Lightness matching and perceptual similarity

Khushbu Y. Patel

Department of Psychology and Centre for Vision Research, York University, Toronto, Canada ✉

Anudhi P. Munasinghe

Department of Psychology and Centre for Vision Research, York University, Toronto, Canada
Department of Psychological and Brain Sciences, University of California, Santa Barbara, CA, USA ✉

Richard F. Murray

Department of Psychology and Centre for Vision Research, York University, Toronto, Canada 🏠 ✉

Lightness constancy is the ability to perceive surface reflectance correctly despite substantial changes in lighting intensity. A classic view is that lightness constancy is the result of a "discounting" of lighting intensity, and this continues to be a prominent view today. Logvinenko and Maloney (2006) have proposed an alternative approach to understanding lightness constancy, in which observers do not make explicit estimates of reflectance, and lightness constancy is instead based on a perceptual similarity metric that depends on both the reflectance and the illuminance of surfaces viewed under different lighting conditions. Here we compare these two views using a novel, free-adjustment reflectance-matching task. We test whether observers can match reflectance in a task where they are free to adjust both the illuminance and the reflectance of the match stimulus over a wide range. We find that observers can match reflectance under these conditions, which supports the view that observers make explicit estimates of reflectance. We also compare performance in this free adjustment task using physical objects and computer-rendered images as stimuli. We find that lightness constancy is good in both cases, but with some evidence of a glow-related artifact with computer-rendered stimuli.

Introduction

People can reliably perceive the colors of surfaces in complex scenes, even though the spectrum and intensity of the light that a surface patch reflects to the eye depends not only on the reflectance properties of the patch itself, but also on illumination conditions and scene geometry. A special case of this ability is lightness perception, the ability to perceive the reflectance of black, white, and gray surfaces. Reflectance is the proportion of incident light reflected by a surface, as measured in photometric units, and lightness is perceived reflectance. Even the most basic principles of lightness perception are still contested. Some theories hold that lightness perception is based on low-level image properties such as bandpass energy (e.g., Blakeslee & McCourt, 1999; Shapiro & Lu, 2011), while others claim that midlevel features such as cues to surface boundaries and lighting conditions play a crucial role (e.g., Adelson, 2000; Bloj et al., 2004; Gilchrist, 2006; Murray, 2013). Our understanding of lightness perception is still so tentative that there is even room for basic questions about how to describe percepts of simple gray surfaces. Logvinenko and Maloney (2006) used difference scaling methods to investigate the perceptual dimensions of matte gray surfaces, and developed a quantitative model of how reflectance and illumination contribute to the perceptual similarity or dissimilarity of surface patches. Figure 1 illustrates their model. The *x*-axis represents the reflectance of an achromatic surface patch, and the *y*-axis represents illuminance (i.e., intensity of incoming illumination). Both axes are scaled logarithmically. The white dot represents the reflectance and illuminance of a reference patch. Any other surface patch has some degree of perceptual similarity to the reference patch, and the gray curves show iso-similarity contours according to Logvinenko and Maloney's model (their equation 5). Human observers judge all points on each iso-similarity contour to be equally similar to the reference patch. The iso-similarity contours illustrate several of Logvinenko and Maloney's findings. First, when measured in logarithmic units, reflectance differences contribute more to perceptual dissimilarity than illumination differences do, so iso-similarity contours

are closer to the reference patch in the *x* direction than in the *y* direction. Second, the diamond-like shape of the contours shows that similarity is given approximately by a city-block metric: To find the approximate similarity between the reference patch and another patch we take a weighted sum of the difference between them in log reflectance and in log illuminance. A third, subtler phenomenon is that two reflectances are seen as slightly less similar under high illuminance than under low illuminance. This is indicated by the inward bowing of the upper half of the diamonds and the outward bowing of the lower half.

Logvinenko and Maloney (2006) pointed out that in addition to establishing fundamental facts about the perceptual dimensions of achromatic surfaces, these findings also suggest a new theory of lightness matching. In a typical asymmetric lightness-matching task the observer sees a reference patch under one illuminant, as well as a set of match patches under a different illuminant, and the observer's task is to choose the match patch that has the same reflectance as the reference patch. We can represent the reference patch as the white dot in Figure 1, and the match patches as the black dots along a horizontal iso-illuminance line. The usual view of lightness matching is that the observer estimates the reflectance of the reference and match patches (and investigating exactly how the observer computes reflectance is often the goal of such experiments), and then chooses the match patch with the reflectance estimate closest to that of the reference patch. Logvinenko and Maloney suggested that, in fact, observers might not estimate reflectance at all in such tasks. They pointed out that because typical lightness matching experiments show all match patches under a single illuminant, the observer does not have to compute reflectance in order to make a reflectance match. Instead, the observer can simply choose the match patch that is perceptually most similar to the reference patch according to the similarity measure illustrated in Figure 1, and this match will have approximately the same reflectance as the reference patch. That is, the observer solves a constrained optimization problem, matching reflectance by maximizing perceptual similarity along the iso-illuminance line, without computing a separate estimate of reflectance per se.

Expanding on this theory, Logvinenko and Maloney (2006) suggested that there are at least two dimensions to achromatic surface appearance: lightness, which depends mostly on surface reflectance, and surface brightness, which depends mostly on the intensity of incident illumination. Earlier researchers have also argued that lightness and brightness are distinct perceptual dimensions (e.g., Arend & Spehar, 1993); lightness is usually defined as perceived reflectance, and brightness as perceived luminance. One of Logvinenko
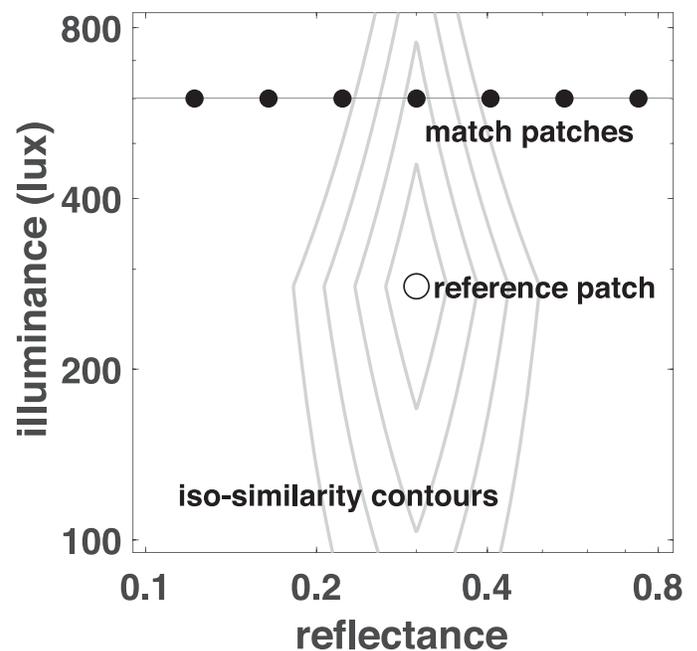

