

**REALISM AND FEATURES SUPPORTING LIGHTNESS CONSTANCY
IN VIRTUAL SCENES**

KHUSHBU Y PATEL

**A DISSERTATION SUBMITTED TO
THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY**

**GRADUATE PROGRAM IN PSYCHOLOGY
YORK UNIVERSITY
TORONTO, ONTARIO**

JULY 2024

© Khushbu Patel, 2024

Abstract

Lighting and surface properties play an important role in visual perception. Our visual system decodes two-dimensional retinal images to discern potential three-dimensional scenes. A particular challenge in the context of achromatic lights and surfaces is *lightness constancy* — the ability to maintain a consistent perception of an object’s reflectance despite varying illumination conditions. Although humans are generally adept at maintaining lightness constancy, it is not perfect. This dissertation examines lightness constancy within both real-world and virtual environments, including flat-panel displays and virtual reality (VR).

Chapter 2 evaluates lightness constancy through an asymmetric lightness matching task across a shadow boundary, using physical surfaces, a flat-panel display, and an immersive VR environment. While the VR condition exhibited realistic levels of lightness constancy, participants showed significantly lower levels of constancy in flat-panel display compared to the physical environment. Notably, participant variability was more pronounced in both virtual environments. In Chapter 3 the study extends to lightness matching across various 3D orientations using both physical surfaces and in VR. The findings reveal that lightness constancy is significantly weaker in VR compared to physical environments. Building on Chapter 3, Chapter 4 further evaluates lightness constancy using the same task, but incorporates more realistic and accurate rendering techniques, along with realistic materials, in VR. Contrary to expectations, the results show no notable improvement in lightness constancy, underscoring the persistent challenges in achieving realism for tasks evaluating lightness across 3D orientations in VR.

Despite robust 3D shape, lighting, and depth cues available in VR, constancy is significantly worse in VR and our understanding of lightness perception fails to explain why. This discrepancy highlights a gap in our knowledge, pointing to potentially overlooked factors critical for accurate lightness perception, such as fine material details or subtle surface textures. In conclusion, the findings in this dissertation suggest that VR is a reasonable proxy for real-world scenarios in tasks when lightness is judged across a shadow boundary, but current technology falls short in replicating realistic lightness constancy when image luminance varies from one location to another due to differences in 3D orientation relative to a light source.

Dedication

To family, given and chosen.

Acknowledgements

I extend my deepest gratitude to my supervisor and mentor, Dr. Richard Murray, for his unwavering support, encouragement, guidance, and patience throughout my decade of study under his mentorship. His guidance has been invaluable to my personal and professional growth.

I am also thankful to my dissertation committee members, Dr. Laurie Wilcox and Dr. Laurence Maloney, and Dr. Krista Ehinger, for thoughtful, encouraging, and indispensable advice on this dissertation and throughout our collaborative work. Additionally, I appreciate Dr. Rob Allison, Dr. Kevin Lande, and Dr. Michael Barnett-Cowan for being a part of my committee and showing abundant enthusiasm for my work. Also, a shoutout to the Graduate Department of Psychology for consistently being the calm in our academic storm.

My heartfelt thanks go to my family, especially my parents, whose love, support and sacrifices have been foundational to my journey. I am equally grateful to my in-laws, who, despite being far away, have provided steadfast support throughout my studies. I am particularly grateful to my lab mate and brother, Jaykishan Patel, for his guidance and persistent motivation, and to my husband, Angshuman Ghosh, for his unwavering belief in me and his support. I extend a warm thank you to all my friends whose continuous support has been essential to my success.

This research was funded by an NSERC Alexander Graham Bell Scholarship CSG-D and a VISTA scholarship. I am grateful to these funding agencies for making my research possible.

Contents

Abstract	ii
Dedication	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	x
List of Figures	xi
1 Introduction	1
1.1 Basic Problem of Lightness Perception	1
Key Terminology and Concepts	2
The Ambiguity Problem	4
1.2 Review of Models in Lightness and Brightness Perception	4
Lightness Theories	5
Brightness Theories	10
Intrinsic Image Decomposition	13
1.3 Evaluating Lightness Constancy: Relevant Methodologies and Empirical Findings	15
Lightness Matching Across 2D Shadow Boundaries	15
Lightness Matching Across 3D Orientation Differences	17

1.4	Lightness Perception in Virtual Reality	20
	History and current uses of VR	20
1.5	Current Objectives	22
2	Lightness constancy in reality, in virtual reality, and on flat-panel displays	24
2.1	Abstract	25
2.2	Introduction	25
	Lightness perception	26
	Asymmetric lightness matching	29
	Previous work	31
2.3	Methods	34
	Participants	34
	Stimuli	34
	Procedure	39
2.4	Results	42
2.5	Discussion	47
	Appendices	55
	A	56
	B	57
	C	62
	D	64
	E	66
3	An Equivalent Illuminant Analysis of Lightness Constancy with Physical Objects and in Virtual Reality	67

3.1	Abstract	68
3.2	Introduction	68
	Lightness perception in real and virtual environments	70
	Equivalent illumination models	72
	Present study	74
3.3	Experiment 1	75
	Methods	75
	Results	90
	Discussion	95
3.4	Experiment 2	98
	Methods	98
	Results and discussion	99
3.5	General discussion	101
3.6	Open Practices Statement	109
	Appendices	111
	F ICE parameters in Experiments 1 and 2	112
	G Calculation of Thouless Ratio	113
	H	114
4	Evaluating lightness constancy in a high-definition VR environment	115
4.1	Abstract	116
4.2	Introduction	116
	Equivalent illumination models	118
	Present study	119
4.3	Methods	120
	Participants	120
	Stimuli	120

Procedure	123
Analysis	123
4.4 Results	126
Lightness matches	126
Orientation matches	127
4.5 Discussion	127
Conclusion	130
Appendices	132
I	133
5 Conclusion	134
5.1 How Realistic is Lightness Constancy in VR?	134
5.2 Role of Immersion	137
5.3 Future Directions for VR	138
5.4 Concluding Statement	140
References	141

List of Tables

2.1	Mean Thouless ratios by condition, with bootstrapped standard errors and confidence intervals. The mean for each environment is averaged over reference reflectances and participants, and the mean for each reference reflectance is averaged over environments and participants.	44
3.1	Thouless ratios in previous studies	105

List of Figures

1.1	A visual representation of the concepts of illuminance, reflectance, and luminance as they interact on a surface.	2
1.2	Sequential products calculated based on the retinex theory (Land & McCann, 1971).	7
1.3	The simultaneous contrast illusion. The central gray squares are identical in luminance, but the square against the dark surround appears lighter than the one against the white surround, demonstrating the influence of surrounding luminance on lightness perception.	8
1.4	Visualization of linear filters. Taken from (Kim et al., 2018).	12
1.5	Illustration of apparatus used for a shadow boundary task. (a) physical stimulus (b) computer-generated stimulus. The images are taken from (K. Y. Patel et al., 2018)	16
1.6	An illustration of how a change in 3D orientation of an object’s surface affects the intensity of light reflected from it. The image is taken from (Ripamonti et al., 2004).	18
2.1	The snake illusion (Adelson, 2000). The four diamonds are physically identical, but the top two appear lighter than the bottom two.	28

