An equivalent illuminant analysis of lightness constancy with physical objects and in virtual reality

Khushbu Y. Patel[1,*], Laurie M. Wilcox[1], Laurence T. Maloney[2], Krista A. Ehinger[3], Jaykishan Y. Patel[1], and Richard F. Murray[1]

[1]Department of Psychology and Centre for Vision Research, York University
[2]Department of Psychology, New York University
[3]School of Computing and Information Systems, The University of Melbourne

* Corresponding author: khushbuypatel1234@gmail.com

Journal-formatted paper available at
https://doi.org/10.3758/s13428-025-02688-1

Abstract

Several previous studies have found significant differences between visual perception in real and virtual environments. Given the increasing use of virtual reality (VR) in performance-critical applications such as medical training and vision research, it is important to understand these differences. Here we compared lightness constancy in physical and VR environments using a task where viewers matched the reflectance of a frontoparallel match patch to the reflectance of a reference patch at a range of 3D orientations relative to a light source. We used a custom-built physical apparatus and four VR conditions: (1) All-Cue (replicating the physical apparatus), (2) Reduced-Depth (no disparity or parallax), (3) Shadowless (no cast shadows), and (4) Reduced-Context (no surrounding objects). Lightness constancy was markedly better in the physical condition than in all four VR conditions. Surprisingly, viewers achieved a degree of lightness constancy even in the Reduced-Context condition, despite the absence of lighting cues. In a follow-up experiment, we re-tested the All-Cue and Reduced-Context conditions in VR with new observers, each participating in only one condition. Here we found lower levels of constancy than in the first experiment, suggesting that experience across multiple experimental settings and possibly exposure to the physical apparatus during instructions had enhanced performance. We conclude that even when robust lighting and shape cues are available, lightness constancy is substantially better in real environments than in virtual environments. We consider possible explanations for this finding, such as the imperfect models of materials and lighting that are used for rendering in real-time VR.

Introduction

In recent years, virtual reality (VR) has become an increasingly popular medium for many applications, such as gaming (Linowes, 2020) and cinema (Marantz, 2016), and it has also found use as a tool for studying human visual perception (Hibbard, 2023; Scarfe & Glennerster, 2015, 2019). Computer-generated images provide flexible control of stimulus properties such as luminance, contrast, and texture, which can be difficult to manipulate precisely in the real world. For example, features such as shadow position (H. Adams, Stefanucci, Creem-Regehr, & Bodenheimer, 2022; Rensink & Cavanagh, 2004), cast shadows (H. Adams et al., 2021), image contrast (Boyaci, Maloney, & Hersh, 2003), and shading of 3D objects (Ripamonti et al., 2004) can be controlled in rendered scenes, facilitating experiments that advance our understanding of visual perception. Furthermore, VR offers a more immersive and potentially realistic 3D environment than traditional flat-panel displays, making it well-suited to studies that involve spatial navigation, lighting perception, and other visual processes influenced by the environment.

Realism is an important goal in many applications of VR, meaning that we would like viewers to perceive a virtual scene just as they would perceive the corresponding real scene. Commercial VR technology currently has significant limits on stimulus properties such as image resolution, field of view, and the dynamic range of luminance and chromaticity. It also depends on image rendering algorithms that make many approximations, since the most physically accurate methods are not computable in real time (Pharr, Jakob, & Humphreys, 2023). As a result, viewers rarely mistake virtual environments for real environments, and performance in virtual environments often falls short of performance in real environments. Realism remains an aspiration at this point, and VR environments must be carefully evaluated for specific applications.

One important factor in realism is color constancy, which is the ability of a viewer to correctly perceive the intrinsic color of surfaces despite variations in illumination and context. Lightness constancy is a special case of color constancy, that addresses the perception of achromatic surface colors (i.e., black, white, and shades of grey). Lightness perception raises many of the same theoretical and empirical problems as color perception more generally. For example, ambiguity is a key obstacle to the accurate estimation of achromatic surface color: the luminance of a surface depends on both the intensity of illumination at a surface (*illuminance*) and the proportion of light reflected by the surface (*reflectance*)[1]. In complex, realistic environments, viewers are able to overcome this ambiguity and estimate surface reflectance to some extent.

In a recent study, we compared lightness constancy in a real scene and in a carefully constructed VR replica, and found that constancy was about equally good in the two cases

[1] For more precise definitions of these photometric terms, see McCluney (1994).

(Patel, Wilcox, Maloney, Ehinger, Patel, et al., 2024). This study used a simple task where participants compared the reflectance of two test patches at the same 3D orientation, on a single planar background surface, separated by a single shadow boundary. In contrast, Morgenstern, Geisler, and Murray (2014) compared previous studies that examined performance in real and virtual environments, using more demanding tasks that required viewers to discount illumination across a range of 3D surface orientations. Their findings suggested that in such tasks, lightness constancy may be notably worse in computer-generated scenes than in real scenes. An important caveat concerning this comparison, though, was that the previous studies they compared were designed independently, and the real and virtual environments differed in many potentially important details.

The goal of the present study is to evaluate lightness constancy in virtual scenes, using a task that requires viewers to match reflectance across 3D orientations and across objects. We compare performance in a virtual scene to performance in a carefully matched physical scene, and we also evaluate the extent to which shadows and depth cues contribute to lightness constancy in VR.

Lightness perception in real and virtual environments

Psychophysical work on VR has given good reason to be cautious about assuming that people see virtual scenes the same way they see real scenes. Previous studies have found discrepancies between real and virtual environments in such diverse and fundamental tasks as navigation (Kimura et al., 2017), size perception (Rzepka et al., 2023), depth judgement (Hartle & Wilcox, 2022), and estimating egocentric distance (Creem-Regehr, Stefanucci, & Bodenheimer, 2022).

Lightness and color perception are computationally difficult tasks that require viewers to disentangle the effects of lighting and surface reflectance. Estimating reflectance requires taking into account factors such as surface orientation, material properties, and the lighting distribution in a scene (Boyaci, Doerschner, & Maloney, 2006; Doerschner, Boyaci, & Maloney, 2004; Flock & Freedberg, 1970; Kraft & Brainard, 1999). To estimate lighting conditions, viewers draw on a range of cues such as cast shadows, surface shading, and specular highlights (Boyaci et al., 2006). Thus we might expect that differences between stimuli in real and virtual environments lead to differences in lightness and color constancy.

Some previous work has shown that people can achieve realistic levels of constancy in virtual environments. Blakeslee, Reetz, and McCourt (2008) found similar levels of lightness constancy using simple 2D stimuli in VR and in a physical apparatus constructed with paper and lights. Patel, Wilcox, Maloney, Ehinger, Patel, et al. (2024), also using a simple 2D stimulus, found that lightness constancy was about equally good in VR and in a carefully matched physical environment, although individual differences were more pronounced in VR. Gil Rodríguez et al. (2022, 2024) found realistic levels of color constancy in a VR environment, though without a comparison to a matched physical scene. Lighting perception is an important part of lightness constancy, and previous studies have found similarities between perception of lighting in real and virtual environments. For example, Radonjić et al. (2016) found that lighting discrimination thresholds were practically identical in a stereoscope and in a physical apparatus.

Other studies have found discrepancies between lightness in real and virtual environments. Patel, Munasinghe, and Murray (2018) reported a paradoxical effect where Thouless ratios, a measure of constancy, were significantly *higher* with computer-rendered stimuli on a flat-panel monitor than with a physical apparatus illuminated by real lights; they speculated that this may have been due to the computer-generated stimuli appearing to glow when displayed on a monitor in a dark room.

As noted above, Morgenstern et al. (2014; 2015) also found large discrepancies between real and virtual environments, based on a re-analysis of previous studies. Boyaci et al. (2003) used computer-generated scenes in a stereoscope to examine whether viewers compensate for the directional distribution of lighting when estimating the reflectance of surfaces at various 3D orientations. Bloj et al. (2004) independently carried out a similar study, but used a physical apparatus with real objects and lights. Both studies modelled their findings using variants of the equivalent illumination model (EIM) that we describe below, and showed that viewers do take into account the 3D distribution of lighting when estimating reflectance. Lightness constancy was not perfect in either study, and parameters of the EIM allowed the authors to quantify the degree of constancy. Morgenstern et al. (2014) re-analyzed these findings using a single form of the EIM to make them directly comparable, and showed that lightness constancy was dramatically better in Bloj et al.'s task (which used physical stimuli) than in Boyaci et al.'s (which used a stereoscope).

Díaz-Barrancas, Cwierz, Pardo, Pérez, and Suero (2020) and Díaz-Barrancas, Cwierz, and Pardo (2021) described hyperspectral imaging techniques for VR, which is one approach to improving realism and may improve surface color perception in virtual environments.

Equivalent illumination models

Equivalent illumination models (EIMs) are a class of mathematical models of surface color perception (Brainard, 1998; Brainard & Maloney, 2011; Murray, 2021). These models assume that perception is a process of inferring surface properties (such as reflectance) from image properties (such as luminance), guided by a physically realistic model of the interaction between light and surfaces. They are therefore normative models that take the approach of 'inverse optics' (Pizlo, 2001).

Boyaci et al. (2003) and Bloj et al. (2004) developed EIMs for perception of lightness of 3D surfaces, which are equivalent to the following model. Suppose that achromatic lighting consists of a directional light source with maximum illuminance D, in a direction given by unit vector $\mathbf{s}$, and also an ambient light with illuminance A in all directions. An achromatic Lambertian surface with reflectance ρ and surface normal $\mathbf{n}$ produces the following image luminance:

$$L = \frac{\rho}{\pi} \left[D \max(\mathbf{n} \cdot \mathbf{s}, 0) + A \right] \tag{1}$$

That is, the surface receives illuminance D from the directional source, scaled by $\mathbf{n} \cdot \mathbf{s} = \cos\theta$, where θ is the angle between the surface normal $\mathbf{n}$ and lighting direction $\mathbf{s}$. It also receives illuminance A from the ambient source, regardless of orientation. The image luminance L is then the total illuminance scaled by the surface reflectance ρ, and also scaled by a factor π that follows from the definitions of SI units for luminance and illuminance (McCluney, 1994).

A viewer who knows the lighting conditions and surface orientation can therefore potentially infer surface reflectance from image luminance:

$$\rho = \frac{\pi L}{D \max(\mathbf{n} \cdot \mathbf{s}, 0) + A} \tag{2}$$

This equation is the foundation of the EIM considered here. The model assumes that viewers estimate lighting conditions and surface orientation, and infer reflectance following this equation. Precisely how viewers estimate lighting and orientation is beyond the scope of the model, but nevertheless, as we will see, the model assumes that errors in lighting and orientation perception play a key role in accounting for failures of lightness and color constancy.

This model has a clear physical interpretation. The numerator of equation (2) is the luminous exitance (i.e., luminous flux *emitted* per square meter) of a Lambertian surface with luminance L. The denominator is the illuminance (i.e., luminous flux *received* per square meter) of the assumed lighting. Thus the model simply estimates reflectance as the ratio of reflected to incident light.

Human lightness and color constancy are rarely perfect, and one goal of modelling is to account for variations in constancy across variations in stimulus properties such as lighting conditions and surface orientation. Boyaci et al. (2003), Bloj et al. (2004), and subsequent studies found that such variations are often described well by the EIM with appropriate parameter settings (Brainard & Maloney, 2011). EIMs have been successful at accounting for lightness and color constancy in a range of moderately complex scenes, both physical and computer-generated, with a variety of lighting conditions and material types (Bloj, Kersten, & Hurlbert, 1999; Bloj et al., 2004; Boyaci, Doerschner, & Maloney, 2004; Boyaci et al., 2006, 2003; Doerschner, Boyaci, & Maloney, 2007). Interestingly, these studies have advanced the claim that lightness constancy fails largely because viewers tend to strongly over-estimate the ambient component of lighting, and to a lesser extent, because their orientation estimates are biased towards the frontoparallel plane. (This frontoparallel bias is well-known from previous work, e.g., Saunders and Chen (2015).) Both of these tendencies imply that when two surfaces with the same reflectance have different orientations relative to a light source, viewers will mistakenly interpret the resulting luminance difference partly as a reflectance difference. Hence viewers will have imperfect lightness constancy.

Present study

Understanding realism in VR is crucial for applications like cinema, gaming, and training, which involve perception of complex scenes, and it is also important for applications in vision research. Here we take a step in this direction by evaluating lightness constancy in carefully matched real and virtual environments, in a task like the one examined by Morgenstern et al., where participants match the lightness of surfaces at a range of 3D orientations in a moderately complex scene. We use an EIM to give an interpretable description of participants' behaviour in this task, and to quantify their level of lightness constancy. Under some conditions, the question of what constitutes 'constancy' in a virtual environment may be difficult, since the ground truth is simulated. In the present case, we will consider viewers to have good constancy in VR if they make lightness judgements in VR that track the true reflectance of objects in the corresponding physical scene.

