement in lightness constancy is achieved through an improvement in the visual system’s ability to represent the light field or whether a low-level explanation involving contrast and luminance distributions across space and scale can account for the data. In Experiment 4, the 2-D equivalent surrounds were phase-scrambled versions of the 3-D rocky surrounds (see Figure 12). Scrambling the phase spectrum information ensures that the 2-D and 3-D surrounds are matched not only in terms of their luminance histograms, but also in terms of their spatial frequency content. The phase-scrambled surrounds tend to evoke impressions of grainy textures, which appear to be perceived predominantly as variations in pigment (particularly the specular highlights). If a representation of the light field via shading patterns and/or specular highlights is crucial for lightness constancy, then observers should exhibit better lightness constancy when test patches are embedded in 3-D rocky surrounds. Alternatively, if the enhanced lightness constancy of the glossy surfaces arises from the distribution of contrasts across spatial scales, no
difference in lightness constancy between the two conditions should be observed.

**Methods**

**Observers**

Twenty first-year psychology students participated in Experiment 4. All observers were naïve and had not participated in any of the previous experiments.

**Apparatus and stimuli**

The task was the same as in Experiments 1 and 2. Phase-scrambled surrounds were created by Fourier transforming the 3-D rocky images and replacing the phase spectrum information with that of white noise. Thus, phase-scrambled and 3-D rocky surrounds had equated pixel luminance histograms and power spectrums. As in Experiment 2, the regions immediately surrounding the central patch in the 3-D rocky and phase-scrambled stimuli were matched in average luminance.

**Procedure**

The layout of trials and instructions were the same as in Experiment 1 and 2. There were four surround-type conditions: 3-D matte, 3-D glossy, phase-scrambled matte equivalent, and phase-scrambled glossy equivalent. The surround Munsell conditions were the same as in the previous experiments. However, observers only performed lightness judgments on the 11 test patches common to all the surround types in Experiment 1 (see Table A2), i.e., no very low-contrast test patches were used. This resulted in 264 trials for each observer.

**Results and discussion**

The results of Experiment 4 are presented in Figure 13. The data reveal that test patch lightness settings differ between matte and glossy conditions, replicating the results of Experiment 1. However, lightness constancy does not differ between 3-D rocky and phase-scrambled surround conditions. Figure 14 plots the average difference scores and the standard deviation between adjacent Munsell conditions, calculated the same way as in Experiments 1 and 2. A lower score indicates better lightness constancy.

Binomial sign tests were carried out on the average difference scores and standard deviation scores. The results show that regardless of whether surrounds are 3-D rocky or phase-scrambled, observers exhibit better lightness constancy in the glossy compared to the matte condition, \( p = 0.006 \) (3-D rocky condition), \( p < 0.001 \) (phase-scrambled condition; Figure 14). The results also show no difference in lightness constancy between the 3-D rocky and phase-scrambled conditions, \( p = 0.11 \) (matte condition), \( p = 0.27 \) (glossy condition). This suggests that, for these types of stimuli at least, the visual system is not using information about the light field generated by specular reflections to improve lightness constancy for glossy surfaces; observers perform just as well when control surfaces are created with the same contrast and luminance distributions as naturalistic 3-D surfaces.

The results of this experiment also replicate the trends observed in the matte condition of Experiment
For both 3-D rocky and phase-scrambled conditions, there is a tendency for lighter test patches to be matched more consistently than darker test patches (indicated by the negative slopes of the matte data curves [closed circles] in Figure 14). Furthermore, as in Experiment 1, data curves in the matte condition are shifted vertically up in relation to ground truth (Figure 13, top row, compare data curves to the diagonal solid black line).

1: For both 3-D rocky and phase-scrambled conditions, there is a tendency for lighter test patches to be matched more consistently than darker test patches (indicated by the negative slopes of the matte data curves [closed circles] in Figure 14). Furthermore, as in Experiment 1, data curves in the matte condition are shifted vertically up in relation to ground truth (Figure 13, top row, compare data curves to the diagonal solid black line).