Figure 1. A perceptual similarity measure for achromatic surfaces. The white circle represents the reference patch in a lightness-matching task, and the gray curves are iso-similarity contours relative to the reference patch, based on Logvinenko and Maloney's (2006) similarity metric. The black circles represent match patches with various reflectances at a fixed illumination level. For this illustration, we chose an arbitrary value for the parameter of the similarity metric that controls the slight pinching of the upper half of the iso-similarity contours and the bowing of the lower half.

and Maloney's contributions is to show that two such dimensions emerge spontaneously from a multidimensional scaling analysis that makes no prior assumptions about the perceptual dimensions of achromatic surfaces. Furthermore, it is sometimes ambiguous in the literature whether brightness is meant to be a property of an image or of a reflective surface depicted in an image; in Logvinenko and Maloney's account, brightness is unambiguously a property of the surface (hence "surface brightness"). Finally, brightness is usually defined as perceived luminance, but in Logvinenko and Maloney's account, surface brightness depends mostly on incident illumination, not on the luminance of a surface, which depends on both illuminance and reflectance.

Logvinenko and Maloney (2006) also suggested that observers are unable to judge these two perceptual dimensions independently—for example, they are unable to abstract the lightness values of two surfaces under different illuminants and compare the lightness values directly. When asked to choose a match patch whose reflectance is the same as that of a reference patch, observers simply choose the match patch that is

most similar to the reference patch according to the measure illustrated in Figure 1. Logvinenko and Maloney noted that this hypothesis explains why observers are sometimes dissatisfied in asymmetric matching tasks, and report that no available match patch perfectly matches the reference patch (e.g., Adelson, 1993, footnote 8). According to the similarity-maximizing theory, the observer chooses the match patch that is most similar to the reference patch (as in Figure 1), but because the reference and match patches are seen under different illuminants, even this optimal match has some nonzero level of dissimilarity to the reference patch. The match patch that the observer chooses may have the same lightness as the reference patch, but the observer is unable to judge this perceptual dimension alone. Thus the matching task is designed in such a way that the observer's attempt to choose a perfectly similar match cannot succeed.

This theory of asymmetric lightness matching takes a fundamentally new approach to lightness perception, in that it suggests that observers do not actually estimate reflectance, or even a proxy of reflectance (e.g., Blakeslee & McCourt, 1999). It implies that observers' ability to match reflectance depends strongly on the design of the experiment: When all match patches are on a single iso-illuminance line, an observer who does not perceive reflectance can nevertheless match reflectance by maximizing a measure of perceptual similarity that depends on both reflectance and illuminance.

To develop a test of this theory, we reasoned that if reflectance matches are not based on reflectance estimates, and if people's ability to match reflectance is a byproduct of how lightness matching experiments are typically designed, then it should be possible to find a new experimental design that makes it difficult or impossible to match reflectance by minimizing perceptual dissimilarity. Here we test whether observers can match reflectance in a task where they are free to adjust both the illuminance and the reflectance of the match stimulus over a wide range. In this task, the match stimuli do not all lie on an iso-illuminance line, so there is no constraint that allows observers to make a correct reflectance match by choosing the stimulus that is least dissimilar to the reference stimulus. If observers have no estimate of reflectance, but only a similarity measure that depends on both reflectance and illuminance (as in Figure 1), then in a task where they have control over both reflectance and illuminance they should find it difficult or impossible to make a reflectance match. To anticipate, we find that observers can make reflectance matches in such a free-adjustment task, and this result supports the alternative hypothesis that observers use explicit estimates of reflectance to make lightness matches.

Experiment 1

Methods

Participants

There were six observers. One was author KP. The others were unaware of the purpose of the experiment and participated for payment. The observers were 19 to 24 years old, and four were female. In both experiments reported here, all observers gave written informed consent and reported normal vision, and all procedures were approved by the Office of Research Ethics at York University.

Stimuli

The stimuli were modeled after those in Logvinenko and Maloney's (2006) experiment 1. Observers viewed a panel of white poster paper that measured 61 cm horizontally and 50 cm vertically, at a viewing distance of 170 cm, subtending 20.3° × 16.7° (Figure 2). The panel had a reflectance of 0.78 (i.e., 78%), and was printed with a field of small, randomly placed circles with diameters ranging from 0.05 to 0.30 cm and reflectances ranging from 0.08 to 0.75. The panel had two hexagonal apertures with maximal diameters of 5.0 cm (1.7°), through which papers of various reflectances could be seen (details below). The distance between the centers of the two apertures was 25.5 cm (8.2°). Six trapezoids with reflectances of 0.07, 0.10, 0.21, 0.29, 0.45, and 0.74 were placed around each aperture. The trapezoids were 6.9-cm long (2.3°), and 2.8-cm wide (0.9°) at the shorter end and 7.8-cm wide (2.6°) at the longer end. The gap between the apertures and the trapezoids was 1.0-cm (0.3°) wide.

We used computer-controlled motors to show papers of various reflectances through the two hexagonal apertures. Immediately behind the reference aperture (Figure 2, right-hand hexagon), flat against the back of the main textured panel described above, was a circle of poster paper with a diameter of 31.0 cm. The circle was divided into three equal sectors with reflectances of 0.20, 0.31, and 0.49. A computer-controlled motor was able to rotate the circle so that a uniform section of paper with one of the three reflectances was visible through the reference aperture. Immediately behind the match aperture (Figure 2, left-hand hexagon), flat against the back of the panel, was a conveyor belt–like loop of poster paper, 152-cm long and 11-cm wide. The reflectance of the conveyor belt ranged continuously from 0.06 to 0.80. Reflectance increased nonlinearly along the belt, in such a way that reflectances were evenly spaced according to the Munsell scale (Schanda, 2007). A computer-controlled motor was able to advance the conveyor belt so that any section of the
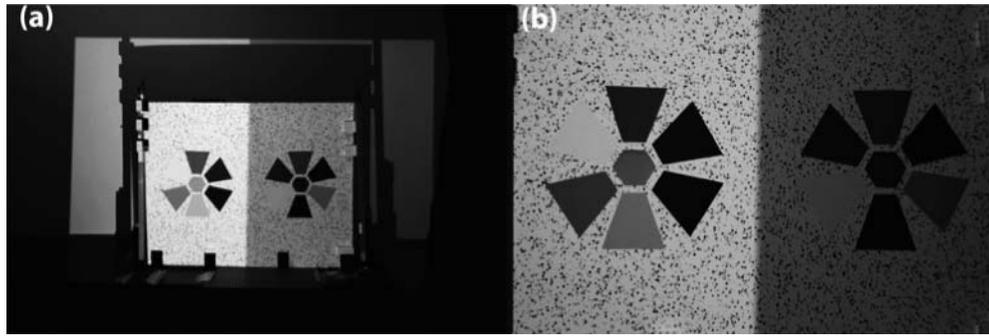
Figure 2. Apparatus for Experiment 1. (a) The observer's view. (b) Close-up of the central panel. The observer adjusted the illuminance on the left-hand side and the reflectance of the left-hand hexagon (the match hexagon). The illuminance on the right-hand side was fixed, and the reflectance of the right-hand hexagon (the reference hexagon) varied from trial to trial.

belt was visible through the match aperture. Just 5 cm of the 152-cm belt was visible through the match aperture at any time, so the visible segment of paper always had approximately uniform reflectance, and the reflectance gradient was not perceptible.