2.2	A typical asymmetric lightness matching experiment. (a) The reference patch has a fixed reflectance and illumination level. The match patch has an adjustable reflectance, and is shown under a different illumination level. The observer adjusts the match patch so that it appears to have the same reflectance as the reference patch. (b) Lightness matches for three different Thouless ratios. At high Thouless ratios (good constancy), the match reflectance is not strongly affected by the illuminance at the match patch. At low Thouless ratios (poor constancy), the match reflectance decreases substantially as illuminance increases.	31
2.3	A schematic top-down view of the testing room and apparatus in the physical environment.	36
2.4	(a) Photograph of the physical room and apparatus, taken from the viewing position used by observers in the experiment. (b) A screen capture of the VR environment. (c) The panel apparatus rendered and displayed on a flat-panel monitor.	37
2.5	Results for two typical observers in the lightness matching task. Small dots represent reflectance matches on individual trials, and large filled dots represent the median match across trials. Straight lines are sum-of-squares fits, constrained to pass through the open circles that represent the reference stimuli. The three colours in each panel show data for the three reference reflectances. The horizontal dashed lines with slope zero represent perfect constancy ($\tau=1$), and the diagonal dotted lines with slope -1 represent luminance matching ($\tau=0$).	43
2.6	(a) Thouless ratios as a function of environment and reference reflectance. Each triad of blue dots connected by straight lines represents a single observer, and red dots represent means across observers. (b) Mean Thouless ratios by environment. (c) Mean Thouless ratios by reference reflectance.	44

2.7	Thouless ratios in the VR and flat-panel conditions are strongly correlated.	46
2.8	Thouless ratio versus median response time. The black 'x' indicates the median response time and Thouless ratio for each environment. The standard error for the medians was approximately the same size as the marker itself, and so are not displayed in this plot.	47
2.9	(a) The penalty function f_2 . (b) The model's median match settings and the corresponding Thouless ratios.	51
B.1	Results for additional observers in the lightness matching task. See caption of Figure 2.5 for details.	58
B.2	Results for additional observers in the lightness matching task. See caption of Figure 2.5 for details.	59
B.3	Results for additional observers in the lightness matching task. See caption of Figure 2.5 for details.	60
B.4	Results for an additional observer in the lightness matching task. See caption of Figure 2.5 for details.	61
D.1	Taken from Murray et al., 2022	65
3.1	The physical apparatus seen from participants' viewing position.	76
3.2	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.	77
3.3	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.	83

3.4	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.	89
3.5	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.	91
3.6	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.	93
3.7	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.	94
3.8	(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.	97
3.9	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).	100

3.10	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.	103
3.11	Mean Thouless ratios in Experiments 1 and 2, averaged by condition. .	104
3.12	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. As in figure 3.4, the large red dots are mean reflectance matches, the green fitted curve is the EIM model fit and the blue curve is the stimulus luminance, scaled to make a least-squares fit to the mean reflectance matches.	109
4.1	VR scenes from the participant's viewpoint. (a) The Full-Context condition. (b) The Reduced-Context condition, which showed only the reference patch and the match apparatus.	122
4.2	(a) Reflectance matches as a function of the orientation of the reference patch for a typical participant in each condition. Small red dots are reflectance matches on individual trials, and large red dots are means. Green fitted lines show least-squares fits of the EIM model. Horizontal black lines indicate the reflectance of the reference patches. Blue curves show stimulus luminance, scaled to make a least-squares fit to mean reflectance matches. (b) Mean values of the fitted ICE parameter, averaged over participants. Error bars show standard errors. The black dashed line shows the true ICE value of the simulated lighting in the VR scene.	125
4.3	Mean Thouless ratios in Full- and Reduced-Context conditions. Error bars show standard errors.	129

5.1	Thouless ratios from all studies in the dissertation. For Chapters 3 and 4, the 'left' and 'right' conditions are averaged for the relevant All-Cue and Reduced-Context conditions.	135
5.2	Integrated overview of orientation slopes across conditions from Chapters 3 and 4 in the dissertation. Again, the 'left' and 'right' conditions are averaged for the relevant All-Cue and Reduced-Context conditions. . .	135

Chapter 1

Introduction

One of the primary functions of the visual system is to decode and interpret properties of objects in our environment that hold practical significance. The visual system translates images on the retina — a pattern of light intensity modulated by physical parameters like retinal position, wavelength, and time — into perceptions that allow us to discern attributes such as distance, shape, surface colours, and textures. These attributes are crucial for tasks like navigation, object identification, and interaction with our surroundings. colour perception plays a pivotal role in this process, aiding not only in distinguishing objects with similar shapes but also in segmenting objects from their backgrounds and recognizing visual scenes (Von Helmholtz, [1925](#)).

1.1 Basic Problem of Lightness Perception

The human visual system exhibits a remarkable ability to perceive object properties across a variety of lighting conditions. For example, a white t-shirt appears white under both bright and dim light, despite the vastly different amounts of light it reflects to our eyes. *Lightness constancy* is the ability to maintain the same perceived lightness, i.e., the perceived whiteness or blackness of a surface despite large changes in lighting intensity (Von Helmholtz, [1925](#)). Lightness constancy is a special case of colour constancy where the focus is solely on the value component, which effectively reduces

the complexities of hue and saturation. Understanding lightness constancy has proven challenging, and there is still no completely successful model of lightness perception.

This introductory chapter establishes the groundwork for an in-depth exploration of lightness constancy in virtual environments. I begin by identifying the key factors that influence our perception of lightness followed by a detailed review of various concepts and models for lightness perception. I also present a summary of empirical evidence related to lightness constancy. A major goal of this dissertation is to evaluate lightness constancy in virtual reality (VR). Therefore, following the review of lightness constancy literature, I provide an overview of VR's historical development, its evolution, and its current role in research. The chapter concludes with an outline of the subsequent chapters in this dissertation, detailing the structure and specific areas of focus.

Key Terminology and Concepts

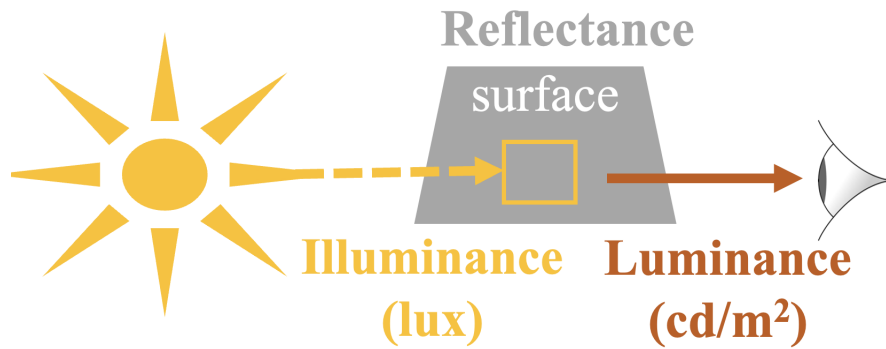


Figure 1.1: A visual representation of the concepts of illuminance, reflectance, and luminance as they interact on a surface.