General discussion

Summary

The aim of the present study was to investigate whether the visual system uses cues to the illuminant created by complex mesostructure and specular highlights to improve lightness constancy of an embedded flat, matte test patch. The control experiments tested whether any benefits of shading and specular reflections could be attributed to articulation, the choice in matching patch surround, or differences in the energy across different spatial scales. In Experiment 1, observers judged the lightness of test patches surrounded by four different surrounds: flat matte, flat glossy, rocky matte, and rocky glossy. When the surrounds were rocky, there was a smooth, monotonic relationship between perceived lightness and test patch reflectance with glossy surfaces yielding better lightness constancy than matte surfaces. When the original rocky displays were compared to 2-D nonsurface control displays with matched pixel histograms (Experiment 2), lightness constancy was better for the rocky glossy compared to the 2-D glossy equivalent displays. However, this difference in lightness constancy was eliminated when the control displays had equated power spectra to the rocky displays via phase scrambling (Experiment 4). When surrounds were flat,
there was a nonlinear “step” or crispening (Takasaki, 1966) in the pattern of matches as test patch reflectance passed through that of the surround. Furthermore, there is an asymmetry between increment and decrement settings (see Figure 3, top four panels). Lightness settings for decrements tend to decrease as surround reflectance increases. However, the induction for increments is independent of surround color.

**Relation to previous work**

Contrary to the hypothesis that originally motivated our experiments, our results do not provide support for explanations of lightness perception based on illumination estimation, such as EIMs (Bloj et al., 2004; Boyaci, Maloney, & Hersh, 2003; Boyaci, Doerschner, & Maloney, 2004; for a review, see Brainard & Maloney, 2011). Our results suggest that, when computing test patch lightness, there is no benefit in observers’ lightness judgments for stimuli that contain surface relief, shadows, shading, and specular highlights. Rather, a low-level explanation involving contrast and luminance distributions across space and scale appear to be sufficient in explaining the pattern of results described above.

One theory of lightness perception that does not require an explicit representation of the illuminant is anchoring theory (Gilchrist, 2006; Gilchrist et al., 1999). Anchoring theory proposes that the visual system assigns a fixed lightness value (white) to the highest luminance in a scene, which serves as a lightness anchor. Other values in lightness are computed relative to this anchor point. It is unclear how these principles can be extended to naturalistic displays, such as those in the present study. The highest luminance in most natural scenes will not be generated by the surface with the highest diffuse reflectance but rather by specular reflections of the primary light source. Moreover, it is unclear how anchoring theory could account for the difference in lightness constancy between the variegated displays (Experiment 2) and the phase-scramble displays (Experiment 4), which have equated luminance histograms and therefore the same highest luminance.

One important consideration when interpreting the high-relief data is that all surfaces in the present study were rendered under the same illuminant. Consequently, test patch reflectance covaried with luminance, meaning that observers could have performed the task by matching perceived luminance (brightness), not lightness. Previous research has shown that lightness and brightness matches do not differ when illumination is homogeneous, and despite variations in illumination due to vignetting and interreflections, globally our displays were illuminated uniformly (i.e., there were no illumination boundaries or gradients). Thus, models that predict brightness may be able to account for the results. One such class of model involves spatial filtering with contrast normalization, such as Blakeslee and McCourt’s oriented difference-of-Gaussian multiscale filtering model (Blakeslee & McCourt, 1999, 2001, 2004; Blakeslee, Pasieka, & McCourt, 2005; see Shapiro & Lu, 2011, for an alternative model). While such models have the potential for dealing with more natural stimuli than anchoring theory, currently there is no principled way to specify the parameters of such models for arbitrary stimuli.
Another general criticism of spatial filtering models is that they do not predict effects of illumination and transparency on brightness/lightness perception. Lightness and brightness can differ substantially when illumination is inhomogeneous (Blakeslee & McCourt, 2012), suggesting that perceptual decomposition into layers (or scission; see Anderson, 1997; Anderson & Winawer, 2005, 2008) may affect lightness perception under some circumstances. Future studies should investigate whether low-level accounts involving distributions of contrasts can extend to lightness effects in naturalistic displays with illumination boundaries and/or gradients or whether midlevel (layered image decomposition) mechanisms may be involved.