We created the main panel, three-sector circle, and conveyor belt by printing them on a poster printer that we had characterized by measuring the mapping from grayscale RGB values to reflectances. We measured reflectances using a Minolta photometer (LS-110; Konika Minolta, Tokyo, Japan) and a Labsphere Spectralon 99% diffuse reflectance standard (SRS-99-020; Labsphere, North Sutton, NH). Under constant illumination, we measured the luminance l_S of the reflectance standard and the luminance l_T of the target surface whose reflectance was being measured, and assigned the target surface the reflectance $r_T = 0.99 l_T/l_S$.

The apparatus was illuminated by overhead fluorescent lights, and by an LCD data projector (Epson PowerLite HC2040) placed 104 cm in front of the panel. We characterized the data projector using a photometer and reflectance standard (same models as above) to measure the mapping from grayscale RGB values to illuminances at the reference and match apertures. During the experiment, lighting on the right-hand side of the main panel was fixed, with an illuminance of 314 lux at the reference aperture, meaning that a surface with a reflectance of 1.0 would have a luminance of 100 cd/m^2. Lighting on the left-hand side of the panel was adjustable via a computer program that controlled the data projector, with illuminance at the match patch adjustable between 628 and 2,512 lux, meaning that a Lambertian surface with a reflectance of 1.0 would have a luminance between 200 and 800 cd/m^2. There was very little light scatter: Increasing the illuminance from minimum to maximum on the match side increased the illuminance on the reference side by just 4%. The lighting boundary between the left- and right-hand sides was created by the data projector, which projected a two-tone image

that was brighter on the left than on the right. We defocused the projector slightly so that no pixelation was visible in the projected image, and so that the lighting boundary was penumbra-like and unlikely to be mistaken for a reflectance boundary.

Procedure

Each observer ran in one 30-min session that included 120 trials. At the start of each trial, the printed circle behind the reference aperture was rotated to show one of the three reference reflectances, randomly chosen with the constraint that the same reflectance was not shown on consecutive trials. Also at the start of each trial, the conveyor belt behind the match aperture was advanced to show a random initial reflectance, and the data projector set the illuminance at the match aperture to a randomly chosen value in the adjustable range given above. After these initial settings were made, a short beep indicated to the observer that the apparatus was ready for their response. The observer used two keys on a computer keyboard to increase or decrease the match reflectance via the conveyor belt, and two other keys to increase or decrease the match illuminance via the data projector. The adjustments were effectively continuous, with no artificially introduced steps in reflectance or illuminance. A brief tap of the reflectance keys moved the conveyor belt 1 to 2 cm of its 152-cm length, which adjusted reflectance by about 1%. Illuminance steps were limited only by the 256 gray levels of the projector. The observer was instructed to adjust the paper and lighting at the match aperture so that the paper appeared to be the same shade of gray paper as the paper visible through the reference aperture, but to be illuminated at a different lighting intensity than the reference aperture. On a few practice trials before the main experiment, we trained the observer to adjust both reflectance and illuminance on every trial. The observer had unlimited time to respond, and pressed a fifth key to indicate that they had finished the trial. A short beep acknowledged their

response, and then the next trial began. The median duration of the free-adjustment phase of the trial was 16 s, across all observers and trials. The relatively long response time was largely due to the slow speed of the apparatus.

As described above, the adjustable match illuminance range was higher than the fixed reference illuminance. We chose this design so that observers could not adjust the match illuminance to equal the reference illuminance, because then reflectance matching could be accomplished via luminance matching. In the present design, observers always had to match reflectance across two regions of very different illuminance, which provides a stronger test of lightness constancy. Our results support the theory that observers make lightness matches by estimating and comparing reflectance, so we expect that we would find similar results if the match illuminance range was lower than the reference illuminance, but this would need to be confirmed experimentally.

We discarded all trials where observers did not adjust both reflectance and illuminance, since on these trials observers were effectively choosing a stimulus along an iso-illuminance or iso-reflectance constraint line, instead of freely adjusting both reflectance and illuminance as required by the experimental design. On average, observers adjusted both dimensions on 70% of the trials.

Results and discussion

Figure 3 shows the reflectance and illuminance of the observers' match settings (small circular data points), along with the three reference patches (large circular data points). The yellow, blue, and green points show the conditions where the reference reflectance was 0.20, 0.31, and 0.49, respectively. The match settings can be compared with two lines on the graphs: The vertical dashed lines show iso-reflectance lines, where points have the same reflectance as the reference stimuli, and the diagonal dotted lines show iso-luminance lines, where points have the same luminance as the reference stimuli. If an observer has perfect lightness constancy, they will choose matches along the iso-reflectance lines. If an observer has no lightness constancy, and simply matches the luminance of the reference and match stimuli, then they will choose matches along the iso-luminance lines. Here observers' matches were close to the vertical iso-reflectance lines, but with some influence of illuminance, indicating imperfect discounting of illumination and hence partial lightness constancy, as is commonly found in studies of lightness constancy and color constancy (e.g., Kraft & Brainard, 1999). The Thouless ratio is a common measure of lightness constancy (Thouless, 1931), defined as:

$$\tau = \frac{\log(r_M) - \log(r_0)}{\log(r_R) - \log(r_0)} \quad (1)$$

Here r_R is the reference patch reflectance, r_M is the match reflectance chosen by the observer, and r_0 is the match reflectance that would be chosen by an observer who has no lightness constancy and simply matches image luminance. A Thouless ratio of one indicates perfect lightness constancy, and a value of zero indicates that the observer has no constancy and is matching luminance. In the appendix, we show that on log-log axes, the match points corresponding to a fixed Thouless ratio τ form a straight line that has slope $(\tau - 1)^{-1}$ and passes through the point $(\log r_R, \log i_R)$ that represents the reflectance r_R and illuminance i_R of the reference patch. In Figure 3 we show a sum-of-squares fit of a straight line to each observer's match settings, constrained to pass through the reference patch; this shows the best fit of the model outlined in the Appendix, where the Thouless ratio is assumed to be constant across illuminance levels. Figure 3 also shows the Thouless ratio τ corresponding to each fitted line. Estimates of τ ranged from 0.67 to 0.94, and bootstrapped 95% confidence intervals ranged from ±0.02 to ±0.06. These Thouless ratios indicate relatively good lightness constancy. For comparison, Gilchrist (2006) calculated the Thouless ratios in some of Katz's classic experiments on lightness constancy, and found values ranging from 0.35 to 0.70.