Luminous flux is the power of light emitted per second in all directions, weighted by the spectral sensitivity of the human eye across the visible spectrum, which spans from 400 to 700 nm. It is measured in lumens (lm). Luminous flux per steradian is referred to as *luminous intensity*, with its SI unit being the candela (cd), defined as one lumen per steradian ¹. *Illuminance* is the total luminous flux incident on a surface per

¹A steradian can be defined as the solid angle subtended at the centre of a unit sphere by a unit area on its surface. For a general sphere of radius r , any portion of its surface with area $A = r^2$ subtends one steradian at its centre.

unit area and is measured in lux (lx), with one lux being one lumen per square meter. *Luminance*, which is the luminous flux leaving a surface per steradian per unit area, is measured in candelas per square meter (cd/m²). The luminance received by the eye from a surface is influenced by the light’s illuminance on the surface, the surface’s orientation, and its reflectance characteristics (depicted in Figure 1.1). *Reflectance* is the ratio of light reflected by a surface to the light incident upon it, and it is an intrinsic property of the surface that remains constant under all lighting conditions. As reflectance is a dimensionless ratio, it is often expressed as a percentage or a fraction and has no SI units.

The luminance L of a matte surface can be expressed as the product of its illuminance I and its reflectance R :

$$L = \frac{R \times I}{\pi} \quad (1.1)$$

The division by π normalizes the calculation to accommodate the diffusion of light over a hemisphere, which is characteristic of matte surfaces. The relationship in Equation 1.1 implies that luminance is directly proportional to both illuminance and reflectance, illustrating how different combinations of these variables can yield equivalent luminance values. For example, a low-reflectance dark gray surface under bright illumination might have the same luminance as a high-reflectance light gray surface under dimmer light.

Brightness, defined as the subjective perception of luminance, ranges from *dim* to *bright*. Lightness, which corresponds to perceived reflectance, spans from *black* to *white* (Adelson, 1993; Arend & Spehar, 1993). The distinction is significant: brightness pertains to the perceived intensity of light, while lightness concerns how much of that light a surface appears to reflect, irrespective of the illumination.

The Ambiguity Problem

The visual system faces the complex challenge of interpreting luminance information captured by the photoreceptors in the retina. This process involves distinguishing between the properties of the illuminant and the surface. The inherent difficulty in this task, known as the "inverse problem", arises because multiple three-dimensional configurations can produce the same two-dimensional retinal image, creating intrinsic ambiguity in the perception of object properties. This ambiguity presents a significant challenge in achieving lightness constancy, where the visual system must consistently perceive the lightness of surfaces despite changes in illumination that can drastically alter the reflected light.

Researchers in vision science have developed computational models to explain our perception of lightness and brightness based on retinal images, incorporating environmental influences and internal visual mechanisms (e.g., Blakeslee and McCourt (2012) and Land (1977)). These models are particularly insightful for understanding lightness constancy, demonstrating how the visual system filters and retains essential information to maintain consistent perception of an object's lightness across varying lighting conditions. Initially formulated for simple scenarios, these models now provide a foundation for exploring more complex visual phenomena related to both lightness and brightness. In the next section, I review the models that contribute to our understanding of lightness and brightness perception.

1.2 Review of Models in Lightness and Brightness Perception

Helmholtz's seminal work laid the foundation for understanding the interplay between light's physical properties and visual perception. He highlighted the importance of cognitive factors in sensory interpretation, advocating for an empirical approach that integrates physiological, cognitive, and experiential factors in lightness perception.

Despite significant research advancements since Helmholtz’s pioneering work (Von Helmholtz, 1867), a complete understanding of these perceptual attributes within the human visual system remains elusive.

Low-level visual models, essential for understanding image and light perception, focus on the processing of fundamental visual information such as edges and colour gradients. Here, I discuss these models, beginning with the basics of light and image processing and advancing towards more sophisticated interpretations of lightness and brightness perception.

Lightness Theories

This section summarizes three foundational theories that explain how we perceive lightness: the luminance ratio principle, retinex theory, and anchoring theory.

Luminance Ratio Principle

Wallach’s seminal work explored the impact of local luminance ratios on perception (Wallach, 1948). The luminance ratio principle posits that the perceived lightness of a patch is determined by its luminance relative to the surrounding area, rather than by its absolute luminance (Heinemann, 1955; Wallach, 1948, 1976). Wallach’s classical experiments demonstrated that an object’s perceived lightness remains constant if the luminance ratio between the object and its background is maintained, even when absolute luminance varies. He showed that a disk with constant luminance could appear as any shade from white to black, depending on the luminance of the surrounding annulus. This suggests that the visual system might rely exclusively on relative luminance amounts for lightness perception, eliminating the need for absolute luminance levels (Wallach, 1948).

Subsequent research by Flock and Freedberg (1970), Noguchi and Masuda (1971), and Arend and Goldstein (1987) supported Wallach’s findings and significantly extended the applicability of the luminance ratio principle to a broader range of illumination conditions, including extreme ratios over a million to one (Jacobsen & Gilchrist, 1988). This

marked a considerable expansion from Wallach's initial tests, which were limited to an 8-to-1 illumination range at most. However, luminance ratio alone cannot fully explain perceived lightness in complex scenes, particularly those involving three-dimensional shapes. For example, an issue arises when the luminance ratio varies along a curved surface of an object relative to a constant background, making the simple application of the luminance ratio principle inadequate (Pessoa et al., 1996). Furthermore, this principle struggles in scenes with varying local illumination because it cannot handle comparisons between differently lit regions effectively (Gilchrist et al., 1999). Nevertheless, Wallach's foundational work thus set the stage for more sophisticated theories of lightness perception, such as the retinex theory, which is explored in the following section.

Retinex Theory

The retinex model, developed by Land and McCann (1971), builds on the luminance ratio theory by normalizing luminance ratios within an image. This model begins by identifying points in the image where luminance changes sharply. At the core of the retinex model is a series of computations that "walk" through the image, comparing luminance values of adjacent areas. This process, known as the calculation of sequential products, is crucial for determining the reflectance ratios along the path of the walk. After identifying the brightest or whitest point in the image, the model normalizes all other luminance values relative to this point. Essentially, the model adjusts the luminance values throughout the image in relation to the brightest point, enabling it to estimate the reflectance, or the proportion of light a surface reflects compared to the total light it receives.

Figure 1.2 demonstrates the retinex model's methodology in a step-by-step manner. Initially, the image is analyzed to identify areas where luminance shifts abruptly, with these changes and their luminance values marked on the left side of the figure. The center column shows the sequential ratios calculated across these luminance edges.