The results from the homogeneous displays are in agreement with previous studies that have reported heightened discrimination for test patches that are close to the surround color (Krauskopf & Gegenfurtner, 1992) or luminance (Whittle, 1992). The crispening observed in the low-relief displays implies a qualitative difference in the way test patches are perceived when embedded in homogenous versus inhomogeneous surrounds. Some researchers have suggested that lightness perception may involve more than one perceptual dimension (Ekroll & Faul, 2013; Logvinenko & Maloney, 2006; Vladusich, 2012, 2013; Vladusich et al., 2007). Ekroll, Faul, and colleagues (Ekroll & Faul, 2009, 2012a, 2012b, 2013; Ekroll, Faul, & Niederée, 2004; Ekroll, Faul, & Wendt, 2011) observed crispening in homogeneous colored center-surround displays and suggested that impressions of transparency contributed to the appearance of the central patch. There is a significant body of work in lightness and color that has shown that layered image decompositions can induce large lightness and color induction effects (Anderson, 1997; Anderson & Khang, 2010; Anderson & Winawer, 2005, 2008; Wollschlager & Anderson, 2009). In Ekroll and Faul’s 2013 study, observers varied the transmittance and/or the color of an adjustable patch on a variegated surround to match a target on a uniform surround. The variegated and uniform surrounds had the same average luminance, which the authors argue eliminates the influence of von Kries adaptation to the crispening effect. They found that targets embedded in uniform surrounds were better matched when observers were allowed to vary both the physical color and transmittance of the matching patch on the variegated surround. Furthermore, transmittance and saturation settings were inversely related to the chromatic contrast between the target patch and its surround; at low chromatic contrasts, homogeneous center-surround stimuli appeared to trigger impressions of transparency with the target region being perceptually divided into a saturated, transparent filter layer and a background layer the same color and saturation as the surround. Support for this view was also provided by Wollschlager and Anderson (2009), who showed that similar forms of image decomposition can be evoked in textured displays when the center and surround have similar textures and satisfy the conditions for transparency (see also Anderson & Khang, 2010). Future research could explore whether this finding extends to lightness displays. Furthermore, allowing observers to adjust transmittance might shed light on the asymmetry between increment and decrement settings because the ambiguity of these homogeneous displays allows arbitrary amounts of luminance to be attributed to the background and filter layers.

If perceptual transparency does affect the lightness of test patches embedded in homogeneous surrounds, as suggested by the crispening effect and Ekroll and Faul’s (2013) findings, then it suggests that two separate mechanisms may be responsible for the perception of lightness in complex and homogenous center-surround displays: The effects of complex surface mesostructure and surface optics on lightness are well explained by low-level contrast and luminance distributions across space and scale; conversely, lightness in homogeneous displays, which are descriptively simple, may involve segmentation of surfaces into layered image representations. Paradoxically, the homogeneous displays might evoke more complex scene representations than the ostensibly complex rocky displays.

Regardless of whether lightness representations differ for homogeneous and inhomogeneous displays, the difference in the pattern of data between these two types of displays has important implications for a large body of literature investigating the simultaneous contrast (SC) effect. Previous studies have compared the size of the SC effect in uniform versus variegated center-surround displays. It is often reported that SC is enhanced when surrounds are articulated (Adelson, 2000; Arend & Goldstein, 1987; Bressan & Actis-Grosso, 2006; Gilchrist et al., 1999; Schirillo, 1999), suggesting either poorer or better lightness constancy for articulated displays, depending on the experimental context. These displays are similar to the low-relief (Experiment 1) and variegated (Experiment 2) displays used in the present study. Our results suggest that the enhanced SC effect for variegated displays may be a result of researchers evaluating lightness constancy using only a few test patch and surround values (or a failure to adequately equate the luminance of the surrounds). One advantage of the factorial combination of conditions used in the present study is that it reveals the complicated nature of the SC effect on homogeneous surrounds. For uniform displays, the SC effect clearly depends on the test patch’s contrast against its surround (crispening). At low contrasts, these test patches seem to take on a qualitatively
different appearance compared to test patches embedded in rocky or variegated displays. Therefore, it may not be meaningful to compare lightness constancy between uniform and articulated SC displays for single target and surround combinations.

One final point to note is the size of the SC effect in our study. Perceived lightness of identical targets differed by as much as 4 Munsell steps when embedded in different surrounds. Even in the rocky glossy conditions, perceived lightness could differ by 2 Munsell steps or more. These effects are much larger than what is normally reported in the literature, which is about 0.5 to 1 Munsell step (Gilchrist, 2006). We are unsure why the effect is so strong for these stimuli although we suspect it has to do with the large range of test patch and surround reflectance values used.