Shadows can have a strong effect on lightness (Adelson, 1993; Boyaci et al., 2006), as can binocular disparity and motion parallax (Kitazaki, Kobiki, & Maloney, 2008). Here we also examine the effect of these cues on lightness constancy in a virtual environment, in order to evaluate the importance of rendering these stimulus properties accurately for realistic lightness perception in VR.

Thus we evaluate lightness constancy (a) in a physical apparatus, and in a VR headset (b) with a standard rendering of shadows and depth cues, (c) without cast shadows and (d) without binocular disparity or motion parallax. As a limiting case, we also include (e) a VR condition without the array of objects that provided participants with cues to lighting in the other conditions; here we expect participants to simply match luminance instead of reflectance, as they have no cues to lighting conditions. Equation (2) shows that according to the EIM, errors in surface orientation perception can cause errors in lightness perception, so in each condition we also include an orientation matching task to evaluate whether participants perceive orientation accurately. In the VR conditions, we used an Oculus Pro headset driven by an NVIDI GeForce GTX 1060 graphics card, and virtual scenes were generated using Unity 2019.4.

Experiment 1

Methods

Participants. There were 12 participants, none of whom were aware of the purpose of the experiment. Nine were female, three were male, and ages ranged from 19 to 31 years old. Participants gave written informed consent, and were paid for their participation. All reported normal or corrected-to-normal acuity, and reported no known anomalies in color vision. All procedures were approved by the Office of Research Ethics at York University.

Stimuli.

Physical environment. In the Physical condition, the participant sat in a 3.00 m × 1.75 m room and viewed the apparatus on a table, from a viewing distance of 74 cm (Figures 1 and 2). The equipment had two components: the reference apparatus and the match apparatus.

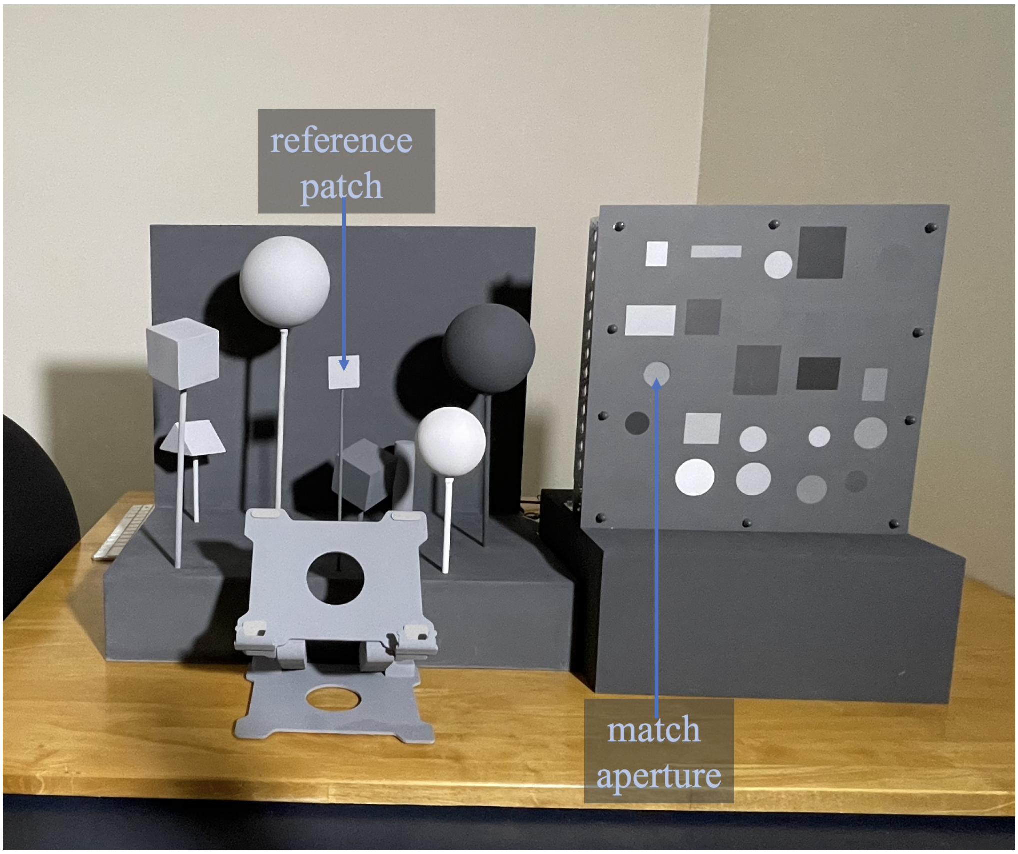

Figure 1. The physical apparatus seen from participants' viewing position.

The reference apparatus was located on the left side of the table (Figure 1). It consisted of a reference patch and several geometric objects attached to a wooden frame. The reference patch was a 3 × 3 cm grey metal square, magnetically attached to the top of a vertical metal rod. The reference patch was removable, and a different reference patch with a randomly chosen reflectance was attached to the vertical rod by the experimenter on each trial (details under Procedure). The vertical rod was attached to a computer-controlled servo motor, which rotated the rod and reference patch to a range of azimuth angles. The reference patch was upright at all azimuths, i.e., its surface normal was parallel to the ground plane. The reference apparatus also included several geometric objects, supported at a range of heights on wooden rods, each painted a shade of uniform grey, with reflectances ranging across objects from 0.11 to 0.82. The geometric objects provided cues that the participant could potentially use to estimate local lighting conditions.

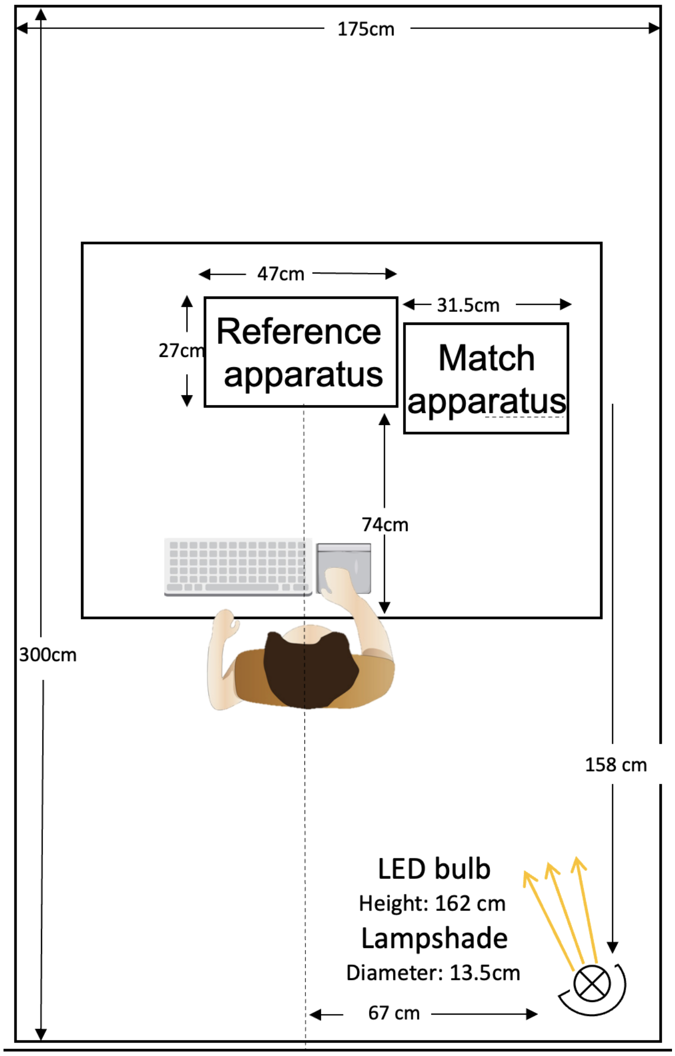

Figure 2. A schematic top-down view of the physical room and apparatus. The match apparatus was placed slightly in front of the reference apparatus, so that the reference and match patches were equally distant from the participant.

We painted all components of the reference apparatus with mixtures of three shades of matte grey paint (Rust-Oleum Chalked Ultra Matte Paint; shades Charcoal, Aged Grey, and Linen White; Rust-Oleum, San Diego, CA, U.S.A.). We painted the square reference patches using an airbrush to produce a smooth and untextured finish, and the rest of the apparatus with spray tools and paintbrushes. We used mixtures of the paints to produce a range of reflectances. Measurements of luminance as a function of slant relative to a collimated light source showed that the painted reference patches were closely approximated by a Lambertian shading model (see Supporting Information, Figure S1).

The match apparatus was located on the right side of the table (Figure 1). The front of the apparatus was an achromatic paper panel, 31.5 cm square. The panel showed a printed image with background reflectance 0.16 and randomly placed circles and rectangles with reflectances ranging from 0.07 to 0.71. On the left side of the apparatus was a 2.5 cm diameter circular aperture that we refer to as the ‘match patch’. Immediately behind the aperture, flat against the back of the panel, was a disk of diameter 25 cm. The front of the disk was a paper annulus whose reflectance ranged continuously from 0.06 to 0.80. A computer-controlled servo motor rotated the disk to adjust the part of the annulus (and hence the reflectance) seen through the aperture. Only a small area of the annulus was visible through the aperture at any time, so the reflectance gradient was not perceptible. The same poster printer and type of paper were used to create the printout on the front of the panel and the annulus with graded reflectance. The square panel sat on a 14 cm high, matte grey wooden base.

We measured the reflectance of components of both apparatuses using a photometer (model LS-110; Konika Minolta, Tokyo, Japan) and a 99% diffuse reflectance standard (Spectralon SRS-99-020; Labsphere, North Sutton, NH). We measured the luminance L of a target surface location, and the luminance L_{99} of the reflectance standard placed at the same location. Both measurements were made under the illuminant used in the experiment. We calculated the reflectance of the target location as $r = 0.99\,L/L_{99}$.

The room and apparatus were illuminated by an LED bulb (Philips A-Line Energy Saving Light Bulb, 3000 lumen, 5000K; Signify N.V., Eindhoven, Netherlands). The bulb was positioned above, behind, and to the right of the participant (Figure 2). Relative to the reference patch, the bulb had an azimuth of approximately 22° to the right, and an elevation of approximately 22°.

Lighting conditions are be said to be *directional* if incident light comes mostly from a single direction, and *diffuse* if light comes from a wide range of directions (Murray & Adams, 2019; Pont & Koenderink, 2007). We adjusted the lighting diffuseness at the reference and match apparatus by changing the distance between the LED bulb and the apparatus (longer distances made the light more diffuse) and adjusting the length of a black paper cylinder that we placed around the LED bulb in spotlight-like fashion (shorter cylinders made the light more diffuse). Our goal was to use a naturalistic level of diffuseness in the experiment.

Morgenstern et al. (2014) measured lighting in a wide range of natural scenes, and quantified diffuseness using a novel measure that they called 'illuminance contrast energy' (ICE). The average ICE value in the scenes they examined was 0.55. To illustrate the meaning of this value: if a scene is illuminated by a distant point light source and a uniform ambient light source, an ICE value of 0.55 means that the incident light from the point source on a surface directly facing it is three times as intense as the light on a surface at any orientation from the ambient source (see Morgenstern et al., equation (3)).

We estimated lighting diffuseness at the apparatus with the following procedure. We used an imaging photometer (Lumetrix 400A; Westboro Photonics, Ottawa, Canada) to take a luminance-calibrated photograph of the large white sphere located above and to the left of the reference patch (Figure 1). We fitted a model to this image, where a numerical optimization routine assumed that the lighting consisted of a directional source and a uniform ambient source (which was only an approximation), and searched for the direction and intensity of the directional source, and the intensity of the ambient source, that best accounted for the pixel-by-pixel luminance of the image. Using the fitted directional and ambient intensity parameters, Morgenstern et al.'s equation (3) shows how to calculate ICE. We adjusted the lighting distance and LED bulb cylinder length to produce an ICE value of 0.56, which according to Morgenstern et al.'s measurements is a typical value in natural scenes. This setting is not a crucial part of our procedure, but we wished to ensure that the lighting conditions in the apparatus were at least approximately typical of natural scenes.

We used a spectrophotometer (SpectraScan PR-655; JADAK, North Syracuse, NY) to measure the chromaticity of the lighting, which yielded coordinates $x = 0.32$ and $y = 0.33$ in xy color space. The luminance of the reflectance standard, positioned at the same location as the reference patch during the measurement, was $41.5\,\text{cd/m}^2$.

For the orientation matching task (see Procedure for details), we placed a small LCD monitor (7" diagonal display) in front of the reference apparatus, in the grey monitor stand visible in Figure 1. The monitor displayed two lines: a fixed horizontal line at the centre of the display, and an adjustable line that the participant could rotate using a trackpad to indicate the orientation that the reference patch would have if viewed from above. The adjustable line rotated about its midpoint at the centre of the display. Both lines were 7 cm long, and were white on a grey background. This monitor was only placed in the scene during the orientation matching task, and not during the lightness matching task.