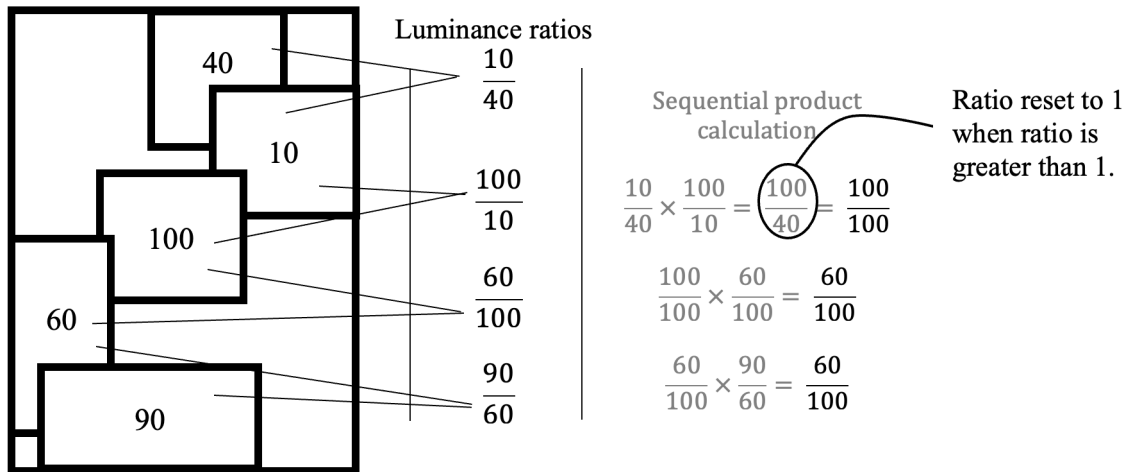


Figure 1.2: Sequential products calculated based on the retinex theory (Land & McCann, 1971).

When a ratio exceeds 1, as depicted in the right column, the sequential products are reset, allowing the model to adjust the base luminance ratio to 1, taking this as the lightest point. The normalizing step is what sets the retinex model apart from simple luminance ratio methods. By adjusting the luminance ratios relative to the whitest patch, the model inherently considers the effect of lighting, offering a perception of absolute lightness that goes beyond mere relative comparisons (Land, 1977; Land & McCann, 1971).

The Retinex model effectively addresses distinct edges, where there are sharp luminance changes. The model ignore the smooth gradients of luminance where changes are small, gradual changes in luminance. In scenarios where luminance changes gradually, the model resets, effectively ignoring these subtle transitions. Consequently, to properly handle these gradients, additional adaptations or supplementary algorithms are often required, since the Retinex model inherently does not detect gradients.

Anchoring Theory

Anchoring theory (Gilchrist et al., 1999), proposes that lightness perception can be understood by organizing an image into *frameworks*—distinct regions with approximately constant lighting. An image is first segmented into lighting frameworks.

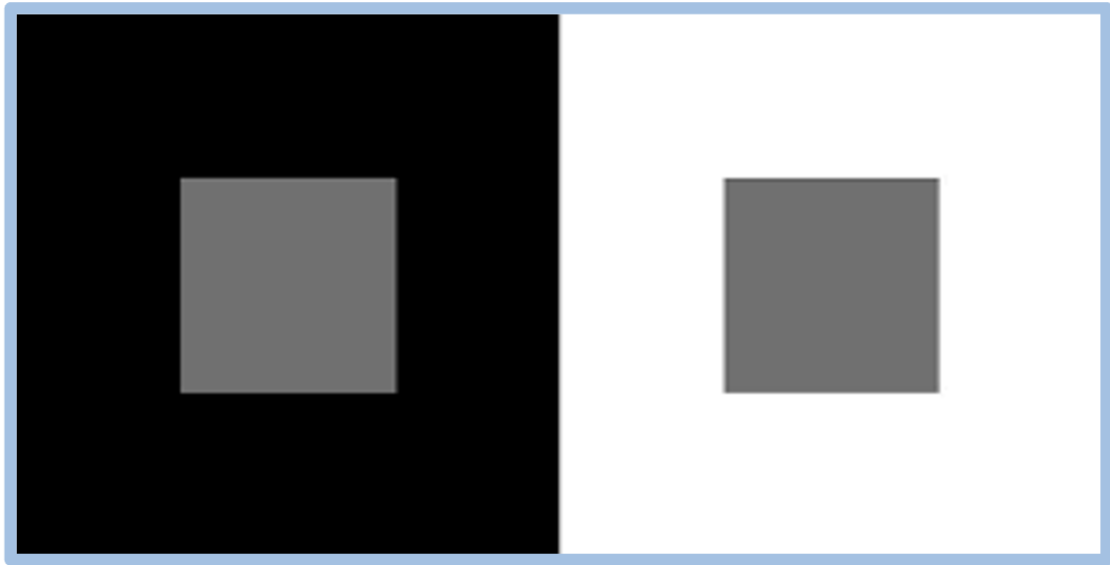


Figure 1.3: The simultaneous contrast illusion. The central gray squares are identical in luminance, but the square against the dark surround appears lighter than the one against the white surround, demonstrating the influence of surrounding luminance on lightness perception.

These are distinguished as the *global framework*, which encompasses the entire scene, and *local frameworks*, which are specific to smaller segments. Secondly, within each framework, the lightness of objects is evaluated relative to the *anchor*—the brightest area within that framework, serving as a reference point. Finally, lightness values for each patch in the image are calculated based on a weighted sum of the reflectances from local and global frameworks.

To illustrate, for the simultaneous contrast illusion shown in Figure 1.3, the anchoring theory can explain why the gray square appears lighter on a dark background than on a light one. According to anchoring theory, in the local framework with the dark background on the left side, the square is the whitest area in the framework. Conversely, the square appears darker against a lighter background on the right, with the background now being the whitest area in the framework. In the global framework, the two squares are equally gray, while the background on the left and right are the darkest and whitest areas in the framework. Anchoring Theory posits that perceived lightness is calculated as a weighted average of the lightness values in the global and local frameworks. Given

this averaging, the square against the dark background does not appear white due to the influence of the global context, but it is perceived as lighter due to its local contrast compared to the right square .

Additionally, anchoring theory incorporates the highest luminance rule and scale normalization to define the anchor point that maps stimulus luminance onto the perceived lightness scale. According to the highest luminance rule, the highest-luminance surface in an image is mapped to white, with other surfaces scaled relative to this maximum luminance following the ratio principle (Wallach, 1948). Scale normalization then dictates how the range of stimulus luminance is adjusted to the range of perceived grays. This rule suggests that the visual system modulates the perceived lightness spectrum to conform to the canonical reflectance range of 30:1—from typical white to typical black. Consequently, this adjustment leads to the expansion of the perceived lightness spectrum when the luminance range of the stimulus is less than 30:1 and to its compression when the range exceeds this standard (Gilchrist, 2006; Radonjic & Gilchrist, 2014).

Although anchoring theory offers a structured way to predict perceived lightness and is generally predictive, it does not always precisely reflect human lightness estimates (Blakeslee et al., 2009; Bressan, 2006; Bressan & Actis-Grosso, 2001; Rudd & Zemach, 2005). Notably, as a conceptual model, anchoring theory lacks a computational framework for applying it to arbitrary images, which can lead to ambiguous predictions. This gap suggests the need for more refined models or computational approaches that can consistently handle the complexities of lightness perception across varied visual contexts.