Conclusions

The present study does not provide support for lightness perception based on illumination estimation. Lightness perception in putatively complex displays is well explained by low-level distributions of contrasts across space and scale. However, the crispening effect in the low-relief displays suggests that different mechanisms play a role in the appearance of test patches embedded in homogeneous surrounds. Further investigation is required to determine whether lightness perception in homogeneous displays involves midlevel segmentation of surfaces into layered image representations and whether low-level explanations can extend to complex displays with inhomogeneous illumination.

Keywords: lightness, brightness, lightness constancy, reflectance, albedo, material perception, surface perception, natural light field, mesostructure, gloss, illumination estimation, contrast distributions, center-surround displays, simultaneous contrast, simultaneous lightness contrast, crispening

Acknowledgments

This research was supported by grants from the ARC to B. L. Anderson.

Commercial relationships: none.

Corresponding authors: Alexandra C. Schmid and Barton L. Anderson.

Email: asch9222@uni.sydney.edu.au; barton.anderson@sydney.edu.au.

Address: School of Psychology, The University of Sydney, NSW, Australia.

Footnotes

1This material is not limited to the physical light-scattering properties of plastic; rather, it can be used for a wide variety of reflective materials, such as surfaces made of concrete, wood, paint, etc.

2Rendering with two ambient reflections allows shadowed areas of the rocky surfaces to be indirectly illuminated by other parts of the surface as would occur in natural scenes. This was needed to test the hypothesis that observers use the amount of shadow “filling-in” as a cue to surface lightness.

3Follow-up paired t tests using Sidak-corrected alpha values of 0.0398 per test indicated that the matte surround induced a larger step than the glossy surround only for the low-relief condition, t(4) = 3.62, p = 0.02, not the high-relief condition, t(4) = −1.74, p = 0.16. Additionally, the low-relief surround induced a larger step than the high-relief surround only when the surround was matte, t(4) = 5.86, p = 0.004, not glossy, t(4) = 0.84, p = 0.45. Sidak-corrected alpha values were calculated as x = 1 − (1 − FWER)1/n = 1 − (1 − 0.15)1/4 = 0.0398, where FWER is the family-wise error-rate for the ANOVA.

4Sidak-corrected alpha values were calculated as x1 = 1 − (1 − x)1/n = 1 − (1 − 0.05)1/2 = 0.0253.

5For simplicity, only p values for tests on the average difference scores are reported. However, the results are identical for standard deviation scores.

References


Appendix A: How surrounds were created for Experiment 1

Each surface was created in Blender using an 800 × 800 mesh. The textures in each surround were generated with the displace modifier, a tool in Blender that displaced vertices in a mesh based on the intensity of a texture. Various textures were used to deform the surfaces: the inbuilt cloud, marble, and stucci textures as well as textures from images of rocks and rough paper. The image of rough paper used is displayed in Figure A1. The rocky texture image can be found at http://junk-paris-stock.deviantart.com/art/macro-rock-texture-13-119245673. Table A1 shows the modifiers that were used for low- and high-relief surfaces and the order in which they were applied. Note that although the rough paper texture was used to displace vertices in the low-relief surrounds, this effect was extremely subtle, so the rendered images
were essentially homogeneous. Also note that, although multiple textures were used to deform the high-relief surfaces, we refer to them as “rocky” because of their rocky appearance after rendering.

Appendix B: Tone mapping of RADIANCE HDR images

The procedure used to tone-map each HDR image is as follows: The diffuse component was linearly compressed by transforming luminance values below 140 cd/m² with the equation

\[ L_i^D = \frac{H_i^D}{H_{\text{max}}^D} \cdot L_{\text{max}}^D, \]

where \( L_i^D \) is the transformed luminance associated with the diffuse component for each pixel \( i \), \( H_i^D \) is the original HDR luminance associated with the diffuse component for each pixel \( i \), \( H_{\text{max}}^D \) is the maximum HDR luminance attributed to diffuse reflectance, and \( L_{\text{max}}^D \) is the maximum luminance assigned to diffuse reflectance in the tone-mapped (transformed) image. \( H_{\text{max}}^D \) was constant for all images and was equal to 140. \( L_{\text{max}}^D \) was constant for all images and was equal to 53.59. Thus, the brightest regions of diffuse shading in the tone-mapped image had a luminance of approximately 53.59 cd/m².