VR environment. In the VR conditions, the participant viewed a scene in a Meta Quest Pro headset (106° FOV, 1920 × 1800 pixel resolution per eye), driven by an NVIDIA GeForce GTX 1060 graphics card on a PC running Windows 10.

The virtual scene (Figure 3) was modelled after the room and apparatus in the physical environment, and rendered in Unity using the Built-in Render Pipeline (Unity Technologies, 2020, Version 2019.4.38f1). The virtual reference and match apparatus were carefully matched to the physical apparatus. For the reference apparatus, the simulated size, position, and reflectance of the reference patch, surrounding geometric objects, and backing panels were matched to the physical apparatus. For the match apparatus, the simulated size, position, and reflectance pattern of the front panel were matched to the physical apparatus, as were the simulated size and position of the aperture. The reference and match apparatus were rendered as a Lambertian material (Unity material type Legacy/Diffuse). The rest of the virtual scene was also matched to the physical scene, though not as precisely. The walls and furniture in the virtual room also had approximately the same simulated size, position, and color as their physical counterparts. These surrounding objects were created with Unity's 'Standard' material type, and we made small adjustments to their chromaticity, as well as their 'metallic' and 'smoothness' properties, until the virtual objects appeared similar to the physical objects. We also measured luminance at a coarse grid of locations and adjusted the virtual scene so that the luminances of the rendered RGB values were approximately proportional.

The virtual lighting consisted of a directional light source (Unity light type Directional) at the same angle relative to the reference patch as the LED bulb in the physical environment (azimuth 22°, elevation 22°), and also an ambient light source (the Unity Skybox). We gave the directional and ambient light intensities a 3:1 ratio, which as explained above results in an ICE value of 0.55, matching the diffuseness in the physical apparatus. We used Unity's 'Soft shadows' setting for the directional light source.

Luminances were higher in the VR scenes than in the physical scene. In the VR scene, the luminance of the reference patch, when assigned a reflectance of 0.99 and a frontoparallel orientation, was approximately 89 cd/m^2, whereas, as noted above, the luminance of a similar patch in the physical scene was 41.5 cd/m^2. However, luminances on the reference and match apparatuses in the VR scenes were proportional to the corresponding luminances on the physical apparatuses. We also configured the virtual lighting so that the luminance at several locations in the scene, on the walls surrounding the panel apparatus, was proportional to the luminance in the physical scene, to within approximately 10% error. The chromaticity of achromatic stimuli in VR was x=0.32 and y=0.33.

We calibrated the VR headset to ensure that the luminances displayed were proportional to rendered achromatic RGB values (i.e., the three integers 0-255 that represent each pixel). Murray, Patel, and Wiedenmann (2022) provide details of the calibration procedure we followed.

The experiment used four variations of this VR environment. In the All-Cue condition, the scene was rendered as described above (Figure 3 (a)). In the Reduced-Depth condition, binocular disparity was set to zero and the virtual camera position was fixed (i.e., did not track head movements) in order to eliminate motion parallax. As a result, the displayed scene was always the same, regardless of head movements. In the Shadowless condition, no cast shadows were rendered (Figure 3 (b)). In the Reduced-Context condition, only the reference patch and the front panel of the matching apparatus were visible (Figure 3 (c)). In this last condition, the scene elements were shown against a grey background that matched the luminance of the back panel of the reference apparatus in the other VR conditions, so that the local contrast of the reference patch was the same as in the other conditions.

As in the Physical condition, for the orientation matching task a small monitor was added to the scene, in the stand in front of the reference apparatus, showing a fixed horizontal line and an adjustable line that the participant could rotate to indicate the perceived orientation of the reference patch.

Procedure. Before starting the experiment, the participant made a few practice lightness matches and orientation matches in the Physical condition (see below for details). This familiarized them with the task, and providing instructions while they used the physical apparatus made it easier to convey that the goal was to match the perceived shade of grey surface color (i.e., reflectance), and not some other stimulus property such as illumination or luminance. Performance in matching tasks can be highly dependent on instructions (Blakeslee et al., 2008), so for consistency the experimenter read the

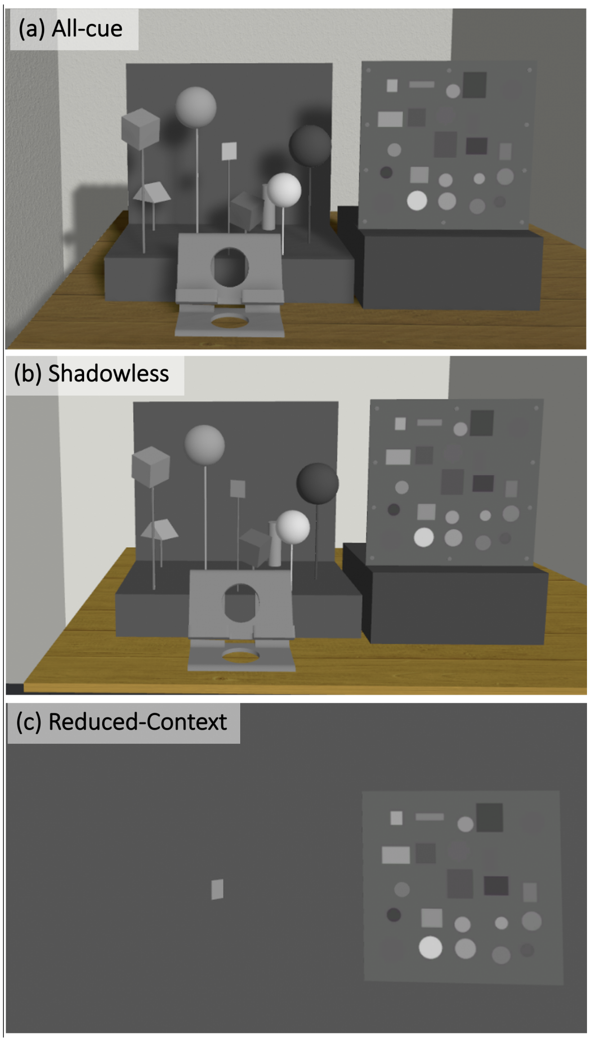

Figure 3. Screen captures of VR scenes from the participant's viewpoint. (a) The All-cue condition, modelled after the physical testing room. The Reduced-Depth condition appeared the same, but was configured for zero disparity and no motion parallax. (b) The Shadowless condition had no shadows. (c) The Reduced-Context condition, which showed only the reference patch and the match apparatus.

instructions from a prepared script. After the instructions and practice trials, the participant ran in the five conditions (the Physical condition and the four VR conditions) in a randomly chosen order. Each participant completed the experiment over two or three days with at least a 20-minute break between conditions.

Physical condition. The participant sat at a table supporting the experimental apparatus, and their head position was stabilized by a chinrest. The reference and match patches in the apparatus described above were approximately at eye level.

In each condition, participants first performed a lightness matching task, followed by an orientation matching task.

At the beginning of each trial of the lightness matching task, the participant pulled a cord which caused a small black curtain to block their view of the apparatus. While the view was blocked, the experimenter placed a new reference patch on the apparatus. The reflectance of the reference patch was either 0.41 or 0.58, chosen according to a computer-generated, pseudo-random ordering of stimuli. There were four practically identical reference patches available for each of the two reference reflectances, and the experimenter used the patches in a computer-generated, pseudo-random order to reduce the probability that the participant would learn to identify individual reference patches from small imperfections. After the experimenter attached the reference patch, the servo motor rotated the patch to one of seven randomly chosen azimuths: -50°, -33°, -17°, 0°, 17°, 33°, or 50°. Here 0° indicates an orientation facing the participant, and positive angles indicate that the surface normal was directed to the right of the participant. The reflectance at the match aperture was initialized to a random value within the range displayable on the apparatus (0.06 to 0.75). The participant then released the cord, and the curtain moved to reveal the apparatus. The participant adjusted the reflectance of the match patch by swiping up or down on a trackpad. The participant was instructed to match the paper visible through the match aperture to the shade of grey paint on the reference patch. The participant had unlimited time to make this setting, and pressed the spacebar on a keyboard to indicate that they had completed the match. A short beep acknowledged their response, and then the next trial began. There were two reflectances and seven azimuths, with each combination repeated ten times, for a total of $2 \times 7 \times 10 = 140$ trials.

At the beginning of each trial of the orientation matching task, the participant pulled a cord and the curtain blocked their view of the apparatus. The reference patch (always the same patch, with reflectance 0.58) was then randomly rotated to one of the seven azimuths used in the lightness task. The participant released the cord and the curtain moved aside to reveal the apparatus. As explained under Stimuli, a small monitor displayed a fixed horizontal line, and also a match line whose orientation the participant could adjust using the trackpad. The participant was instructed to adjust the match line to the orientation the reference patch would have if viewed from above. The participant pressed a spacebar when they had completed the match. The next trial then began. There were seven azimuths, each repeated five times, for a total of $7 \times 5 = 35$ trials.

VR conditions. The four VR conditions (All-Cue, Shadowless, Reduced-Depth, and Reduced-Context) were largely the same as the Physical condition, except that the participant viewed a virtual scene in a VR headset. We did not use a chinrest in this condition, in order to avoid contact between the headset and chinrest. Instead, a green cube was positioned at the intended eye position in the virtual scene. Participants were instructed to adjust their head position until the cube was no longer visible, ensuring a consistent viewpoint. This arrangement encouraged the participant to keep their head at a fixed viewing position, while still allowing small head movements, much as with the chinrest in the Physical condition.

In the lightness matching task, the simulated reflectance and orientation of the reference patch were set to randomly chosen values from the reflectances and orientations used in the Physical condition. The change in stimulus properties was instantaneous, so blocking and unblocking the apparatus with a virtual curtain was not necessary. The participant used two VR hand controllers to adjust the match reflectance to match the reference patch reflectance. The left joystick made fast changes, and the right joystick made slower changes. The participant pressed a button on the right controller to indicate that they had completed the match.

In the orientation matching task, the orientation of the reference patch was set to a randomly chosen value from the orientations used in the lightness task. The participant adjusted the orientation of the match line using the two VR controllers, and pressed a button on the right controller to indicate that they had completed the match.

Analysis. Given a participant's reflectance matches ρ_i for a set of surfaces with luminances L_i and normals $\mathbf{n}_i$, we can fit equation (2) to find the lighting parameters (D, $\mathbf{s}$, A) that are most consistent with the matches. In practice, the EIM has been used in a slightly different manner.

First, the participant's reflectance matches are usually fitted only up to a global scale factor, which we can accomplish by introducing a free multiplicative parameter k into equation (2):

$$\rho_h = \frac{\pi L k}{D_h \max(\mathbf{n}_h \cdot \mathbf{s}_h, 0) + A_h} \tag{3}$$

Here we also add the subscript h to several variables, to indicate that they are not the true physical properties of the scene as in equations (1) and (2), but rather the human participant's possibly biased or noisy estimates of these properties.

Second, variations in surface orientation are often limited to the azimuth of the surface normal $\mathbf{n}$, as in the present experiment. To establish a coordinate system, we assume that the participant is located on the $+x$-axis, looking towards the origin, and the $+z$-axis is upwards. The perceived surface normal can then be given in spherical coordinates with azimuth ψ_n and elevation ϕ_n:

$$\mathbf{n}_h = (\cos\phi_n \cos\psi_n,\ \cos\phi_n \sin\psi_n,\ \sin\phi_n)$$

as can the perceived lighting direction $\mathbf{s}_h$:

$$\mathbf{s}_h = (\cos\phi_s \cos\psi_s,\ \cos\phi_s \sin\psi_s,\ \sin\phi_s)$$

The dot product in equation (3) can be written as

$$\mathbf{n}_h \cdot \mathbf{s}_h = \cos\phi_n \cos\phi_s \cos(\psi_n - \psi_s) + \sin\phi_n \sin\phi_s$$

and equation (3) becomes

$$\rho_h = \frac{\pi L k}{D_h \max(\cos\phi_n \cos\phi_s \cos(\psi_n - \psi_s) + \sin\phi_n \sin\phi_s, 0) + A_h}$$

Following earlier analyses (Bloj et al., 2004; Boyaci et al., 2003), we assume that the participant correctly estimates that $\phi_n = 0$, i.e., the reference patch is upright. Then,

$$\rho_h = \frac{\pi L k}{D_h \max(\cos\phi_s \cos(\psi_n - \psi_s), 0) + A_h} \tag{4}$$

$$= \frac{L}{\alpha(\max(\cos(\psi_n - \psi_s), 0) + \beta)} \tag{5}$$

where

$$\alpha = \frac{D_h \cos\phi_s}{\pi k}$$

$$\beta = \frac{A_h}{D_h \cos\phi_s}$$

Here we use the fact that $\cos\phi_s \geq 0$ for $-90° \leq \phi_s \leq 90°$, to move this factor outside the max function. The main parameters of interest in equation (5) are ψ_s, which is the participant's estimate of the lighting azimuth, and β, which depends on both the participant's estimate of the ambient-to-directional lighting ratio A_h/D_h and their estimate of the lighting elevation ϕ_s.