Low-level theories primarily focus on the initial stages of visual processing, emphasizing the processing of fundamental visual information such as edges and colour gradients, without accounting for changes in illumination. In contrast, mid-level theories, like the anchoring theory, address the limitations of low-level theories by incorporating more complex scene interpretations that factor in variations of illumination within a scene. Mid-level theories integrate simple image properties such as blurred edges (Shapiro & Lu, 2011), transparency (Beck et al., 1984; Metelli, 1970), and X-junctions

(Adelson, 1993) to segment scenes into different illumination zones. Within these zones, lightness perception is guided by comparisons to the local 'white' standard, which serves as a baseline for the visual system to quantitatively assess and adjust the perceived lightness of various areas (Gilchrist et al., 1999).

Brightness Theories

So far, I have focused on theories of lightness perception. Now, I will shift to discussing theories of brightness. Although these theories address different aspects of visual perception, understanding brightness is crucial for comprehensively grasping how we interpret light and colour in our environment.

Lateral Inhibition

As the luminance ratio principle is primarily associated with theories of lightness, lateral inhibition is a fundamental concept in understanding brightness perception. In the simultaneous contrast illusion, as shown in Figure 1.3, a patch of the same colour appears different depending on its surrounding colours. Cornsweet (1970) attributed this illusion to the ability of retinal neurons to suppress the activity of neighboring neurons, thereby enhancing edge contrast. This phenomenon of lateral inhibition has inspired spatial filtering models such as the high-pass filter model by Shapiro and Lu (2011) and the Oriented Difference of Gaussians (ODOG) model by Blakeslee and McCourt (1999, 2003, 2008) and Blakeslee et al. (2009). The subsequent sections will delve deeper into these models, illustrating how they build upon lateral inhibition to enhance our understanding of visual perception.

High Pass Filter Model

The high pass filter model (Shapiro & Lu, 2011) model leverages the principle that illumination changes within scenes are often indicated by low-pass image features—given that lighting usually varies smoothly except at shadow boundaries, which are typically

blurred (Land, 1977). This model selectively removes low-spatial-frequency components from images, focusing on enhancing high-frequency details that are crucial for brightness perception. While it is primarily designed to model brightness, it also has potential applications in understanding lightness, drawing on principles similar to those seen in the neural process of lateral inhibition.

Central to this model is a filter with a positive peak in the center and a surrounding square negative region, forming a center-surround receptive field, akin to the on-center off-surround receptive fields found in the ganglion cells of the eyes. High luminance in the center elicits a positive response, while high luminance in the surrounding area produces a negative response. Therefore, the filter is adept at highlighting luminance transitions in an image. A key innovation of the high pass filter model is its use of a flexible scale for the box filter, which adjusts based on the size of the test patch. This adaptability allows the filter to maintain optimal effectiveness, adjusting to cover just enough area to accurately detect changes in luminance without unnecessary overlap. This feature distinguishes it from classic lateral inhibition models that use a fixed filter size. By applying this center-surround receptive field mechanism to every pixel in an image, the model aids in the detection of luminance edges, allowing it to analyze the brightness response in the area of interest within the image.

Oriented Difference-of-Gaussian (ODOG) Model

The ODOG model, developed by Blakeslee and colleagues (Blakeslee & McCourt, 1999, 2012; Blakeslee et al., 2016), stands as a prominent model in brightness perception research. It is a low level model that is a more intricate variant of spatial filtering model like the high pass filter (Shapiro & Lu, 2011), effectively using orientation-based filters to interpret brightness. In this model, the on-center filter has a circular shape, and the off-surround filter is elliptical. These filters vary in orientation and are tuned to distinct spatial frequencies. A visualization of these filters is shown in Figure 1.4. The combination of different orientations and tunings allows the filters to detect changes in

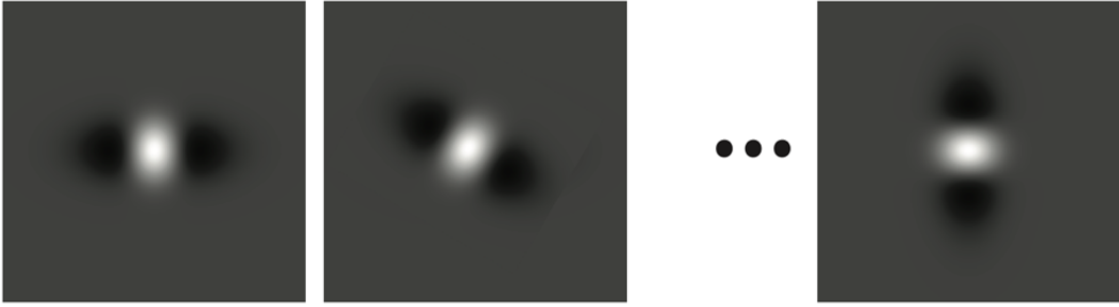


Figure 1.4: Visualization of linear filters. Taken from (Kim et al., 2018).

luminance in a similar way to receptive fields in the visual system.

The model operates by applying filters to an image, generating pixel-specific responses based on the on-center and off-surround filters. The ODOG model performs two steps to estimate brightness. Firstly, it attenuates low spatial frequencies, recognizing that higher frequencies primarily encode reflectance information, whereas lower frequencies often signify changes in illumination (Horn, 1974; Land & McCann, 1971). Second, the model performs contrast normalization across different orientations to balance the output strength from each orientation. This normalization step boosts the response from orientations with the weakest initial response to the image and decreasing the response from those with the strongest response.

Remarkably, the model effectively accounts for certain brightness phenomena traditionally attributed to mid-level factors like shadows and transparency, such as in White’s illusion and Adelson’s snake illusion (Blakeslee & McCourt, 2012). However, the ODOG model is not without its limitations. It struggles to explain some brightness illusions, pointing to the need for additional, higher-level processing in our visual system for a complete understanding (Blakeslee & McCourt, 2012; Economou et al., 2007; Kim et al., 2018).

Adaptive Filtering Model

The adaptive filtering model (Dakin & Bex, 2003) is a spatial frequency model that emphasizes the statistical properties intrinsic to natural scenes, particularly the

regularity found in their amplitude spectra. The model seeks to effectively approximate the image that likely generated the observed responses across the spatial frequency structures of the image. The origins of this property are thought to be linked to spatial correlations in both illuminant and material properties, as well as the prevalence of edges in natural scenes. Natural scenes typically exhibit a $1/f^\alpha$ spectrum, where the amplitude scales inversely with spatial frequency. This relationship is expressed as:

$$amplitude(f) = c * (1/f^\alpha)$$

Hence, the idea is that the visual system has a built-in mechanism to make all image frequencies have amplitudes proportional to the $1/f$.

The adaptive filtering model utilizes a set of scaled filters, similar to those in the primate visual system, which are calibrated to produce consistent responses across different spatial frequency channels. This consistency is achieved by adjusting the filters to match the characteristic decline of amplitude with increasing spatial frequency found in natural scenes. This approach enables the model to explain various lightness and brightness illusions, such as simultaneous contrast, White’s illusion, and the Craik–Cornsweet–O’Brien effect (Dakin & Bex, 2003). A drawback of this model is assuming uniformity in the statistical properties of natural scenes, which do not always present consistent spectral characteristics (Dakin & Bex, 2003).