An alternative method of quantifying lightness constancy is to find the average of the Thouless ratios calculated from individual match points. This approach gave very similar results to the model-fitting method described above, for all observers and conditions. For example, observer EH in Experiment 1 had average point-by-point Thouless ratios of 0.72, 0.66, and 0.83 in the three reference stimulus conditions, which are similar to the fitted Thouless ratios of 0.75, 0.67 and 0.87 (Figure 3).

At the end of this experiment, and also at the end of Experiment 2, we informally asked observers if they were satisfied with their lightness matches, and whether there were trials on which a good match was not possible. Observers described the task as an easy one, where it was always possible to make a good match.

Clearly observers were able to match reflectance in this task where they freely adjusted both reflectance and illuminance. This is consistent with Rutherford and Brainard (2002), who also found that observers had reasonably good lightness constancy in tasks where they could adjust both reflectance and illuminance. The set of possible match stimuli was not on a single iso-illuminance line, and so observers cannot have been making reflectance matches by maximizing similarity between the reference patch and the match patch, using a similarity measure like the one shown in Figure 1. It appears that instead, observers compute reflectance (e.g., Gilchrist et al., 1999) or some proxy of reflectance
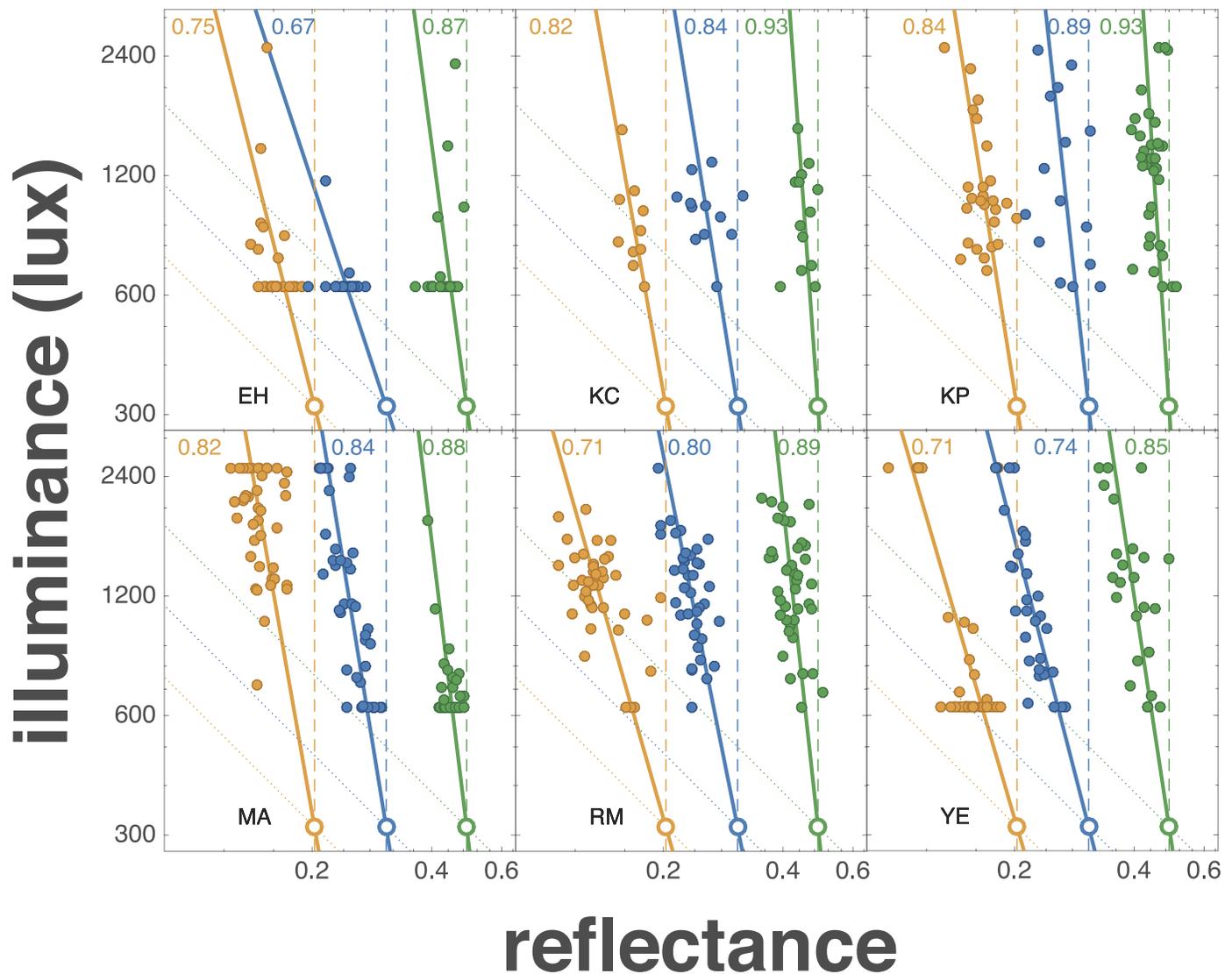
Figure 3. Results of Experiment 1. The small circles show reflectance and illuminance matches for all six observers. Yellow, blue, and green represent conditions with reference reflectances 0.20, 0.31, and 0.49, respectively. The white dots with a colored outline represent the reference patches. The solid lines are linear fits to the data, constrained to pass through the point representing the reference patch. Thouless ratios are shown beside each line. The vertical dashed lines are iso-reflectance lines and the diagonal dotted lines are iso-luminance lines.

(e.g., Blakeslee & McCourt, 1999), and choose the match stimulus for which this estimate has approximately the same value as for the reference stimulus.

Experiment 2

Having built an apparatus to run Experiment 1 with real paper and light sources, we had an opportunity to test whether we would have found similar results using computer-generated stimuli. It is increasingly common for vision researchers to use rendering software to create precisely controlled images of complex scenes

(e.g., Morgenstern, Murray, & Harris, 2011), but there have been concerns that even high-quality computer-generated images may not be realistic enough to elicit the same visual processing as real scenes (Morgenstern, Geisler, & Murray, 2014; Radonjić et al., 2016). To investigate this issue, we repeated Experiment 1 using computer-generated renderings of a similar apparatus.