In the Results section, we report an analysis where we fit equation (5) to each participant's reflectance matches, with α, β, and ψ_s as free parameters, and with the reference patch azimuth ψ_n set to the true stimulus azimuth on each trial. Assuming that the participant perceives the lighting elevation $\phi_s = 22°$ correctly (we address this assumption in the General Discussion), we use the fitted value of β to find the ambient-to-directional lighting ratio that is consistent with the participant's responses:

$$A_h/D_h = \beta \cos\phi_s$$

Morgenstern et al. (2014) define ICE at a given point in space as the coefficient of variation of the illuminance over all incoming directions at that point:

$$\lambda = \left(\frac{1}{4\pi} \int_0^{\pi} \int_0^{2\pi} \left(\frac{E(\psi,\phi) - \bar{E}}{\bar{E}} \right)^2 \sin\psi \, d\phi \, d\psi \right)^{1/2}$$

where $E(\psi, \phi)$ is the illuminance in the direction with spherical coordinates (ψ,ϕ), and $\bar{E}$ is the mean illuminance over all directions. They show that an ambient-to-directional lighting ratio A/D can be converted to ICE as follows:

$$\lambda = \begin{cases} \frac{\sqrt{5/48}}{(A/D)+0.25} & D > 0 \\ 0 & D = 0 \end{cases}$$

Results

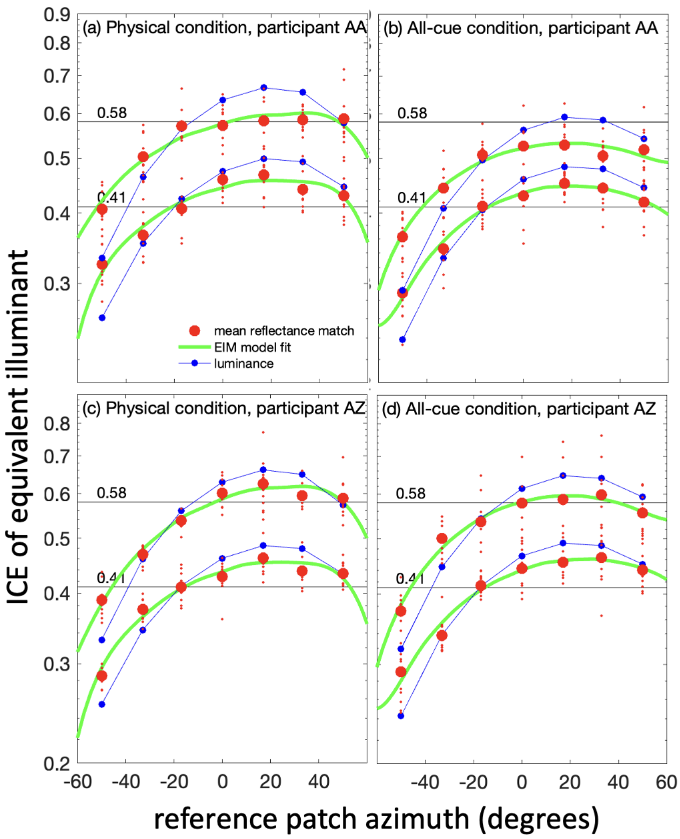

Figure 4. Reflectance matches as a function of the azimuth of the reference patch for two typical participants in the Physical and All-Cue conditions. The small red dots are reflectance matches on individual trials, and the large red dots are means. The green fitted lines show the EIM model fit to the mean reflectance matches. The horizontal black lines indicate the reflectance of the reference patches. The blue curve is the stimulus luminance, scaled to make a least-squares fit to the mean reflectance matches.

Lightness matches. Figure 4 shows reflectance matches as a function of the azimuth of the reference patch, for two typical participants in the Physical and All-Cue conditions. (We provide data and code for these and other analyses as Supporting Information at https://doi.org/10.17605/OSF.IO/3GXRA.) If participants had perfect lightness constancy, their matches would not vary with azimuth, and would fall along a horizontal line. If participants simply matched luminance, their matches would follow the inverted-U pattern of blue data points, which show the stimulus luminance, scaled to match the participant's reflectances as well as possible in the sum-of-squares sense. Instead, participants' reflectance matches fell partway between the two extremes of perfect constancy and luminance matching, as is almost always the case in studies of lightness constancy. The green line shows the sum-of-squares fit of the EIM in equation (5) to participants' reflectance matches.

To exclude participants who had difficulty following instructions or maintaining attention, we removed those who had a mean absolute error of 10° or greater in the orientation matching task, averaged over all five conditions. This resulted in exclusion of two participants from the initial group of 12, leaving a sample of 10 for further analysis.

We used a dependent samples t-test to compare the means of the fitted ICE values for the two reference reflectances. This test indicated a significant difference between the means ($t(49.00) = 2.96$, $p < 0.05$). However, the difference was small (mean 0.25 for reflectance 0.41 and mean 0.20 for reflectance 0.58), so we pooled data from the two reflectances in subsequent analyses.

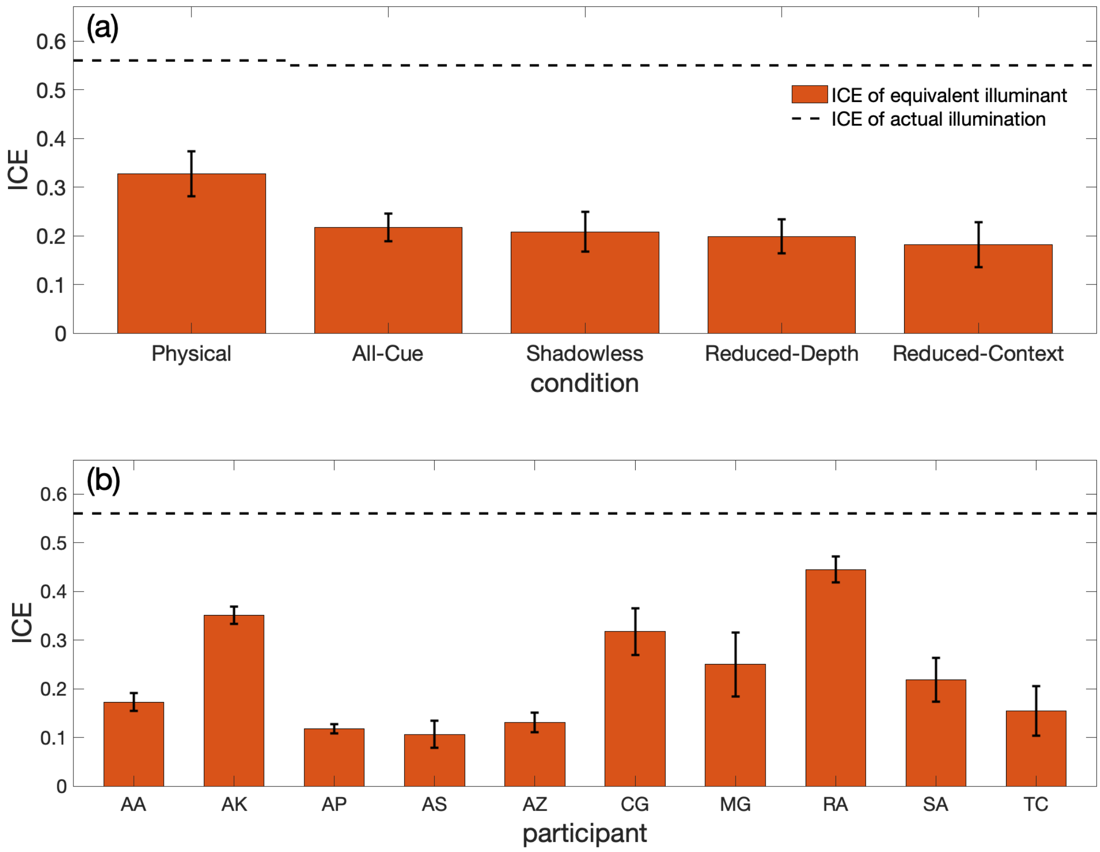

Figure 5. Mean value of the fitted ICE parameter, with error bars showing standard errors. (a) Mean for each condition, averaged over participants. The black dashed line shows the actual ICE values for lighting in the physical (0.56) and VR scenes (0.55). (b) Mean for each participant, averaged over conditions. The black dashed line shows ICE=0.55, the lower of the lighting diffuseness values in the physical and VR conditions.

Figure 5 shows mean ICE values, averaged by condition (top panel) and by participant (bottom panel). The dashed horizontal lines show the true ICE values of the illuminants (0.56 in the Physical condition and 0.55 in the VR conditions). In all conditions, ICE was significantly less than 0.55, as determined by individual one-sample t-tests, with all p values less than the Bonferroni-corrected alpha level of 0.01. Thus participants' reflectance matches were consistent with an over-estimate of lighting diffuseness, as has also been found in previous studies (Brainard & Maloney, 2011;

Morgenstern et al., 2014). Furthermore, ICE was significantly less than 0.55 for all participants, as determined by Bonferroni-corrected one-sample t-tests (all p<0.005), with the exception of three subjects: CG ($p = 0.008$), MG ($p = 0.010$) and RA ($p = 0.017$). In Appendix A we report ICE values for individual participants in all conditions.

A Levene's test for homogeneity of variance of ICE across the five conditions yielded non-significant results ($p = 0.61$), indicating that variance did not vary substantially across conditions. This made it possible to use a repeated-measures ANOVA to compare ICE across conditions. The ANOVA showed a statistically significant effect of condition F(36, 4) = 8.47, $p < 0.05$). Post hoc comparisions with Benjamini/Hochberg corrections indicated that ICE was significantly greater in the Physical condition than in all four VR conditions ($p < 0.05$), and that there were no significant differences across the VR conditions ($p > 0.05$).

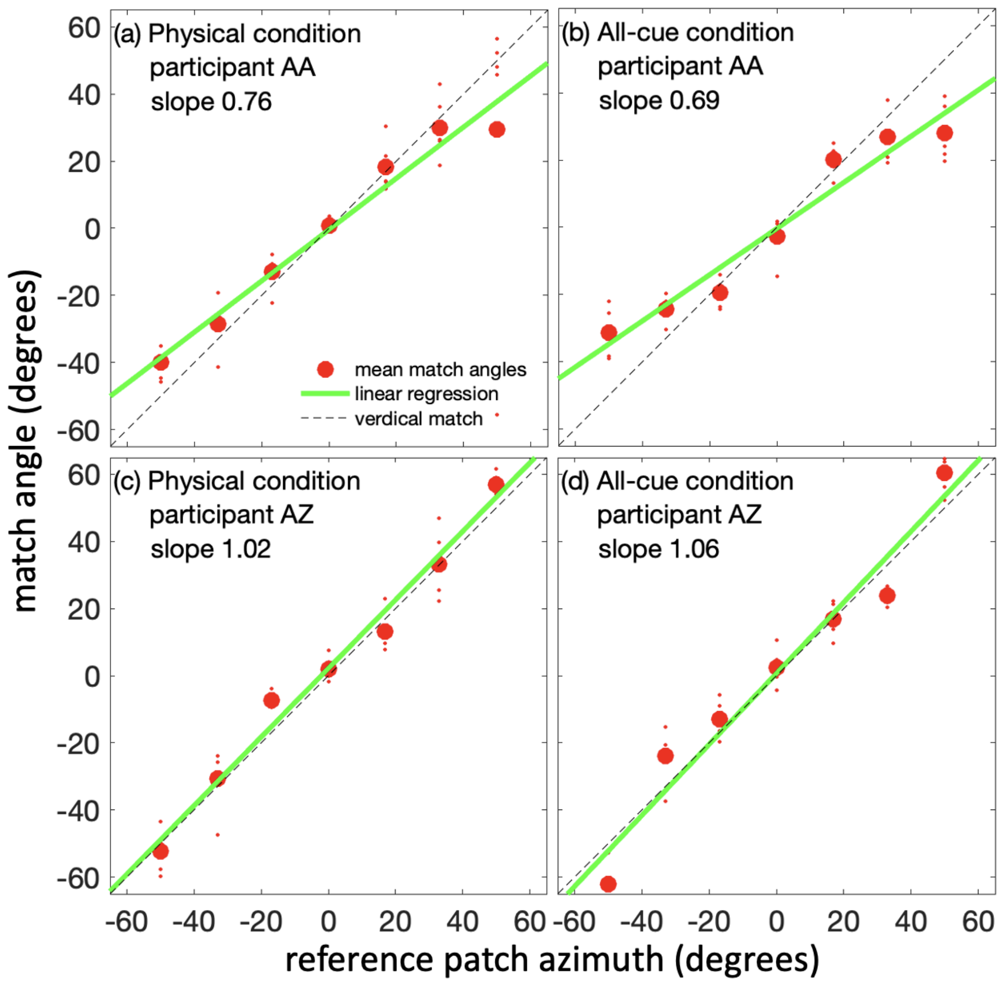

Figure 6. Orientation matches for two typical participants in the Physical and All-Cue conditions. The small red dots show matches on individual trials, and the large red dots are means. The green lines show a least-squares regression of match angle against reference patch angle. The black dashed line represents veridical orientation matches for reference.