Intrinsic Image Decomposition

I will now explore intrinsic image decomposition (IID), a problem relevant to this dissertation topic, explored in the field of computer vision. Although not directly a theory of human perception, IID provides insights into potential processes that could underlie human perception of lightness and colour. IID addresses the complex challenge of separating an image into its constituent layers—specifically, reflectance and illumination (Shi et al., 2017). This problem involves determining how much of the light in an image

can be attributed to the colour of the objects it shows, and how much to the lighting of the scene. This task is particularly challenging as these layers need to be extracted from a single image. Recent advancements in deep learning have significantly enhanced our ability to tackle the IID problem, providing valuable insights into image analysis. To address the ambiguity of extracting reflectance and illumination information, two main types of approaches have been developed. The first is optimization-based methods, which use physical principles as guides, and the second approach involves data-driven methods based on deep learning (Baslamisli et al., 2018; Bonneel et al., 2017; Li et al., 2020; Murray et al., 2023; J. Patel et al., 2023; Wang et al., 2023).

Early IID methods use various constraints like depth and geometry cues (Barron & Malik, 2015; Rother et al., 2011; Shen & Yeo, 2011). These constraints are usually derived from image formation models like the retinex model discussed earlier Land (1977), where IID components are defined in terms of gradients. Inspired by retinex, these models attribute strong gradient changes to reflectance and weaker gradient changes to shading (Baslamisli et al., 2018; Das et al., 2022). Other approaches use constraints based on reflectance properties, gradient distributions, textures, and shapes. To solve the intrinsic decomposition problem, researchers have also used various physically-inspired *priors* (Rother et al., 2011). However, these hand-crafted priors often struggle with complex datasets (Bonneel et al., 2017).

Some researchers use a data-driven approach, meaning the parameters are learned automatically from the dataset, potentially offering a more adaptable and accurate approach to intrinsic image decomposition (Wang et al., 2023). The field has also leveraged neural networks, leading to models that demonstrate optimal performance in colour and lightness constancy under specific conditions (Li et al., 2020; Murray et al., 2023). There have been efforts in exploring the potential of these networks as viable models for understanding human lightness perception (Murray et al., 2023; J. Patel et al., 2023). However, while these models excel in certain settings, their generalization capabilities across varied image types are still being developed. Currently, there is an

ongoing effort to enhance their ability to mimic human-like lightness perception more accurately (Li et al., 2020; Murray et al., 2023; J. Patel et al., 2023).

1.3 Evaluating Lightness Constancy: Relevant Methodologies and Empirical Findings

In this dissertation, I evaluate lightness constancy using two experimental methodologies. I will next discuss these methodologies: (1) lightness matching across a shadow boundary on a flat 2D surface, and (2) lightness matching with a reference patch at various 3D orientations in a complex scene. I will explain these tasks, review relevant literature, and summarize key findings from studies that have utilized these methodologies, providing context for the experiments conducted in this dissertation.

Lightness Matching Across 2D Shadow Boundaries

Shadows are omnipresent in our visual environments and significantly influence our perception of a scene’s layout and material properties. Lightness constancy plays a crucial role in allowing us to effectively discount shadows, enabling accurate perception of surface reflectances even under changing lighting conditions that introduce abrupt transitions, such as at shadow boundaries. This capability ensures stable perceptions of an object’s colour and brightness, regardless of lighting changes.

In the field of lightness perception, one method for assessing a participant’s ability to perceive and interpret lightness under different lighting intensities is to conduct an asymmetric lightness matching or judgment task on a flat 2D surface. In this approach, both the reference and match patches are positioned fronto-parallel to the participant (at 0 degrees orientation). Participants engage in a lightness matching or judgment task under varying illumination for each patch. For demonstration, Figure 1.5 shows a physical and a computer-generated stimulus taken from K. Y. Patel et al. (2018). The illuminance on the right-hand side is fixed throughout the experiment, whereas

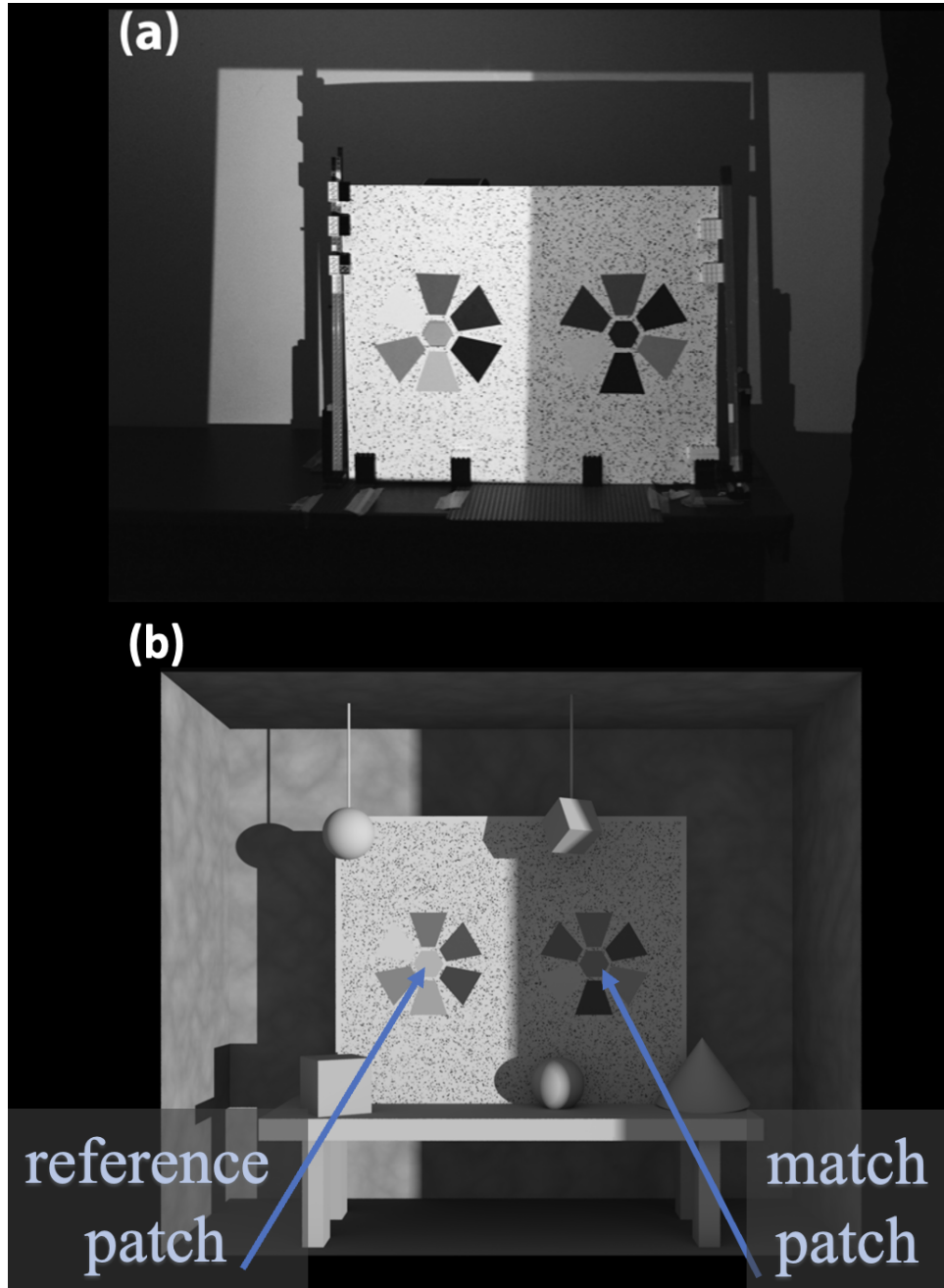


Figure 1.5: Illustration of apparatus used for a shadow boundary task. (a) physical stimulus (b) computer-generated stimulus. The images are taken from (K. Y. Patel et al., 2018)

the reflectance of the right-hand hexagon (the reference patch) varies from trial to trial. Typically, participants adjust the reflectance of the left-hand hexagon (the match hexagon) so that it appears to be painted the same shade of gray as the right-hand hexagon (the reference patch).