Methods

Participants

There were six observers. All were unaware of the purpose of the experiment and participated for
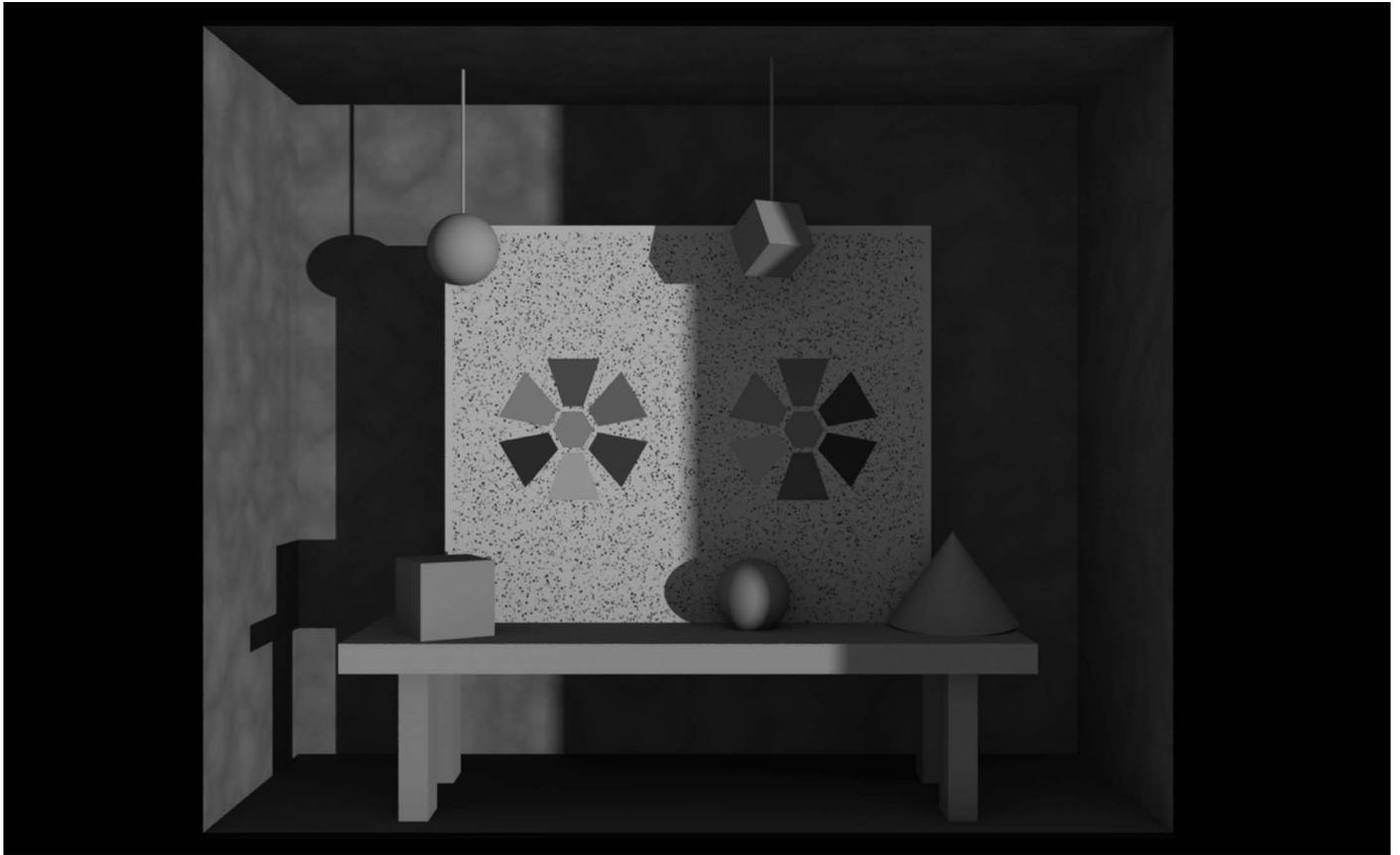
Figure 4. Computer-generated stimulus in Experiment 2. The hexagon in the middle of the darker, right-hand side of the panel is the reference patch, and the hexagon in the middle of the brighter, left-hand side is the match patch.

payment. None had participated in Experiment 1. The observers were 20 to 37 years old, and four were female.

Stimuli

The stimulus was rendered in RADIANCE (Larson & Shakespeare, 1998), and was modeled after the apparatus used in Experiment 1. The stimulus depicted a box containing a frontoparallel panel and some geometric objects on a small table (Figure 4). The panel had the same surface texture as the panel in Experiment 1, and had trapezoids and hexagons with the same relative sizes and in the same arrangement. The panel, trapezoids, and hexagons were the same size in degrees of visual angle as those in Experiment 1, and the simulated reflectances of the panel and trapezoids were also the same. The simulated reflectances of the two hexagons were adjustable, as described under Procedure. We added the geometric objects, which were not present in Experiment 1, in order to give observers a more vivid sense of the spatial arrangement and lighting conditions of the rendered stimulus. Experiment 1 used a paper-and-light stimulus, located on a desk and surrounded by the nearby furniture and walls of the testing room, so the spatial arrangement and

lighting conditions were very clear to observers. We added the geometric objects to Experiment 2 to compensate somewhat for the reduced viewing conditions with rendered stimuli.

We rendered the scene under simulated lighting consisting of a small point-like light source and an ambient source. The point-like source consisted of 20 infinitely distant point light sources, all with the same intensity, distributed over an arc of 0.4°, behind the virtual camera and to the right. This light source illuminated the left half of the display, and cast a shadow on the right half due to an opaque surface that was not visible in the scene. We distributed the point-like source over a small range of directions (0.4°) in order to give the shadow on the panel a clearly visible penumbra, so that it was unlikely to be mistaken for a reflectance boundary. The intensity of the point-like light source was adjustable, as described under Procedure. The ambient light source illuminated the scene uniformly from all directions with a luminance of 135 cd/m^2. That is, the ambient light source functioned as an extended sky, with a visible luminance of 135 cd/m^2 in all directions.

We showed the stimuli on an LCD monitor in a dark room at a viewing distance of 45 cm. The background

luminance surrounding the stimulus on the monitor was 0.08 cd/m². The monitor had a resolution of 2,560 × 1,440 pixels, a pixel size of 0.248 mm, and a nominal refresh rate of 60 Hz. We characterized the monitor's gamma function, using a Minolta LS-110 photometer to measure the relationship between grayscale RGB value and luminance, and the stimulus display software inverted a fitted gamma function to display the required luminances (Brainard, Pelli, & Robson, 2002, p. 179).

Procedure

Each observer ran in one 20-min session that consisted of 120 trials. At the start of each trial, the simulated reflectance of the reference hexagon (Figure 4, right-hand side) was randomly set to one of the three reference reflectances used in Experiment 1 (0.20, 0.31, or 0.49), with the constraint that the same reflectance was not shown on consecutive trials. The simulated illuminance at the reference hexagon was 136 lux, meaning that a surface with a reflectance of 1.0 would have a luminance of 43 cd/m². Also at the start of each trial, the reflectance of the match hexagon (Figure 4, left-hand side) was set to a random value in the range 0.06 to 0.80, and the intensity of the point-like light source was set so that the illuminance at the match hexagon was a random value in the range 275 to 1,102 lux. The observer's task was to adjust the reflectance and illuminance at the match hexagon so that the reflectance appeared to be the same as at the reference hexagon, but the illuminance appeared to be different. Moving a mouse left or right changed the reference hexagon's reflectance in the range 0.06 to 0.80, and moving it up or down changed the illuminance in the range 275 to 1,102 lux. Because of the monitor's limited luminance range, the simulated illuminances were lower in this experiment than the real illuminances in Experiment 1, but they were proportional: Here the simulated illuminance was the same for each reference stimulus, and the lower and upper limits of the adjustable illuminance range for the match stimulus were 2 and 8 times the reference stimulus illuminance, as in Experiment 1. On a few practice trials before the main experiment, we trained the observer to adjust both reflectance and illuminance on every trial. The observer had unlimited time to respond, and clicked the mouse button to indicate that they had finished the trial. A short beep acknowledged the observer's response, and then the next trial began. The median duration of the free-adjustment phase of the trials was 6 s, across all observers and trials. We discarded all trials where observers did not adjust both reflectance and illuminance. On average, observers adjusted both dimensions on 85% of the trials.