Orientation matches. Figure 6 shows orientation matches as a function of the true reference patch orientation, for two typical participants in the Physical and All-Cue conditions. The figure also shows least-squares regression lines. Participants were able to make consistent matches that were an approximately linear function of the true orientation of the reference patch, but the slope of the regression lines often departed from one, indicating non-veridical matches.

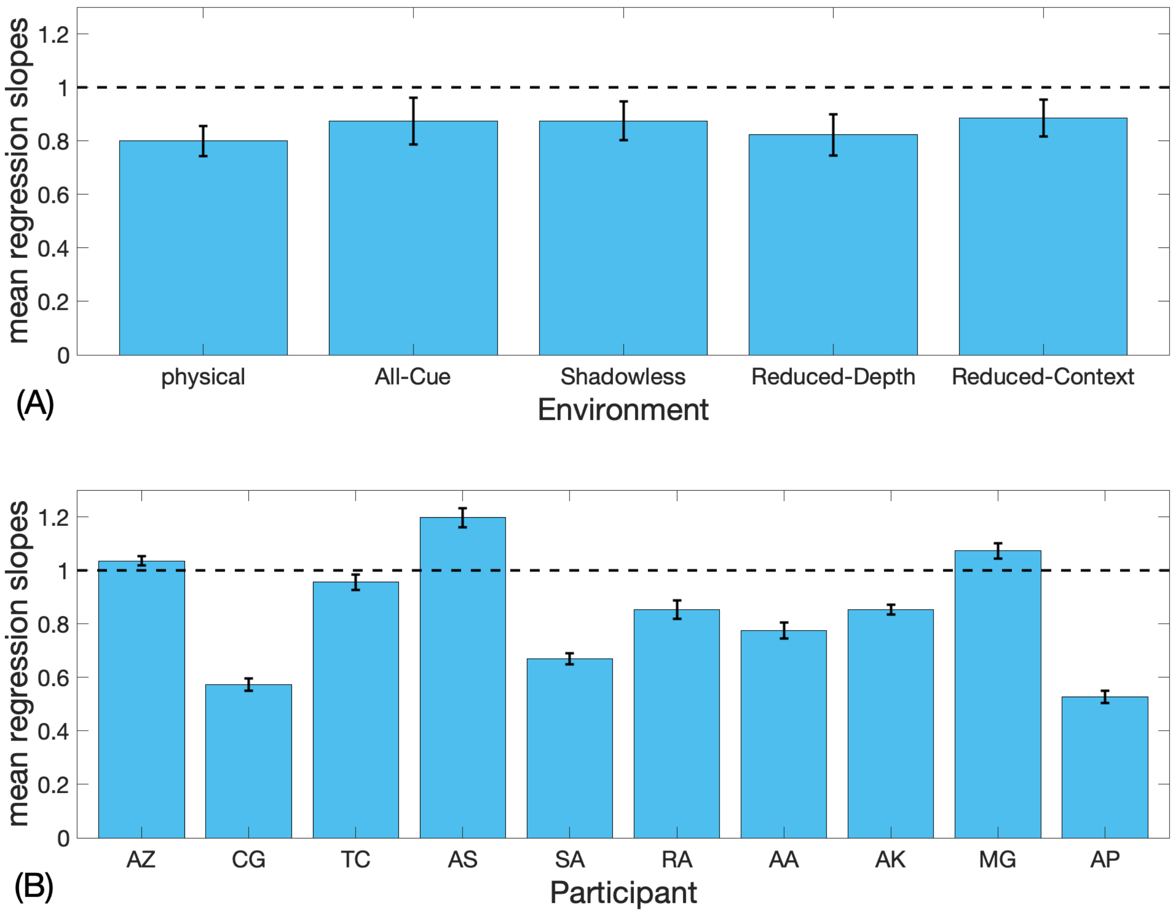

Figure 7. Average regression slopes in the orientation matching task, (a) for each condition and (b) for each participant. Error bars show standard errors. The black dashed line shows a value of 1.0, corresponding to veridical orientation matches.

Figure 7 shows the average regression slope in the orientation matching task for each condition (top panel) and participant (bottom panel). The average slope across all conditions was 0.85, which was significantly less than 1 ($t(49)$ = -4.69, $p < 0.05$), indicating a consistent pattern of underestimation in orientation judgments, as has also been found in previous studies (Bloj et al., 2004; Boyaci et al., 2003; Saunders & Chen, 2015). However, Figure 7(b) also shows that there were substantial individual differences, with some participants reporting orientations that were approximately veridical (regression slopes around 1.0), and others exhibiting a strong bias toward frontoparallel (slopes less than 1.0).

A Levene's test for homogeneity of variance of the regression slopes across the five conditions yielded non-significant results ($p = 0.78$). Consequently, we used a repeated measures ANOVA to compare regression slopes across conditions, and found no significant effect of condition ($F(36, 4) = 2.14$, $p = .08$).

Discussion

Constancy in Physical and VR conditions. In the context of the EIM, saying that a participant overestimated lighting diffuseness (i.e., had a low fitted ICE value) implies that they had imperfect lightness constancy. An overestimate of diffuseness means that as the reference patch orientation varied from trial to trial, the participant responded as though its luminance should change relatively little, and so erroneously attributed the observed luminance change to a change in reflectance. We found that ICE was significantly higher in the Physical condition than in all VR conditions, indicating better constancy in the physical environment. This is consistent with previous studies with similar stimuli and tasks (Morgenstern et al., 2014). We return to this point in the General Discussion, where we consider possible reasons for this finding.

Orientation matches. People often underestimate the slant of planar surfaces, meaning that they show a bias towards frontoparallel orientations (Koenderink, Van Doorn, & Kappers, 1992, 1996; Saunders & Chen, 2015). How would this bias affect the results of experiments where participants judge lightness at a range of slants? Some previous EIM studies have measured this slant bias, and examined participants' lightness matches as a function of *perceived* slant as well as true slant (Bloj et al., 2004; Boyaci et al., 2003). They have typically found that the slant bias was small, and that it had little effect on the conclusions to be drawn from lightness matching data.

We also found that the bias was small, with a mean regression slope of 0.85 in the orientation matching task. We recalculated each participant's ICE value, revising the EIM in equation (5) so that the true reference patch azimuth ψ_n was replaced by perceived azimuth $m\,\psi_n$, where m is the regression slope in the orientation matching task for each participant. The resulting ICE values for individual participants were highly variable. Figure 8a shows ICE for each participant in each condition, calculated using true stimulus azimuth and perceived azimuth. The figure shows that the revised ICE values were lower than the original values for many participants (31% of participants, shown in red), who had orientation matching slopes greater than one, and higher for the rest, who had orientation matching slopes less than one. For a substantial proportion of participants and conditions (21%, shown in green), the revised ICE values were the maximum possible value (1.29), corresponding to completely directional lighting with no ambient component at all (Morgenstern et al., 2014). The recalculated ICE values spanned almost the full range of possible values.

However, we can still give a general picture of how the slant bias may affect lightness judgments. Figure 8b shows mean ICE in each condition, where we recalculated each participant's ICE value using the mean regression slope (0.85) in the orientation matching task, averaged over all participants and conditions, instead of using each individual participant's regression slope. Here the ICE values are markedly higher than in the analysis based on true slant (Figure 5a), and closer to the ICE values of the illuminant

(0.55 and 0.56).

We regard this analysis as tentative. Individual participants' orientation matching slopes were highly variable, and furthermore these slopes were based on explicit judgements of slant, which are not necessarily the same as the implicit judgements that guide reflectance matches. However, this analysis does show that biased slant judgements are a potential reason why participants showed imperfect lightness constancy.

We also examined the effect of outliers in the orientation matching task. Participants occasionally made extreme orientation matches, as can be seen for example in Figure 6(a), which has a small red dot at the bottom right of the panel, to the right of the legend. Among the 50 regression fits (ten participants in five conditions), each based on 35 orientation matches, we found that 36 fits had no matches more than three standard deviations from the least-squares regression line, 13 fits had one outlying match, and one fit had two outlying matches. Thus these outliers were not very common (less than 1% of matches). We compared the regression slopes estimated by least-squares regression (which are the slopes reported above), and by the more robust iteratively reweighted least squares method. We found that using robust regression changed the slopes by an average of just 3.9%, with about half the changes (44%) upwards and half downwards. We conclude that outliers in the orientation matching task had little effect on our findings.

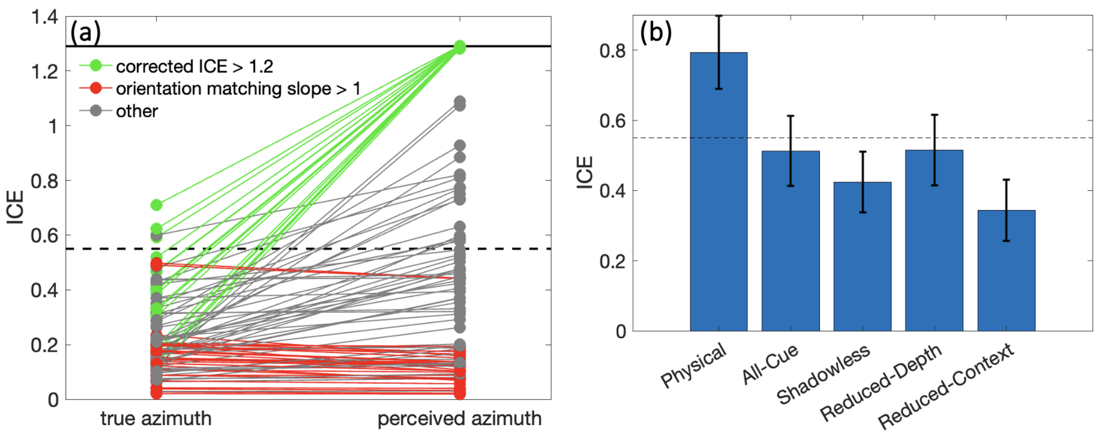

Figure 8. (a) ICE estimates based on actual and perceived orientation of the reference patch, for all participants in all conditions. (b) ICE estimates based on perceived orientation, averaged by condition.

Lightness constancy in the Reduced-Context condition. One surprising result of this experiment was that participants did not have significantly worse lightness constancy in the Reduced-Context condition than in the other VR conditions. The Reduced-Context condition was a control condition where the scene provided no cues to the directional distribution of lighting (e.g., no shaded 3D objects), so we expected that participants would be unable to discount the effect of trial-to-trial changes in the reference patch orientation, and would be forced to follow a suboptimal strategy such as luminance matching. Indeed, this prediction seemed so obvious that we considered not including this control condition, and such a condition has typically not been included in previous studies using the EIM. However, contrary to our expectations, fitted ICE values were about the same in the Reduced-Context condition as in the other VR conditions.

One consequence of this unexpected finding is that we cannot make strong conclusions from the Shadowless and Reduced-Depth conditions. We had expected that constancy would be good in the All-Cue condition, much worse in the Reduced-Context condition, and at some intermediate levels in the Shadowless and Reduced-Depth conditions that would indicate how important shadows and depth cues are for lightness constancy in this task. In fact, constancy was about the same in the All-Cue and Reduced-Context conditions, which left little room for intermediate levels of constancy in the other conditions. These conditions did, however, contribute to the study by showing that the ICE measurements were consistent across three largely similar conditions (All-Cue, Shadowless, and Reduced-Depth; Figure 5).

One possible explanation for the constancy observed in the Reduced-Context condition is that participants learned the directional distribution of lighting in the conditions that they ran in before the Reduced-Context condition (condition order was randomized), and perhaps also during the instructions that were given using the physical apparatus at the beginning of the experiment. Participants may have used this learned lighting information to guide their reflectance matches in the Reduced-Context condition. Adams, Graf, and Ernst (2004) showed that viewers are in fact able to use lighting information learned in one task to guide lightness matches in a different task. A related possibility is that participants' responses in the Reduced-Context condition may have been based on a lighting prior that they learned in everyday life, before participating in the experiment, and that this prior was close enough to the lighting conditions in the experiment to support some degree of lightness constancy.