The specular component (HDR luminance values above 140 cd/m²) was compressed nonlinearly to create smooth fall-off of luminance values that started at \( L_{\text{max}}^S \) (53.59 cd/m²) and peaked at \( L_{\text{max}}^S \), the luminance assigned to the brightest specular highlight and also the brightest luminance of the monitor (64.98 cd/m²; see Figure A2). We achieved this by first subtracting \( H_{\text{max}}^S \) (140 cd/m²) from each pixel and then transforming these values with the equation

\[ L_i^S = (-e^{-\frac{1}{\kappa_m S_i}} + 1) \cdot R, \]

where \( L_i^S \) is the transformed luminance associated with the specular component for each pixel \( i \), \( R \) is the luminance range of the specular highlights and is equal to \( L_{\text{max}}^S - L_{\text{max}}^D \), \( S \) is the slope of the straight line from the linear transformation of the diffuse component and is equal to \( L_{\text{max}}^D / H_{\text{max}}^D \), and \( H_i^S \) is the HDR luminance associated with the specular component for each pixel \( i \). Finally, we added \( L_{\text{max}}^D \) to these specular values, and the result was a tone-mapped image with linearly transformed diffuse shading and nonlinearly transformed specular highlights (Figure A2).

The last step was to display the images using the eight-bit pixel values of the monitor. For this, we made a color look-up table (CLUT) of luminance values corresponding to each eight-bit pixel value (0–255). Each luminance value in the tone-mapped image was transformed into its corresponding CLUT value.

Appendix C: Tables

<table>
<thead>
<tr>
<th>Modifier</th>
<th>Low-relief surfaces</th>
<th>High-relief surfaces</th>
</tr>
</thead>
<tbody>
<tr>
<td>Displace: stucci (inbuilt)</td>
<td>N/A</td>
<td>Strength 0.2</td>
</tr>
<tr>
<td>Displace: marble (inbuilt)</td>
<td>N/A</td>
<td>Strength 0.1</td>
</tr>
<tr>
<td>Displace: clouds (inbuilt)</td>
<td>N/A</td>
<td>Strength 0.07</td>
</tr>
<tr>
<td>Smooth</td>
<td>N/A</td>
<td>Factor 0.5; Repeat 50</td>
</tr>
<tr>
<td>Paper</td>
<td>Strength 0.01</td>
<td>N/A</td>
</tr>
<tr>
<td>Smooth</td>
<td>Factor 0.5; Repeat 50</td>
<td></td>
</tr>
</tbody>
</table>

Table A1. Modifiers used to create the effects in the low- and high-relief surrounds. Notes: The order in which the modifiers are displayed is the order in which they were applied (left column, top to bottom).
<table>
<thead>
<tr>
<th>Surround % reflectance</th>
<th>Surround Munsell values</th>
<th>Surround luminance minimum (cd/m²)</th>
<th>Surround luminance maximum (cd/m²)</th>
<th>Extra test patch % reflectance</th>
<th>Extra test patch Munsell values</th>
<th>Extra test patch luminance values (cd/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>3, 5.2, 9, 13.7, 19.8, 27.2, 36.2, 46.8, 59.1, 73.4, 90</td>
<td>1.95, 2.75, 3.5, 4.25, 5.75, 6.5, 7.25, 8, 8.75, 9.5</td>
<td>1.17, 2.25, 3.34, 5.13, 7.3, 9.83, 13.06, 17.01, 21.34, 26.40, 32.49</td>
</tr>
<tr>
<td>3</td>
<td>1.95</td>
<td>M: 0.031, G: 0.049</td>
<td>M: 2.2, G: 64.98</td>
<td>3.39, 3.82</td>
<td>2.1, 2.25</td>
<td>1.17, 1.49</td>
</tr>
<tr>
<td>9</td>
<td>3.5</td>
<td>M: 0.11, G: 0.12</td>
<td>M: 6.63, G: 64.98</td>
<td>7.47, 8.21, 9.84, 10.7</td>
<td>3.2, 3.35, 3.65, 3.8</td>
<td>2.56, 2.92, 3.67, 4.02</td>
</tr>
<tr>
<td>19.8</td>
<td>5</td>
<td>M: 0.32, G: 0.32</td>
<td>M: 14.29, G: 64.98</td>
<td>17.2, 18.4, 21.1, 22.6</td>
<td>4.7, 4.85, 5.15, 5.3</td>
<td>6.21, 6.56, 7.65, 8.38</td>
</tr>
<tr>
<td>36.2</td>
<td>6.5</td>
<td>M: 0.65, G: 0.65</td>
<td>M: 26.5, G: 64.98</td>
<td>32.4, 34.3, 38.2, 40.2</td>
<td>6.2, 6.35, 6.65, 6.8</td>
<td>11.64, 12.34, 13.77, 14.52</td>
</tr>
<tr>
<td>59.1</td>
<td>8</td>
<td>M: 1.28, G: 1.38</td>
<td>M: 43.44, G: 64.98</td>
<td>53.9, 56.5, 61.8, 64.6</td>
<td>7.7, 7.85, 8.15, 8.3</td>
<td>19.57, 20.32, 22.4, 23.51</td>
</tr>
<tr>
<td>90</td>
<td>9.5</td>
<td>M: 2.63, G: 2.75</td>
<td>M: 61.03, G: 64.98</td>
<td>83.1, 86.5</td>
<td>9.2, 9.35</td>
<td>30.07, 31.51</td>
</tr>
</tbody>
</table>