Results and discussion

Figure 5 shows the reflectance and illuminance of observers' match settings, as well as lines of best fit and Thouless ratios. The results are broadly similar to those of Experiment 1. Thouless ratios ranged from 0.69 to 1.26, and bootstrapped 95% confidence intervals ranged from ±0.02 to ±0.11.

However, the mean Thouless ratio was significantly higher in Experiment 2 than in Experiment 1 (M_1 = 0.82, SD_1 = 0.08; M_2 = 0.94, SD_2 = 0.13; independent samples t test, $t[34]$ = 3.12, $p < 0.01$). This is surprising, because previous studies have typically found weaker lightness constancy with computer-generated stimuli than with real stimuli (e.g., Morgenstern et al., 2014). Observers may sometimes have weaker constancy with computer-generated stimuli because such stimuli are not usually completely realistic, and so observers may tend to match image luminance rather than the simulated reflectance of objects and surfaces depicted in rendered images. In the General discussion we speculate as to why observers in Experiment 2 showed the opposite pattern of results—that is, seemingly better lightness constancy with a simulated experimental apparatus than with a real apparatus.

General discussion

Logvinenko and Maloney (2006) advanced two hypotheses about perception of achromatic surfaces. The first was that achromatic surfaces have at least two perceptual dimensions: lightness and surface brightness. This hypothesis is based on a multidimensional scaling analysis of observers' dissimilarity ratings of achromatic surface patches, which resulted in a perceptual space with two dimensions roughly corresponding to reflectance and illuminance. Thus Logvinenko and Maloney defined lightness and brightness differently than other authors, namely as dimensions that emerge from a multidimensional scaling analysis. For most other authors, lightness is perceived reflectance, and brightness is perceived luminance. To emphasize this difference, Logvinenko, Petrini, and Maloney (2008) referred to the dimensions that emerge from multidimensional scaling as *surface lightness* and *surface brightness*.

To provide some context, it is worth noting that later studies using multidimensional scaling have found different perceptual spaces for achromatic surfaces. Logvinenko, Petrini, and Maloney (2008) arrived at a one-dimensional (1-D) space for percepts of surface patches in the snake illusion (Adelson, 2000), possibly because the stimuli in those experiments did not contain physical lighting boundaries. Madigan and
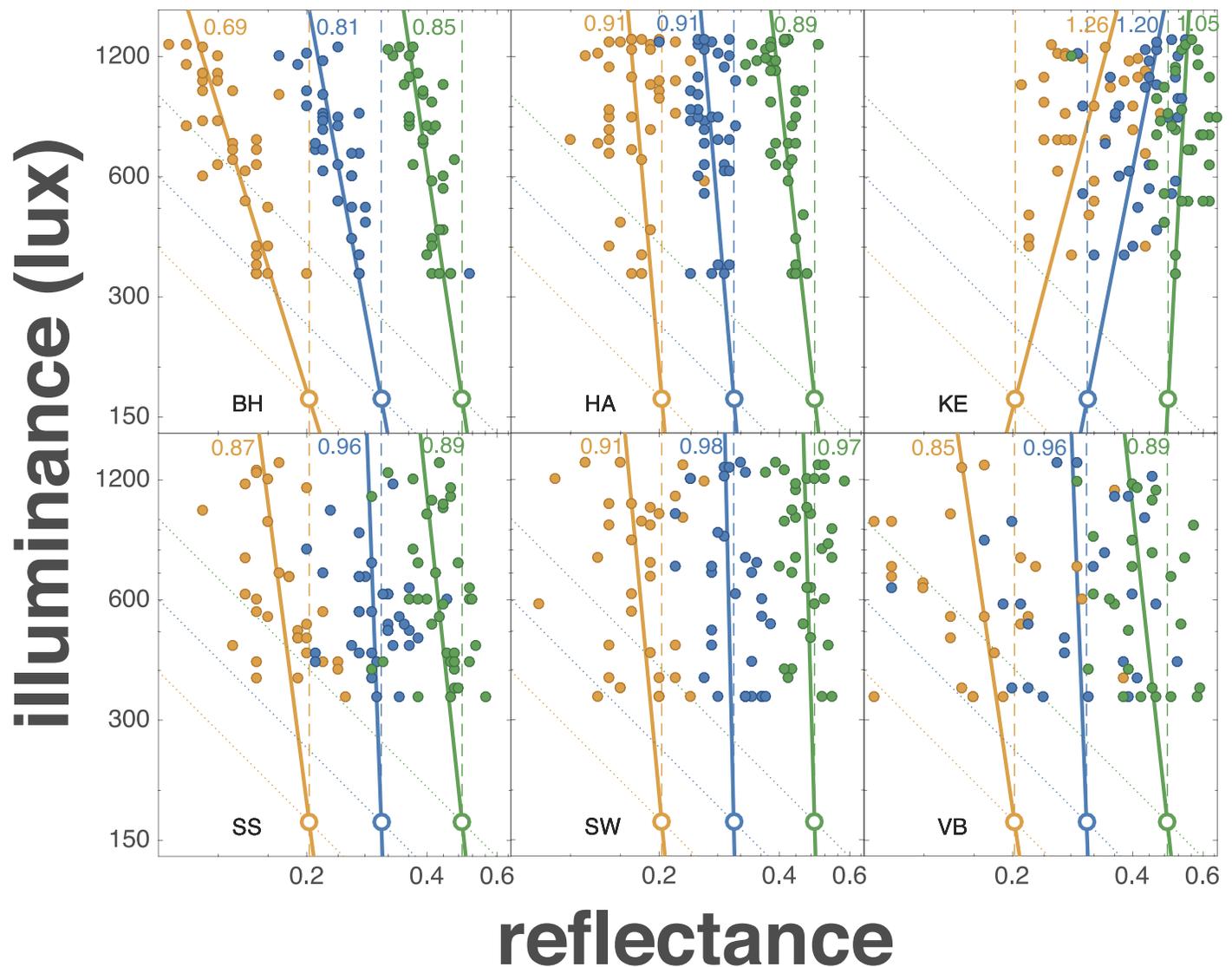
Figure 5. Results of Experiment 2. See caption of Figure 3 for details.