One approach to testing these hypotheses would be to reanalyze the results of Experiment 1 by condition order. However, all participants had experience with the physical apparatus during the instruction phase, including several practice trials, before running in any VR conditions. Instead, we took the more direct approach of running a new experiment with new participants. We ran a second experiment in which (a) each participant ran in only one condition, and (b) lighting direction varied from one condition to another. The All-Cue condition was the same as in Experiment 1. The Reduced-Context-Right condition was the same as the Reduced-Context condition in Experiment 1, and the Reduced-Context-Left condition was the same except that directional lighting came from behind the participant and to the left instead of the right. A new set of 36 participants ran in Experiment 2, with 12 participants in each condition. Thus there was no possibility of participants learning lighting information in one condition and transferring it to others.

Experiment 2

Methods

Participants. There were 36 new, naive participants, with 12 in each of the three conditions. The same screening, payment, and research ethics protocols were followed as in Experiment 1.

Procedure. We used the same VR stimulus generation methods and data analysis methods as in Experiment 1. The procedure was also the same, except that participants were provided with instructions using a demonstration in the VR scene of the condition that they were assigned to, instead of the physical apparatus as in Experiment 1. At no point were the participants in this experiment shown the physical apparatus used in Experiment 1.

The All-Cue and the Reduced-Context-Right conditions were the same as the All-Cue and the Reduced-Context condition in Experiment 1, respectively. The Reduced-Context-Left condition was the same except that the simulated light was behind the participant and 22° to the left instead of the right. We included the Reduced-Cue-Left condition as well as the Reduced-Cue-Right condition in order to test whether the simulated lighting direction would have any effect on participants' responses in these impoverished stimulus conditions. In each condition participants completed the same lightness matching and orientation matching tasks as in Experiment 1.

Results and discussion

Figure 9 shows mean ICE in each condition, calculated using the true azimuth of the reference patch. We reproduce the ICE values from Experiment 1 (Figure 5a) for comparison. Two of the 12 participants in each condition of Experiment 2 were excluded because their mean absolute error in the orientation matching task exceeded 10°. We report fitted ICE values for all participants in all conditions in Appendix **??**.

Figure 9 shows that mean ICE was low in the Reduced-Context-Right (M = 0.07) and Reduced-Context-Left (M = 0.08) conditions, but t-tests showed that both values were nevertheless significantly greater than zero ($t(9) = 9.02$ $p < 0.05$, $t(9) = 4.59$, $p < 0.05$). Thus participants showed a residual degree of lightness constancy in the Reduced-Context conditions even when they had not previously completed conditions that provided stronger lighting cues. We return to this unexpected finding in the General Discussion.

The EIM includes a parameter ψ_s that models the direction of peak lighting intensity that guides the participant's lightness matches (equation (5)). An independent samples t-test showed that the mean value of this parameter was significantly different in the Reduced-Context-Left (M = -25°) and Reduced-Context-Right (M = 21°) conditions ($t(18) = 2.94$, $p < 0.05$). This rules out the hypothesis that participants' lightness matches were completely guided by lighting information they had learned in everyday life, previous to the experiment.

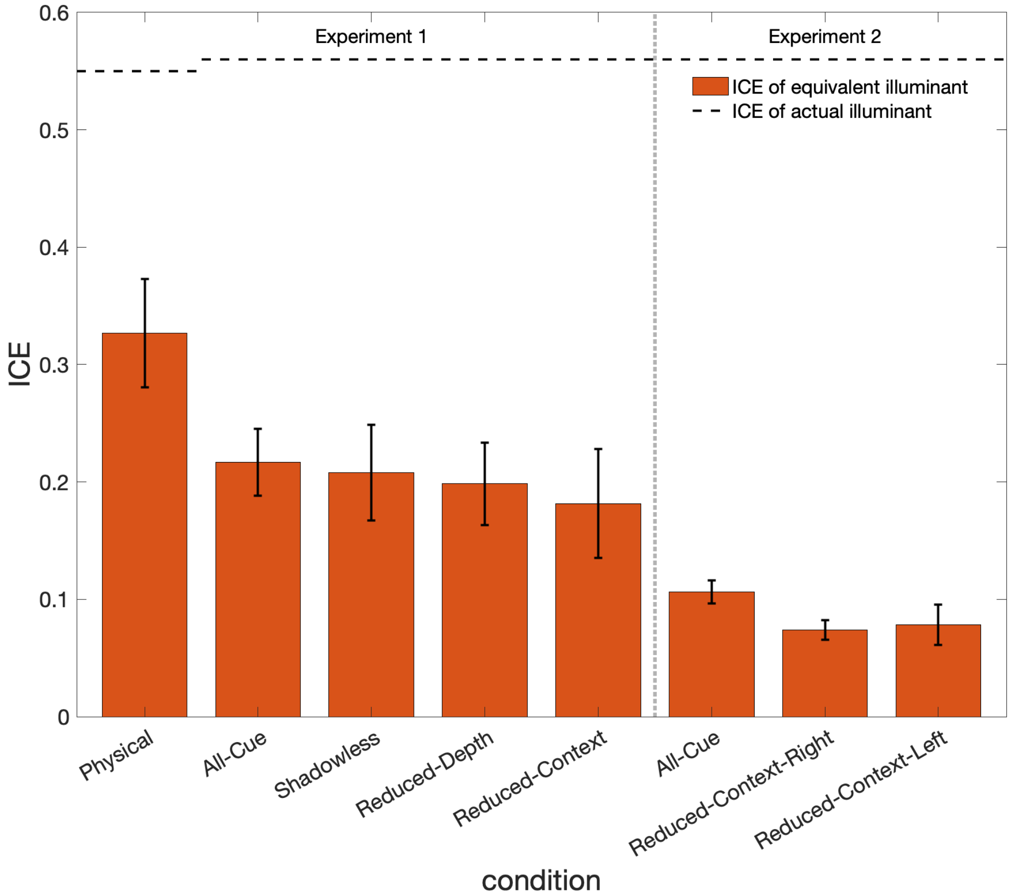

Figure 9. Mean ICE in Experiments 1 and 2, averaged by condition. The black dashed line indicates the ICE of the actual lighting in the Physical (0.55) and VR conditions (0.56).

t-tests showed that mean ICE was significantly greater in the All-Cue condition (M = 0.11) than in the Reduced-Context-Right condition ($t(18) = 2.55$, $p < 0.05$) but not than in the the Reduced-Context-Left condition ($t(18) = 1.42$, $p = 0.18$). This suggests that participants benefited somewhat from the additional lighting cues available in the All-Cue condition. However, perhaps more important is that they did not benefit a great deal: ICE values were not markedly different in the All-cue condition, where many lighting cues were available, than in the Reduced-Cue conditions, where none were available.

t-tests also indicated that mean ICE in the All-Cue condition was significantly greater in Experiment 1 (M = 0.22) than in Experiment 2 (M = 0.11; $t(18) = 3.66$, $p < 0.05$). Furthermore, mean ICE was significantly higher in the Reduced-Context condition in Experiment 1 than in the Reduced-Context-Right condition ($t(18) = 2.28$, $p < 0.05$) but not the Reduced-Context-Left condition ($t(18) = 2.08$, $p = 0.052$) in Experiment 2. These results show that participants in Experiment 1 benefited from having completed several experimental conditions, and possibly also from having viewed the physical apparatus during the instructions.

General discussion

Lightness constancy evaluated via Thouless ratios. How do the levels of lightness constancy reported here compare to typical results in previous studies? We have described participants' lightness matching behaviour using the parameters of the EIM (e.g., ICE for lighting diffuseness), which makes it difficult to directly compare our results to previous studies that did not use the EIM.

The Thouless ratio is a widely used, relatively atheoretical metric for constancy that allows comparisons across studies (Thouless, 1931). Suppose that in a lightness matching task, the correct reflectance match setting is r_1, and the reflectance setting that would result from luminance matching is r_0. If the observer's actual match setting is r_m, then their Thouless ratio is

$$\tau = \frac{\log r_m - \log r_0}{\log r_1 - \log r_0} \tag{6}$$

The Thouless ratio simply reports the relative placement of log r_m between the two reference points of log r_0 and log r_1. It is zero when $r_m = r_0$, indicating no constancy, and one when $r_m = r_1$, indicating perfect constancy.

In a typical design for an asymmetric lightness matching task, we show a reference patch with a fixed reflectance, we vary the illuminance at the reference patch, and we observe how these changes in illuminance affect the viewer's reflectance match setting at a separate match patch. That is precisely the design of the present experiments, with the qualification that instead of varying illuminance by adjusting the intensity of the light source, we vary illuminance by adjusting the orientation of the reference patch relative to the light source.

In Appendix **??** we show that in an asymmetric lightness matching task where illuminance is varied at the reference patch, a plot of match reflectance vs. reference illuminance on log-log axes has slope $1 - \tau$, where τ is the Thouless ratio. Thus we can find an observer's Thouless ratio by plotting match reflectance as a function of illuminance at the reference patch.

Figure 10(a) shows such plots for simulated EIM model observers. Here we assume that the true lighting has directional and ambient components, with an ICE value of 0.55, as in the physical apparatus used above. The model observers also assume that the light has directional and ambient components, but each model observer assumes a different ICE value. We calculated the match reflectance that each observer would assign to a Lambertian reference patch at a range of orientations. In the Supporting Information we provide code for this simulation.

Figure 10(a) plots the resulting match reflectances versus the illuminance at the reference patch for a range of orientations. The model observer with the ICE parameter matching the true illumination (ICE = 0.55) has perfect constancy: its match reflectance does not change as the orientation (and hence illuminance) of the reference patch varies. The slope of this line is zero, so as noted above, the corresponding Thouless ratio is given

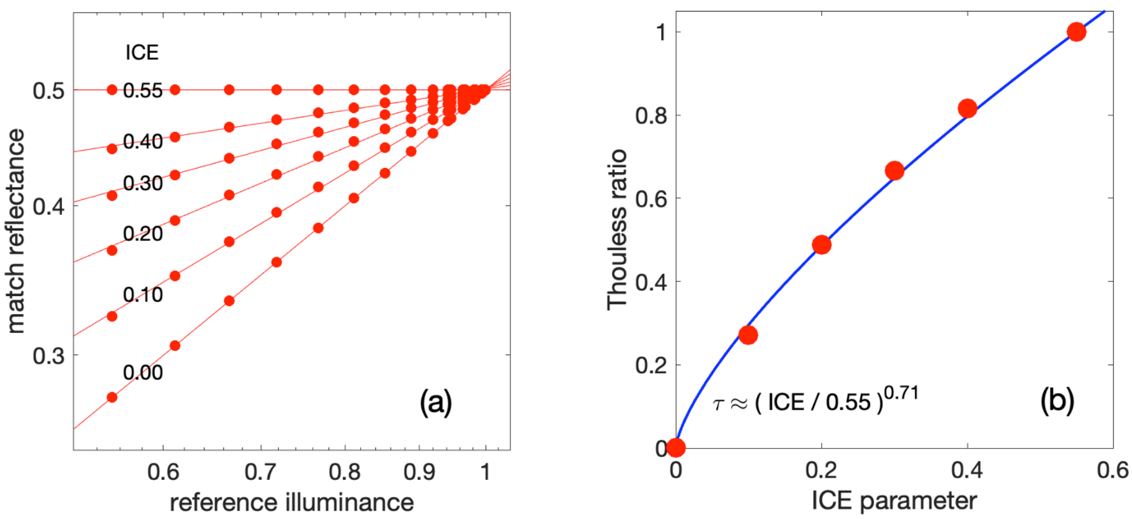

Figure 10. Simulated EIM observers. (a) Match reflectance as a function of reference illuminance for model observers with several ICE parameter values. (b) Thouless ratio as a function of ICE parameter value. The blue line shows the fitted power function.

by $1 - \tau = 0$, meaning $\tau = 1$. The model observer with ICE parameter 0 has no lightness constancy, and instead matches luminance. The slope of this observer's line is one, so $\tau = 0$.

Interestingly, model observers with intermediate ICE parameter values (0.1 to 0.4) also have match settings that are approximately linear functions of illuminance on log-log axes, so they too can be summarized by Thouless ratios given by one minus the slope of each regression line. Figure 10(b) summarizes these results by plotting each model observer's Thouless ratio as a function of its ICE parameter. A least-squares fit of a power function to this data shows that the Thouless ratio is well-fit by $\tau = (\text{ICE}/0.55)^{0.71}$.