Empirical evidence indicates that lightness constancy can vary significantly, from poor to strong, depending on a scene’s complexity. It is generally more robust in complex scenes, while simpler ones tend to exhibit weaker constancy. Adjusting elements like the amount of contextual information in the scene (Gilchrist, 2006), or changing material properties, such as using only Lambertian surfaces as opposed to glossy ones (Boyaci et al., 2006), can greatly affect the observed level of constancy. In complex scenes, lightness constancy is often noted due to the abundance of contextual information that assists the visual system in distinguishing between changes in illumination and surface reflectance. The presence of varied cues, such as shadows, different light sources, and textures, contributes to this enhanced perceptual consistency, thereby improving lightness constancy. Conversely, in settings devoid of such cues, such as only a reference and a match patch are present in a void, participants tend to rely solely on luminance matching (Gilchrist, 1977). Notable, there is divergence from luminance matching behaviour if there are pre-existing priors or assumptions about lighting conditions influencing the participant’s perception (W. J. Adams et al., 2004). Thus there is evidence that the level of lightness constancy can be modulated by varying the complexity of the context, materials, and available lighting information in the scene.

Lightness Matching Across 3D Orientation Differences

Research suggests that the visual system uses internal representations of lighting to predict characteristics in three-dimensional space, enhancing our perception of space and surface properties. Studies such as those by Koenderink et al. (2007) and Xia et al. (2014) demonstrate a sophisticated awareness of how shading varies under different lighting conditions. For example, participants in these studies were able to adjust shading

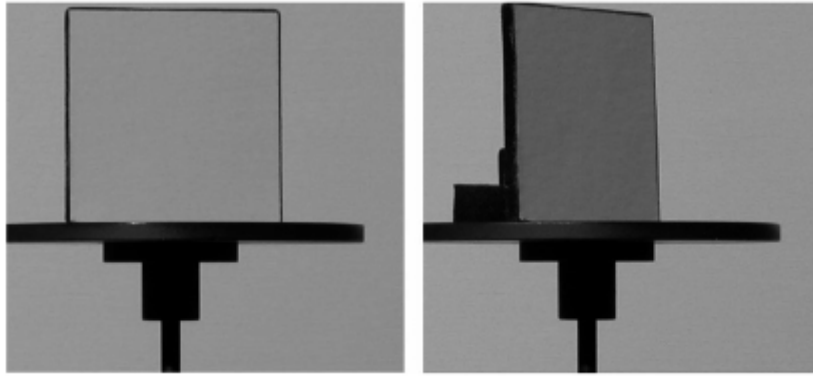


Figure 1.6: An illustration of how a change in 3D orientation of an object’s surface affects the intensity of light reflected from it. The image is taken from (Ripamonti et al., 2004).

on a spherical probe to align with overall scene lighting. Furthermore, studies by Gerhard and Maloney (2010a, 2010b) demonstrate that participants are adept at discerning the direction of a moving light source and distinguishing between luminance changes caused by the light’s movement and those that are not. These findings collectively highlight the visual system’s capacity to accurately interpret and respond to complex lighting dynamics.

The task of lightness matching across 3D orientation differences explores how lightness constant participants are when the orientation of the reference patch changes relative to the light source, rather than altering the light source’s intensity itself. In this experimental setup, the reference patch’s illumination varies as it changes orientation, while the match patch remains fronto-parallel to the participant. The participant is tasked with evaluating the lightness of the reference patch presented at various angles.

Numerous variables in a complex 3D scene affect lightness and colour perception. The arrangement and direction of lighting, as well as the orientation and properties of surface materials, interact with an object’s shape, texture, and position. These factors collectively influence the perception of light and colour within the scene. Figure 1.6 from Ripamonti et al. (2004) shows a flat and matte surface at two different orientations relative to a largely directional light source. Here we see a simple example of how in a complex scene with a largely directional light source, surface orientation affects

the luminance of the patch. Research has shown that the intensity of directional light sources impacts lightness perception and below I summarize some key findings from these earlier studies.

Early experiments used conflicting perspective and stereo cues to study the effect of perceived orientation on lightness, and concluded that orientation did not significantly contribute to lightness perception (Epstein, 1961; Flock & Freedberg, 1970; Redding & Lester, 1980). However, these early experiments used relatively simple stimuli and tasks. More recent research, including work by Gilchrist (1977, 1980) has shown that local contrast and coplanarity of test patches with background planes play an important role in lightness judgment. Overall, these more recent studies demonstrate that the geometric layout of the scene is a critical factor that can affect lightness perception. The perceived orientation of the patch, in relation to the direct light source can also influence participant’s lightness judgments. More recent studies (Gilchrist, 2006; Land & McCann, 1971; Shevell & Kingdom, 2008) and computational models (Brainard & Maloney, 2011; Maloney, 1999) have found that accurately estimating surface colour depends on estimating the chromaticity and intensity of the incident light.

The concept of an equivalent illuminant refers to the visual system’s internal estimation of lighting, originally developed for colour perception and applicable to lightness perception as well (Brainard, 1998; Brainard & Maloney, 2011; Brainard et al., 1997). The equivalent illuminant model (EIM) posits that the visual system generates an internal representation of the illumination conditions, the equivalent illuminant, to estimate the properties of light interacting with surfaces. This equivalent illuminant guides the the viewer’s estimates of colour and lightness. In practical terms, the equivalent illuminant model provides a framework for creating an index to measure lightness or colour constancy (Brainard et al., 1997). According to the model, perfect constancy is achieved when the visual system completely discounts the effect of the illumination, while no constancy implies luminance matching. EIMs quantify lightness constancy by assessing the estimated diffuseness of lighting. A detailed examination of

EIMs is presented in Chapter 3.

1.4 Lightness Perception in Virtual Reality

The pursuit of realism in VR, particularly in the accurate simulation of objects with their inherent play of light, shadow, texture, and colour, is multifaceted and demanding. Despite advancements in immersive technology, there are significant barriers in mimicking the complex interplay of visual elements that we experience in the physical world. Limitations in current hardware, the complexity of creating realistic textures, and the demand for real-time performance are a few of many challenges in achieving realism in VR (Itoh et al., 2019, 2021). Persistent research in this area is essential to bridge these gaps, especially in the context of lightness perception.