Brainard (2014) also arrived at a 1-D space, perhaps because they instructed one group of observers to judge the similarity of reference patch reflectance, and another group to judge the similarity of reference patch luminance (i.e., reflected light intensity), but none judged overall similarity as Logvinenko and Maloney's observers did. We also note that observers in the second group judged the *luminance* of two reference patches, whereas in Logvinenko and Maloney's account the surface brightness dimension corresponds most closely to *illuminance* (i.e., incident light intensity); observers may be more likely to conflate reflectance and luminance than reflectance and illuminance, which may account for the 1-D perceptual space found in Madigan and Brainard's experiment. Finally, Logvinenko (2015) found that in achromatic scenes that contained both cast shadows and attached shadows,

perceptual spaces were three-dimensional. In summary, there has been progress in understanding percepts of achromatic surfaces, but the question of what dimensions such percepts have, and under what conditions, is far from settled.

Logvinenko and Maloney's second hypothesis is that the perceptual dimensions of lightness and surface brightness are not separately accessible to observers, and that observers match lightness by minimizing a dissimilarity metric that depends on both reflectance and illuminance. In the experiments reported here we find that observers can match reflectance reasonably well in a task where both reflectance and illuminance are freely adjustable. The second hypothesis is not easily reconciled with this finding. However, the alternative view—that observers form a separate estimate of surface reflectance, and that their perfor-

mance in lightness matching tasks is based on this estimate—fits well with the behavior we observed in these experiments.

Logvinenko and Maloney noted that in lightness matching tasks, observers sometimes report that no available match stimulus, over the whole range of physically possible reflectances, is a good match for the reference stimulus. Their second hypothesis gave an explanation for this phenomenon, namely that the perceptual dissimilarity between reference and match stimuli cannot be reduced to zero when the two stimuli are seen under different lighting conditions (Figure 1). If we do not accept the second hypothesis, how can we understand the difficulty that observers sometimes have when making lightness matches?

We suggest that the explanation lies in the first hypothesis, that achromatic surfaces have at least two perceptual dimensions. If a reference surface has two or more perceptual dimensions (e.g., lightness, gloss, translucency, surface brightness), then even if a match surface can be equated to the reference surface on one of these dimensions by adjusting a 1-D physical property such as luminance, it may not be matched on the other dimensions. Thus, even when the targeted perceptual dimension (e.g., lightness) is the same for reference and match surfaces, observers may be aware that the overall appearance of the two surfaces remains quite different.

A vivid example of this is the haze illusion, shown in Figure 6. The plus-shaped Regions A and B are physically identical, but they have very different appearances. Region A appears to be a clear aperture, whereas B appears to depict a partly transparent material. Unsurprisingly, we find that increasing or decreasing the luminance of Region A does little to make the regions appear more similar, because at no luminance does Region A have the cues to transparency that determine the appearance of Region B (e.g., the pattern of luminances at X-junctions; Metelli, 1970). We can judge whether Region A appears brighter or darker than Region B, but even when the two regions look about equally bright (as in Figure 6), their total appearance remains very different. In a similar way, we suggest that when observers in a lightness matching task adjust the match stimulus so that it has the same lightness as the reference stimulus, the two stimuli may still differ on other perceptual dimensions, such as surface brightness, with the result that there is an imperfect match in overall appearance. As mentioned earlier, we did not find this to be a problem in our experiments, and whether it occurs presumably depends on the perceptual dimensions of the stimuli being used.

Our second main finding is that observers had good lightness constancy when viewing computer-generated scenes in Experiment 2. The existing literature suggests
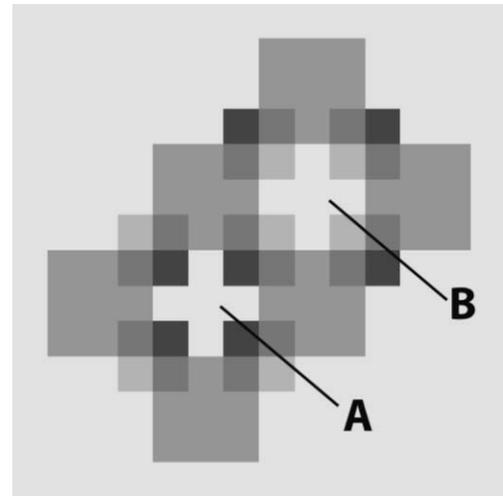

Figure 6. The haze illusion (after Adelson, 2000). The two plus-shaped regions, A and B, are physically identical. Region A appears to be an empty aperture through which the background is visible, whereas Region B appears to be a partly transparent material.

that we should be cautiously optimistic about using rendered scenes in perceptual experiments. Morgenstern et al. (2014) compared the results of previous experiments on lightness constancy, some of which used computer renderings of naturalistic scenes and some of which used physical objects as stimuli. They found that observers showed qualitatively similar lightness matching behavior in the two types of experiments, although constancy was weaker with computer-generated stimuli. In experiments on the closely related topic of illumination discrimination, Radonjić et al. (2016) found that observers showed nearly identical performance with computer-rendered stimuli and real scenes.

However, our findings are surprising in that observers showed significantly *higher* Thouless ratios, and so significantly *better* lightness constancy, with computer-generated scenes than with real scenes. One observer (KE) in Experiment 2 even had Thouless ratios greater than one, which suggests that the high Thouless ratios we found for most observers in Experiment 2 may not have been simply due to the computer-generated stimuli being perceived as very realistic. These results may be related to Agostini and Bruno's (1996) finding that contrast effects are larger in small, bright images in dark environments, such as the computer-generated stimuli in Experiment 2. Also related is McNamara, Chalmers, Troscianko, and Gilchrist's (2000) finding that observers may or may not perceive the intended reflectance in computer-generated images of surfaces, depending on seemingly minor adjustments in the rendering algorithm, such as the number of simulated interreflection bounces.

We speculate that the large, white, textured panel in the rendered images (Figure 4), which was the brightest object in the scene and in the otherwise unlit testing room, may have been perceived to be self-luminous—that is, to be emitting light instead of only reflecting light from external sources. Bonato and Gilchrist (1999) found that when the luminance of a surface that observers perceived to be glowing was increased, the strength of perceived glow increased, and the perceived reflectance of other, nonglowing surfaces in the scene *decreased*. (See Murray [2013] for a Bayesian explanation of why this might be a rational response for the visual system to have.) If observers had perfect lightness constancy, then as the illuminance on the left-hand side of the panel increased, they would have held the match reflectance constant. Instead, Figure 5 shows that as the illuminance on the left increased, observer KE *increased* the reflectance of the match hexagon in order to maintain a constant perceived lightness (i.e., to match the perceived reflectance of the reference hexagon). This is precisely what Bonato and Gilchrist's findings would predict, if the panel appeared self-luminous. Furthermore, we do not believe that observer KE was completely anomalous, because in pilot sessions where the simulated reflectance of the panel was 1.00 (in Experiment 2 it was 0.78), we found that it was relatively common for observers to have Thouless ratios greater than one.