We can use this result to transform ICE values in the present experiments into Thouless ratios. Figure 11 shows the data from the summary in Figure 9, with ICE values transformed into Thouless ratios using the power law relationship. Lightness constancy in the Physical condition was quite good (mean $\tau = 0.68$), considering the demanding task of estimating the reflectance of an isolated 3D patch in a complex scene. Even in simpler tasks that only require viewers to discount a single shadow boundary on a single flat surface, Thouless ratios typically range from 0.8 to 0.9 (e.g., Patel et al., 2018; Patel, Wilcox, Maloney, Ehinger, Patel, et al., 2024). Constancy was notably worse in the VR conditions, with mean Thouless ratios ranging from 0.26 to 0.51. For example, in the All-Cue condition of Experiment 2, the mean Thouless ratio was 0.31, meaning that for each log unit increase in illuminance, participants mistakenly increased the match reflectance by 0.69 log units, and only attributed 0.31 log units of the increase to a change in lighting. These findings are consistent with a follow-up study in which we improved the realism of the VR scenes in these experiments in several ways, but nevertheless still found Thouless ratios with an average of just 0.22 (Patel, Wilcox, Maloney, Ehinger, Singh, &

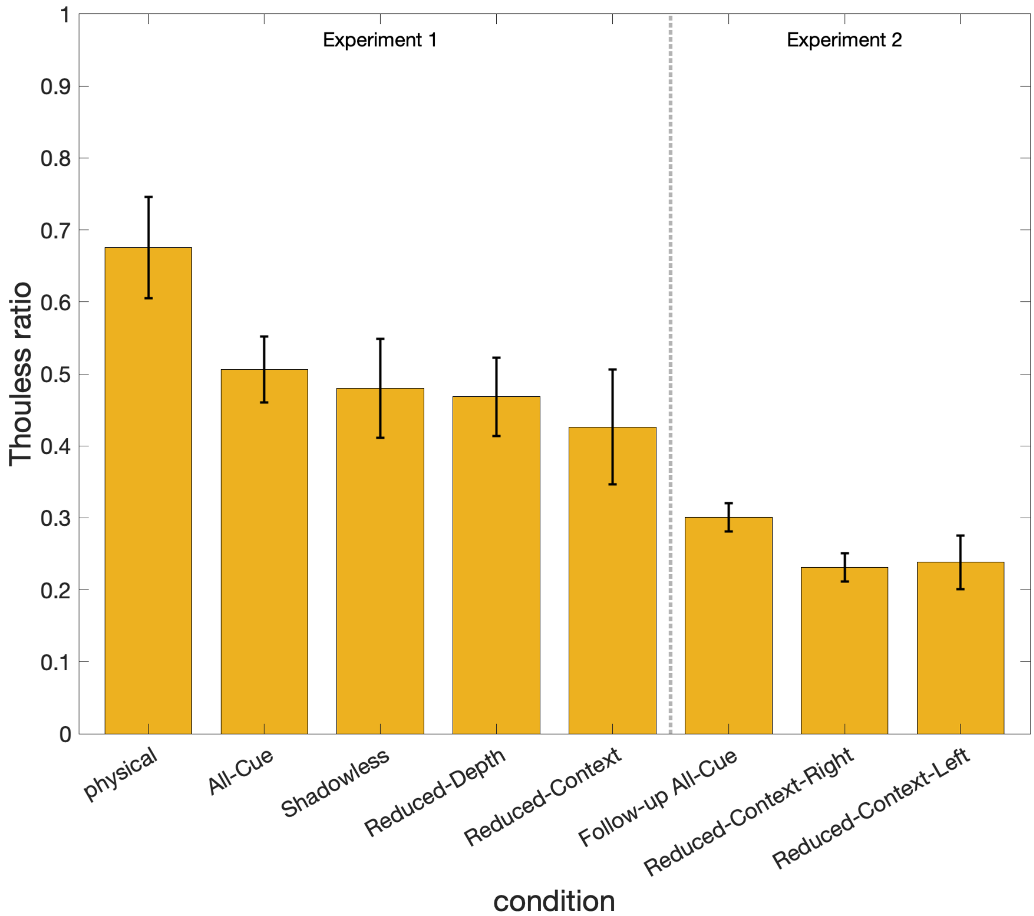

Figure 11. Mean Thouless ratios in Experiments 1 and 2, averaged by condition.

Murray, 2024).

These results are similar to those from previous studies that used the EIM to model lightness constancy. Bloj et al. (2004, Experiments 1 and 2) used a physical apparatus with lighting diffuseness given approximately by ICE = 0.93. The average ICE parameter for their participants was 0.45, which corresponds to a Thouless ratio of 0.64. (See the Supporting Information for power law relationships between ICE and Thouless ratios for the studies discussed in this paragraph; the power law given above is specific to lighting diffuseness ICE = 0.55.) Similarly, Boyaci et al. (2006, Experiment 1) used a stereoscope and computer-generated stimuli where the simulated lighting had diffuseness given by ICE = 0.43. Their participants had average ICE parameter 0.06, for a Thouless ratio of 0.23. Table 1 reports similar analyses of additional previous studies. Morgenstern et al. (2014) compared results across EIM studies, and found that overestimates of diffuseness (and hence failures of constancy) were particularly strong in studies that used virtual scenes. Their conclusions were limited, however, by the fact that they compared results across studies with very different scenes in physical and virtual environments. Our experiments, which carefully matched physical and virtual scenes, show a similar pattern of results, and support Morgenstern et al.'s conclusions.

Table 1
Thouless ratios in previous studies

	True ICE	Perceived ICE	Thouless Ratio	Stimulus Type
Bloj et al. (2004) Expts 1, 2	0.93	0.45	0.60	Physical
Bloj et al. (2004) Expts 3	0.69	0.38	0.69	Physical
Boyaci et al. (2006) Expt 1	0.43	0.06	0.20	Stereoscope
Boyaci et al. (2006) Expt 2	0.76	0.06	0.10	Stereoscope
Boyaci et al. (2003)	0.38	0.09	0.33	Stereoscope

Imperfect lightness constancy. Lightness constancy can vary substantially from one environment to another. As noted by Gilchrist (2006), even in some of Katz's pioneering experiments on lightness constancy, Thouless ratios were often quite low and ranged from 0.35 to 0.7, perhaps because the stimuli were simple scenes that provided few lighting cues. In richer and more complex scenes, though, Thouless ratios and similar color constancy indices are frequently as high as 0.8 or 0.9 (Kraft & Brainard, 1999; Patel, Wilcox, Maloney, Ehinger, Patel, et al., 2024). What are the factors that led to the wide range of Thouless ratios observed in our experiments?

One factor that may have reduced overall constancy is that we used simple Lambertian objects, in order make it easier to match physical and VR scenes. Boyaci et al. (2006) examined the impact of specular highlights, shading, and cast shadows on lightness constancy in a task like ours, where the orientation of a reference patch varied relative to a light source. They found that all three of these cues contributed to viewers' ability to discount lighting when estimating reflectance. Future work on realism in VR should investigate the role that realistic material properties play in viewers' ability to estimate lighting conditions, as it seems plausible that including such materials could provide useful cues to lighting conditions and result in better lightness and color constancy.

Another factor that may have reduced lightness constancy compared to some previous studies is the difficulty of our task. Many studies of constancy have used 'Mondrian' stimuli, where participants match reflectance in relatively simple 2D patterns with single shadow boundaries (Land & McCann, 1971). In EIM studies such as the present one, viewers must integrate information across a scene in order to estimate the directional distribution of lighting, estimate the 3D orientation of a test patch, and combine these estimates to discount the effect of lighting on the luminance of the test patch. Viewers can be highly insensitive to lighting inconsistencies across a scene (Ostrovsky et al., 2005, Wilder et al., 2019; but see Morgenstern et al., 2011), and in that sense it is remarkable that they have any lightness constancy at all in tasks like the present one, where they must integrate lighting information across distinct objects. In particular, we speculate that constancy may have been lower than in some previous studies because here the reference stimulus was an isolated patch supported by a thin pole; it would be interesting to test whether constancy is better for surfaces that are part of complex shaded objects that provide information about lighting in immediately adjacent regions.

An additional factor that may have contributed to low levels of lightness constancy is that participants largely remained stationary. They used a chin-rest in the physical condition, and were asked to keep their head at the position of a fixed green cube in the VR conditions. It is possible that actively exploring the environment (as in Gil Rodríguez et al. (2024), for example) would allow participants to make a better estimate of the lighting conditions, and may improve constancy.

There are also several possible reasons why lightness constancy was better in the Physical condition than in the VR conditions. First, although we made the physical stimuli smooth and Lambertian, there were inevitably imperfections and fine surface textures in the physical scene that were not present in the VR scenes, and these features may have helped viewers to disentangle lighting and surface properties. Second, the VR scenes were generated using Unity's Built-in Render Pipeline, which is not a physically-based renderer, and the approximations that it makes for real-time rendering (e.g., shadow penumbras determined by manually set parameters rather than by the size of the simulated light source) may have reduced viewers' ability to estimate and discount lighting conditions. Third, luminances were higher in the VR scenes than in the physical scene. In both cases, luminances were well into the photopic range, so it seems unlikely to us that this was a significant factor. Nevertheless, sensitivity to fine image features is better at higher luminances (DeValois & DeValois, 1990), and it is possible that perception of subtle material properties and image pixellation played a role in the differences we found between physical and VR conditions. Patel, Wilcox, Maloney, Ehinger, Patel, et al. (2024) briefly review some of the ways in which performance in physical and VR environments has been found to diverge on a range of tasks, and with the current state of VR technology it seems to be the rule rather than the exception that performance in virtual environments falls short of that in real, physical environments.

There were also large and significant differences between Experiment 1, where all participants completed all conditions, and Experiment 2, where each participant completed just one condition. Lightness constancy was markedly better in Experiment 1. One reason may have been that in Experiment 1 all participants were given instructions using the physical apparatus, which may have helped them to understand the task of matching reflectance, and may also have provided them with strong lighting cues that improved their performance in all conditions (W. J. Adams et al., 2004). Another, related reason may have been that experiencing several conditions led to a practice effect, where participants were able to develop and use a better sense of the lighting distribution, which was the same in all conditions.

Like most modelling endeavours, our use of the EIM relies on some assumptions. We have assumed, for example, that participants correctly estimate the elevation of the directional light source ($\phi_s = 22°$) and correctly see the reference patch as upright ($\phi_n =$

0°). With these assumptions, the EIM explains participants' failures of lightness constancy in terms of mistakenly high estimates of lighting diffuseness. Alternative explanations, however, could be in terms of mis-estimates of lighting elevation or other parameters, so it is important to verify these assumptions. Boyaci et al. (2004) found that viewers' surface colour judgements were guided by generally accurate estimates of lighting elevation (ϕ_s), though with a slight upward bias. The same study found that viewers' judgements of the elevation of a test patch (ϕ_n) were highly correlated with the true elevation, and they found only small biases (typically 5° or less) for an upright patch like the one we used here. These findings support our assumptions, and suggest that failures of these assumptions do not explain the large departures from lightness constancy that we observed. Nevertheless, Boyaci et al. used a custom-built stereoscope, not a physical environment or VR headset, so this is an interesting topic for further invesetigation. In any case, our basic empirical finding of poor constancy in VR stands, and does not depend on the ways in which we have used the EIM to arrive at interpretations of this finding.

Lightness constancy in Reduced-Context conditions. A surprising finding in Experiment 1 was that participants showed some lightness constancy even in the Reduced-Context condition, where there were no lighting cues that could have allowed them to estimate and discount illumination. Experiment 2 replicated this result, and also showed that the lighting direction parameter of the fitted EIM tracked the true direction of lighting in the Reduced-Context scene. How is this possible?

We believe that the answer lies in the fact that the EIM attributes all departures from luminance matching to illumination discounting. In work on lightness perception, luminance matching is often taken to be the signature of a complete failure of lightness constancy, but research on brightness perception shows that even when viewers try to match the luminance of two stimulus regions, there are often substantial and complex biases in their responses (Kingdom, 2011). In general, we should not expect reliable luminance matching from participants. Once this is pointed out, the apparent residual constancy in the Reduced-Context conditions is less mysterious: departures from luminance matching are simply interpreted as partial constancy in an EIM analysis.

Figure 12 illustrates this effect, and is modelled after Figure 4 that shows results from human observers. The figure shows results from a simulated model observer that we applied to screen captures from the Reduced-Context condition in our experiments. The model observer does not estimate lighting conditions or surface orientation, but simply passes the stimulus values at the reference and match locations through two different pixelwise nonlinearities. These different nonlinearities could result, for example, from the different contexts at those two locations. The model observer's match setting is the match reflectance at which the outputs of these two nonlinearities (at the reference and match patches) have the same value. The results for this simple model observer are similar to those of human participants in the Reduced-Context conditions. From a fit of the EIM to the model observer's data, we can even make a good estimate of the lighting direction in this scene (it is the angle at which the green fit line reaches a maximum), even though the scene contains no reliable lighting cues. This simulation illustrates how low levels of lightness constancy can result from simple processes, and shows the importance of including control conditions like the Reduced-Context condition in lightness matching experiments, in order to establish a baseline for levels of constancy that have little to do with estimating and discounting illumination.