Table A2. Surround and center patch reflectance and luminance values of the test surfaces used in the experiments. Notes: The first row shows the test patch values that were common to all surround types. The fifth column shows this in percentage reflectance, the sixth column shows this in Munsell values, and the seventh column shows this in luminance values. For all remaining rows, the first column contains the six reflectance values (percentage reflectance) used for the surrounds. The second column shows the values in column 1 transformed to the Munsell scale. The third and fourth columns show the luminance range of the surrounds (M = matte, G = glossy). The fifth column contains the extra two to four reflectance values (percentage reflectance) of the center patches that were very close in lightness and unique to each surround. Two of these values were increments and two were decrements except for the black surround (3% reflectance), which contained only two extra increments, and the white surround (90% reflectance), which contained only two extra decrements. The sixth column shows the values in column 5 transformed to the Munsell scale, and the seventh column displays the luminance values of the test patches.
### Table A3. $F$ values and $p$ values for the increment–decrement difference scores of Experiment 1a. Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. For all tests, $df = (1, 4)$.

<table>
<thead>
<tr>
<th>Surround 3.5</th>
<th>Surround 5</th>
<th>Surround 6.5</th>
<th>Surround 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$</td>
<td>$p$</td>
<td>$F$</td>
<td>$p$</td>
</tr>
<tr>
<td>Relief</td>
<td>19.843</td>
<td><strong>0.011</strong></td>
<td>27.013</td>
</tr>
<tr>
<td>Gloss</td>
<td>0.028</td>
<td>0.874</td>
<td>6.062</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.1</td>
<td>0.768</td>
<td>1.794</td>
</tr>
</tbody>
</table>

### Table A4. $t$ values, $p$ values, and $df$ for the increment–decrement difference scores of Experiment 1b. Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

<table>
<thead>
<tr>
<th>Surround 3.5</th>
<th>Surround 5</th>
<th>Surround 6.5</th>
<th>Surround 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$df$</td>
<td>$t$</td>
<td>$p$</td>
<td>$df$</td>
</tr>
<tr>
<td>LR versus HR-M</td>
<td>19</td>
<td>3.67</td>
<td><strong>0.002</strong></td>
</tr>
<tr>
<td>LR versus HR-G</td>
<td>19</td>
<td>2.43</td>
<td><strong>0.0249</strong></td>
</tr>
</tbody>
</table>

### Table A5. $t$ values, $p$ values, and $df$ for the increment–decrement difference scores of Experiment 3. Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

<table>
<thead>
<tr>
<th>Surround 3.5</th>
<th>Surround 5</th>
<th>Surround 6.5</th>
<th>Surround 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$df$</td>
<td>$t$</td>
<td>$p$</td>
<td>$df$</td>
</tr>
<tr>
<td>LR versus HR-M</td>
<td>19</td>
<td>4.52</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>LR versus HR-G</td>
<td>19</td>
<td>4.57</td>
<td>$&lt;0.001$</td>
</tr>
</tbody>
</table>