Computer monitors are light-emitting surfaces, and in many experimental settings that use computer-generated images as stimuli, this is perceptually obvious: Even if stimuli are rendered and displayed realistically, as judged by current standards, observers are unlikely to mistake virtual objects for real objects, and they can see that virtual objects are not simply reflective surfaces. This was certainly the case in our Experiment 2: We carefully constructed the stimuli by rendering the apparatus from Experiment 1 using physically accurate rendering software, but nevertheless the stimuli were grayscale images viewed at close range (45 cm), without stereo cues, in a dark room. As is usual in such experiments, it was obvious to observers that they were seeing a representation generated by a light-emitting monitor.

Thus we speculate that the unusually high Thouless ratios we found for most observers in Experiment 2 were due to typical (e.g., average Thouless ratio of 0.82 in Experiment 1) or weaker-than-typical lightness constancy being artificially inflated by the glow effects described by Bonato and Gilchrist (1999). It will take further experiments to test this hypothesis, but we provisionally suggest that the high Thouless ratios in Experiment 2 are not strong evidence for normal lightness perception with computer-generated scenes. Furthermore, these findings suggest a problem that will need to be addressed when evaluating the realism of computer-generated displays, namely the possibility that two failures of lightness perception—weak lightness constancy and glow artifacts—may combine to produce behavior that mimics typical lightness constancy.

Keywords: lightness constancy, lightness matching, asymmetric matching task, illumination, free adjustment task

References

Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, *262*, 2042–2044.

Adelson, E. H. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 339–351). Cambridge, MA: MIT Press.

Agostini, T., & Bruno, N. (1996). Lightness contrast in CRT and paper-and-illuminant displays. *Perception & Psychophysics*, *58*(2), 250–258.

Arend, L. E., & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 1. Illuminance variation. *Perception & Psychophysics*, *54*(4), 446–456.

Blakeslee, B., & McCourt, M. E. (1999). A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction. *Vision Research*, *39*(26), 4361–4377.

Bloj, M., Ripamonti, C., Mitha, K., Hauck, R., Greenwald, S., & Brainard, D. H. (2004). An equivalent illuminant model for the effect of surface slant on perceived lightness. *Journal of Vision*, *4*(9):

6, 735–746, https://doi.org/10.1167/4.9.6. [PubMed] [Article]

Bonato, F., & Gilchrist, A. L. (1999). Perceived area and the luminosity threshold. *Attention, Perception, & Psychophysics*, 61(5), 786–797.

Brainard, D. H., Pelli, D. G., & Robson, T. (2002). Display characterization. In J. Hornak (Ed.), *Encyclopedia of imaging science and technology*. Hoboken, NJ: Wiley and Sons.

Gilchrist, A. L. (2006). *Seeing black and white*. New York, NY: Oxford University Press.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., . . . Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, 106(4), 795–834.

Kraft, J. M., & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences*, 96(1), 307–312.

Larson, G. W., & Shakespeare, R. (1998). Rendering with RADIANCE: The art and science of lighting visualization. San Mateo, CA: Morgan Kaufmann.

Logvinenko, A. D. (2015). The achromatic object-colour manifold is three-dimensional. *Perception*, 44(3), 243–268.

Logvinenko, A. D., & Maloney, L. T. (2006). The proximity structure of achromatic surface colors and the impossibility of asymmetric lightness matching. *Perception & Psychophysics*, 68(1), 76–83.

Logvinenko, A. D., Petrini, K., & Maloney, L. T. (2008). A scaling analysis of the snake lightness illusion. *Perception and Psychophysics*, 70(5), 828–840.

Madigan, S. C., & Brainard, D. H. (2014). Scaling measurements and the effect of surface slant on perceived lightness. *i-Perception*, 5, 53–72.

McNamara, A., Chalmers, A., Troscianko, T., & Gilchrist, I. (2000). Comparing real and synthetic scenes using human judgements of lightness. In B. Péroche & H. Rushmeier (Eds.), *Rendering techniques 2000* (pp. 207–218). Vienna: Springer.

Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. *Ergonomics*, 13(1), 59–66.

Morgenstern, Y., Geisler, W. S., & Murray, R. F. (2014). Human vision is attuned to the diffuseness of natural light. *Journal of Vision*, 14(9):15, 1–18, https://doi.org/10.1167/14.9.15. [PubMed] [Article]

Morgenstern, Y., Murray, R. F., & Harris, L. R. (2011). The human visual system's assumption that light comes from above is weak. *Proceedings of the National Academy of Sciences, USA*, 108(30), 12551–12553.

Murray, R. F. (2013). Human lightness perception is guided by simple assumptions about reflectance and lighting. In B. E. Rogowitz, T. N. Pappas, & H. de Ridder (Eds.), *Proceedings of SPIE 8651, Human Vision and Electronic Imaging XVIII*. San Francisco, CA, February 4–7, 2013.

Radonjić, A., Pearce, B., Aston, S., Krieger, A., Dubin, H., Cottaris, N. P., . . . Hurlbert, A. C. (2016). Illumination discrimination in real and simulated scenes. *Journal of Vision*, 16(11):2, 1–18, https://doi.org/10.1167/16.11.2. [PubMed] [Article]

Rutherford, M. D., & Brainard, D. H. (2002). Lightness constancy: A direct test of the illumination estimation hypothesis. *Psychological Science*, 13(2), 142–149.

Schanda, J. (2007). *Colorimetry: Understanding the CIE system*. Hoboken, NJ: Wiley & Sons.

Shapiro, A., & Lu, Z.-L. (2011). Relative brightness in natural images can be accounted for by removing blurry content. *Psychological Science*, 22(11), 1452–1459.

Thouless, R. H. (1931). Phenomenal regression to the real object. I. *British Journal of Psychology*, 21, 339–359.

Appendix

Here we show that in a lightness-matching task, the set of match points corresponding to a fixed Thouless ratio τ forms a straight line with slope $(\tau - 1)^{-1}$ on axes where log illuminance is plotted against log reflectance, as in Figures 3 and 5.

In a lightness matching task, the reference patch has reflectance r_R, illuminance i_R, and if the patch is Lambertian, luminance $l_R = r_R i_R / \pi$. Similarly, the observer's match setting has reflectance r_M, illuminance i_M, and luminance $l_M = r_M i_M / \pi$. An observer who has perfect lightness constancy makes match setting $r_M = r_R$. An observer who has no lightness constancy, and simply matches image luminance, makes match setting r_0, given by $r_0 i_M / \pi = r_R i_R / \pi$, or $r_0 = r_R i_R / i_M$. The Thouless ratio is given by Equation 1 in the main text, and using the above definitions is

equal to

$$\tau = \frac{\log(r_M) - \log(r_R i_R / i_M)}{\log(r_R) - \log(r_R i_R / i_M)}$$

Solving for $\log i_M$ as a function of $\log r_M$,

$$\log i_M = (\tau - 1)^{-1}(\log r_M - \log r_R) + \log i_R$$

Thus for a fixed Thouless ratio τ, the set of all match points $(\log i_M, \log r_M)$ forms a line through $(\log r_R, \log i_R)$ with slope $(\tau - 1)^{-1}$.