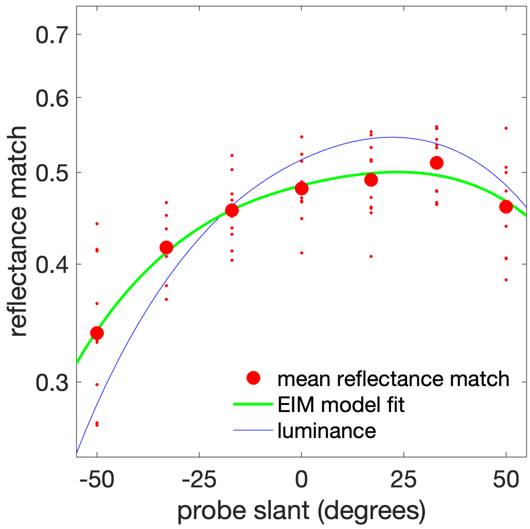

Figure 12. Results from a model observer in the Reduced-Context condition. The model's match reflectance is simply a stochastic, nonlinear function of the reference luminance. The model mimics a degree of lightness constancy (ICE = 0.17, Thouless ratio = 0.43), despite being unable to estimate 3D orientation or lighting conditions.

<u>Concluding remarks.</u> These experiments reveal interesting and important differences between perception in real and virtual environments. VR technology has made rapid advances in recent years, and immersive, interactive environments with a wide range of cues to lighting, shape, and materials can be generated in real time. And yet, we found that even in a simple lightness matching task, performance in VR falls well below performance with physical objects, and in some cases is not much better than luminance matching. Discovering the reasons for this performance difference is a promising research problem, both in order to improve performance in critical tasks in virtual environments, and for our fundamental understanding of how the human visual system estimates properties of the real world from retinal images.

Declarations

Funding: This research was funded by a VISTA doctoral scholarship, NSERC CGS Doctoral scholarship and NSERC CGS MSFSS award to Khushbu Y. Patel, and an NSERC Discovery Grant to Richard F. Murray.

Conflict of interest: All authors declare no conflict of interest.

Ethics approval: Ethical approval for this study was granted by York University's Office of Research Ethics.

Consent to participate: All participants gave written informed consent before starting the study.

Consent for publication: Written consent for publication of behavioral data was obtained from all participants.

Availability of data and materials: Data for all experiments is available on OSF at https://doi.org/10.17605/OSF.IO/3GXRA.

Code availability: Analysis code for all experiments is available on OSF at https://doi.org/10.17605/OSF.IO/3GXRA.

Preregistration: This study was not preregistered.

Author contributions: All authors contributed to the conceptualization, design of methodology, and revision of this manuscript. Khushbu Y. Patel was responsible for programming, conducting experiments, data collection, statistical analysis, and drafting the original version of the published work. Richard F. Murray provided primary supervision, collaborated closely on the statistical analysis, programming of the experiments, revised the drafts of the manuscript and led the calibrations detailed in a separate publication (Murray et al., 2022). Richard F. Murray and Khushbu Y. Patel designed the stimulus base and objects, which were constructed by Christopher Giverin. Richard F. Murray primarily constructed the physical apparatus using the robotic components, while Khushbu Y. Patel painted each element of the stimulus background and matched it in Unity. Laurie M. Wilcox provided guidance during the development of the VR scene. Jaykishan Y. Patel assisted in data collection and programming of the experiments.

References

Adams, H., Stefanucci, J., Creem-Regehr, S., & Bodenheimer, B. (2022). Depth perception in augmented reality: The effects of display, shadow, and position. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (pp. 792–801).

Adams, H., Stefanucci, J., Creem-Regehr, S., Pointon, G., Thompson, W., & Bodenheimer, B. (2021). Shedding light on cast shadows: An investigation of perceived ground contact in AR and VR. *IEEE Transactions on Visualization and Computer Graphics*, *28*(12), 4624–4639.

Adams, W. J., Graf, E. W., & Ernst, M. O. (2004). Experience can change the 'light-from-above' prior. *Nature Neuroscience*, *7*(10), 1057-1058.

Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, *262*(5142), 2042–2044.

Blakeslee, B., Reetz, D., & McCourt, M. (2008). Coming to terms with lightness and brightness: Effects of stimulus configuration and instructions on brightness and lightness judgments. *Journal of Vision*, *8*(11), 1-18.

Bloj, M., Kersten, D., & Hurlbert, A. (1999). Perception of three-dimensional shape influences colour perception through mutual illumination. *Nature*, *402*(6764), 877–879.

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), 6–6.

Boyaci, H., Doerschner, K., & Maloney, L. T. (2004). Perceived surface color in binocularly viewed scenes with two light sources differing in chromaticity. *Journal of Vision*, *4*(9), 664-679.

Boyaci, H., Doerschner, K., & Maloney, L. T. (2006). Cues to an equivalent lighting model. *Journal of Vision*, *6*(2), 106-118.

Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision*, *3*(8), 541-553.

Brainard, D. H. (1998). Color constancy in the nearly natural image. 2. Achromatic loci. *JOSA A*, *15*(2), 307–325.

Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illumination models. *Journal of Vision*, *11*(5), 1–1.

Creem-Regehr, S. H., Stefanucci, J. K., & Bodenheimer, B. (2022). Perceiving distance in virtual reality: theoretical insights from contemporary technologies. *Philosophical Transactions of the Royal Society B*, *378*(20210456), 1-12.

DeValois, R. L., & DeValois, K. K. (1990). *Spatial vision.* Oxford University Press.

Díaz-Barrancas, F., Cwierz, H., & Pardo, P. J. (2021). Real-time application of computer graphics improvement techniques using hyperspectral textures in a virtual reality system. *Electronics*, *10*, 1-15.

Díaz-Barrancas, F., Cwierz, H., Pardo, P. J., Pérez, Á. L., & Suero, M. I. (2020). Spectral color management in virtual reality scenes. *Sensors*, *20*, 1-16.

Doerschner, K., Boyaci, H., & Maloney, L. T. (2004). Human observers compensate for secondary illumination originating in nearby chromatic surfaces. *Journal of Vision*, *4*(2), 3–3.

Doerschner, K., Boyaci, H., & Maloney, L. T. (2007). Testing limits on matte surface color perception in three-dimensional scenes with complex light fields. *Vision Research*, *47*(28), 3409–3423.

Flock, H., & Freedberg, E. (1970). Perceived angle of incidence and achromatic surface color. *Perception & Psychophysics*, *8*(4), 251–256.

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

Gil Rodríguez, R., Bayer, F., Toscani, M., Guarnera, D., Guarnera, G. C., & Gegenfurtner, K. R. (2022). Colour calibration of a head mounted display for colour vision research using virtual reality. *SN Computer Science*, *3*(1), 1–10.

Gil Rodríguez, R., Hedjar, L., Toscani, M., Guarnera, D., Guarnera, G. C., & Gegenfurtner, K. R. (2024). Color constancy mechanisms in virtual reality environments. *Journal of Vision*, *24*(5), 1-28.

Hartle, B., & Wilcox, L. M. (2022). Stereoscopic depth constancy for physical objects and their virtual counterparts. *Journal of Vision*, *22*(4), 1-19.

Hibbard, P. B. (2023). Virtual reality for vision science. In *Topics in behavioral neurosciences* (pp. 1–29). Springer.

Kimura, K., Reichert, J. F., Olson, A., Pouya, O. R., Wang, X., Moussavi, Z., & Kelly, D. M. (2017). Orientation in virtual reality does not fully measure up to the real-world. *Scientific Reports*, *7*(1), 1-8.

Kingdom, F. A. A. (2011). Lightness, brightness and transparency: a quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research*, *51*, 652-673.

Kitazaki, M., Kobiki, H., & Maloney, L. T. (2008). Effect of pictorial depth cues, binocular disparity cues and motion parallax depth cues on lightness perception in three-dimensional virtual scenes. *PLoS ONE*, *3*(9), e3177.

Koenderink, J. J., Van Doorn, A. J., & Kappers, A. M. (1992). Surface perception in pictures. *Perception & Psychophysics*, *52*, 487–496.

Koenderink, J. J., Van Doorn, A. J., & Kappers, A. M. (1996). Pictorial surface attitude and local depth comparisons. *Perception & Psychophysics*, *58*(2), 163–173.

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.

Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Josa*, *61*(1), 1–11.

Linowes, J. (2020). *Unity 2020 virtual reality projects.* Packt Publishing Ltd.

Marantz, A. (2016). Studio 360. *The New Yorker*, April 25, 2016, pp. 86-94.

McCluney, R. (1994). *Introduction to radiometry and photometry.* Artech House, Inc.

Morgenstern, Y., Geisler, W. S., & Murray, R. F. (2014). Human vision is attuned to the diffuseness of natural light. *Journal of Vision*, *14*(9), 1-18.

Morgenstern, Y., Geisler, W. S., & Murray, R. F. (2015). The role of natural lighting diffuseness in human visual perception. In B. E. Rogowitz, T. N. Pappas, & H. de Ridder (Eds.), *Proceedings of SPIE 9394, Human Vision and Electronic Imaging XVIII.* San Francisco, CA, February 9-12, 2015.

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*, *108*(30), 12551–12553.

Murray, R. F. (2021). Lightness perception in complex scenes. *Annual Review of Vision Science*, *7*, 417–436.

Murray, R. F., & Adams, W. J. (2019). Visual perception and natural illumination. *Current Opinion in Behavioral Sciences*, *30*, 48–54.

Murray, R. F., Patel, K. Y., & Wiedenmann, E. S. (2022). Luminance calibration of virtual reality displays in unity. *Journal of Vision*, *22(13):1*, 1-9.

Ostrovsky, Y., Cavanagh, P., & Sinha, P. (2005). Perceiving illumination inconsistencies in scenes. *Perception*, *34*(11), 1301–1314.

Patel, K. Y., Munasinghe, A., & Murray, R. F. (2018). Lightness matching and perceptual similarity. *Journal of Vision*, *18*(5), 1-18.

Patel, K. Y., Wilcox, L. M., Maloney, L. T., Ehinger, K. A., Patel, J. Y., Widenmann, E., & Murray, R. F. (2024). Lightness constancy in reality, in virtual reality, and on flat-panel displays. *Behaviour Research Methods*, 5 March 2024.

Patel, K. Y., Wilcox, L. M., Maloney, L. T., Ehinger, K. A., Singh, S., & Murray, R. F. (2024). Evaluating lightness constancy in a high-definition vr environment. In *Color and Imaging Conference* (Vol. 32, p. 166-172).

Pharr, M., Jakob, W., & Humphreys, G. (2023). *Physically based rendering, fourth edition.* The MIT Press.

Pizlo, Z. (2001). Perception viewed as an inverse problem. *Vision Research*, *41*, 3145-3161.

Pont, S. C., & Koenderink, J. J. (2007). Matching illumination of solid objects. *Perception & Psychophysics*, *69*(3), 459–468.

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), 1-18.

Rensink, R. A., & Cavanagh, P. (2004). The influence of cast shadows on visual search. *Perception*, *33*(11), 1339–1358.

Ripamonti, C., Bloj, M., Hauck, R., Mitha, K., Greenwald, S., Maloney, S., & Brainard, D. H. (2004). Measurements of the effect of surface slant on perceived lightness. *Journal of Vision*, *4*(9), 747-763.

Rzepka, A. M., Hussey, K. J., Maltz, M. V., Babin, K., Wilcox, L. M., & Culham, J. C.

(2023). Familiar size affects perception differently in virtual reality and the real world. *Philosophical Transactions of the Royal Society B*, *378*(1869), 1-14.

Saunders, J. A., & Chen, Z. (2015). Perceptual biases and cue weighting in perception of 3d slant from texture and stereo information. *Journal of Vision*, *15*(2:14), 1-24.

Scarfe, P., & Glennerster, A. (2015). Using high-fidelity virtual reality to study perception in freely moving observers. *Journal of Vision*, *15*(9), 3–3.

Scarfe, P., & Glennerster, A. (2019). The science behind virtual reality displays. *Annual Review of Vision Science*, *5*, 529–547.

Thouless, R. (1931). Phenomenal regression to the real object. *British Journal of Psychology*, *21*(4), 339-359.

Unity Technologies. (2020). *Unity, version 2019.3.3f1.* Retrieved from https://docs.unity3d.com/2020.1/Documentation/Manual/UnityManual.html

Wilder, J. D., Adams, W. J., & Murray, R. F. (2019). Shape from shading under inconsistent illumination. *Journal of Vision*, *19(6):2*.