Previous studies have identified challenges in accurately replicating lightness perception in virtual environments, as rendered on computer screens, compared to real-world scenes (Morgenstern et al., 2014; K. Y. Patel et al., 2018). It remains to be seen whether VR technology, even with its heightened sense of presence and immersion (Berkman & Akan, 2019; Sanchez-Vives & Slater, 2005; Slater et al., 2009), can accurately mimic real-world experiences like lightness perception. This section begins with a brief overview of VR’s historical development and recent advances, followed by a review of related research on lightness perception in VR.

History and current uses of VR

Virtual reality (VR) has origins in the 1960s, stemming from Ivan Sutherland’s pioneering development of *The Ultimate Display* (Sutherland, 1965). Developments continued into the 1970s with head-mounted displays and flight simulators. The mid-1990s saw significant advancements, allowing for immersive VR experiences. Initially limited to research labs and military applications, VR expanded through technical innovations and commercialization (Zimmerman et al., 1986).

These earlier historical advancements provided the foundation for the evolution of modern VR systems. Since the 2010s, substantial advancements in computing power and the increased affordability of graphics processing units (GPUs) have played a crucial role in driving progress. This has led to the emergence of more accessible head-mounted displays (HMDs) and other VR hardware, opening up numerous opportunities for wider adoption and diverse applications (Cipresso et al., 2018).

VR finds applications across diverse domains, including gaming, education and training, healthcare, architecture and design, and entertainment. VR games have become increasingly popular, providing users with immersive experiences (Oyelere et al., 2020; Zyda, 2005). In education and training, VR is utilized to simulate environments and scenarios safely, enabling learners to practice and gain expertise in various fields (Radianti et al., 2020; Rojas-Sánchez et al., 2023). The healthcare industry uses VR for applications such as pain management, physical therapy, and psychological treatment (Dennis & Patterson, 2020). Architects and designers leverage VR to visualize and interact with their creations, enhancing the design process (Zhang et al., 2020). Additionally, VR contributes to entertainment, providing immersive cinema experiences and live events (He & Wu, 2022).

VR also serves as a powerful research tool for investigating human perception and behavior by manipulating visual features like lighting, texture, and colour in controlled environments (Hibbard, 2023). Researchers have utilized these technologies in diverse fields, including psychology (Wilson & Soranzo, 2015), neuroscience (Bohil et al., 2011), and human-computer interaction (Boletsis et al., 2017). For example, researchers can use VR to study the effects of different stimuli (e.g. depth, colour) on human behavior or to investigate the neural mechanisms underlying perception and cognition (Wilson & Soranzo, 2015). The use of VR in research spans over the past 25 years, resulting in a vast number of published papers and a thriving interdisciplinary community of researchers. Moreover, research conducted in VR contributes to driving the advancement of these technologies. Insights gained from studying how individuals interact with VR

environments, as well as identifying limitations and challenges in existing systems, inform the development of more advanced and effective VR technologies.

The potential advantages of VR for research are multifaceted. VR’s capability to control and manipulate visual aspects, such as lighting and texture, which are challenging in real-world setups, underscores its value as a research tool. However, even with its immersive three-dimensional properties, VR has its limitations (Chalmers & Ferko, 2008). It still grapples with simulating genuine presence and eliciting behavioral responses akin to real-world scenarios (Gerschütz et al., 2019). While there has been some research on colour constancy in VR (Gil Rodriguez et al., 2020), there is a notable gap in the literature evaluating lightness perception. This dissertation aims to address this gap by exploring the accuracy of lightness perception in virtual environments and how it compares to performance in physical environments, contributing to the ongoing efforts to enhance the realism and effectiveness of VR technology.

1.5 Current Objectives

Understanding perceptual differences between real and simulated environments is essential, particularly as virtual reality plays an increasingly significant role in areas such as training, education, and therapeutic interventions. This dissertation aims to address this under-explored area of lightness perception in real vs. simulated surfaces, focusing specifically on evaluating lightness constancy. The contributions of this work are expected to significantly advance both theoretical understanding and practical applications in the field. This following section will provide a detailed overview of the dissertation’s chapters.

Chapter 2 presents an in-depth analysis of lightness constancy across various viewing environments, including a physical apparatus with real lights and surfaces, a commercial VR headset, and a flat-panel display. The study centers on how accurate perception of achromatic surface colour is in these settings. We utilize an asymmetric

lightness matching task across a 2D shadow boundary to assess lightness constancy. The results suggest that, despite certain limitations, VR is a feasible and adaptable alternative to traditional displays and can realistically replicate real-world levels of lightness constancy under some circumstances.

Chapter 3 evaluates lightness constancy levels in an asymmetric lightness matching task, across 3D orientation differences. We use a custom-built physical apparatus and four VR conditions: All-Cue (replicating the physical apparatus), Reduced-Depth (no disparity or parallax), Shadowless (no cast shadows), and Reduced-Context (no surrounding objects). We evaluate lightness constancy using EIMs and extensively compare our findings to previous literature. Overall, the results suggests that participants had much worse constancy in VR compared to the physical condition.

In Chapter 4, we used a design similar to that in Chapter 3 but with a more realistic VR scene to assess lightness constancy. This study leverages more advanced and realistic rendering techniques, providing insights into VR’s potential to more accurately simulate real-world lightness perception. Contrary to expectations, the results show limited progress in enhancing lightness constancy.

These studies reveal that while participants can replicate real-world lightness constancy for simpler tasks such as judging lightness across a shadow boundary in VR, they exhibit weaker performance when lightness judgment involves varying a surface’s 3D orientation. It is unclear from our current theoretical understanding of lightness perception why constancy is much worse in the latter task. One might speculate that this discrepancy arises from participants’ inability to estimate orientation. However, our findings contradict this as participants show similar slant estimated across physical and VR conditions. Other potential reasons for weaker lightness constancy in VR along with suggestions for future research are discussed in the concluding chapter. This work enhances our understanding of human lightness perception and demonstrates the need for improved VR rendering techniques to more accurately simulate real-world conditions.

Chapter 2

Lightness constancy in reality, in virtual reality, and on flat-panel displays

This chapter was originally published as:

Patel, K. Y., Wilcox, L. M., Maloney, L. T., Ehinger, K. A., Patel, J. Y., Wiedenmann, E., & Murray, R. F. (2024). Lightness constancy in reality, in virtual reality, and on flat-panel displays. *Behavior Research Methods*, 1–19.

First published in Behavior Research Methods and reproduced with permission from Springer Nature

2.1 Abstract

Virtual reality (VR) displays are being used in an increasingly wide range of applications. However, previous work shows that viewers often perceive scene properties very differently in real and virtual environments, and so realistic perception of virtual stimuli should always be a carefully tested conclusion, not an assumption. One important property for realistic scene perception is surface colour. To evaluate how well virtual platforms support realistic perception of achromatic surface colour, we assessed lightness constancy in a physical apparatus with real lights and surfaces, in a commercial VR headset, and on a traditional flat-panel display. We found that lightness constancy was good in all three environments, though significantly better in the real environment than on the flat-panel display. We also found that variability across observers was significantly greater in VR and on the flat-panel display than in the physical environment. We conclude that these discrepancies should be taken into account in applications where realistic perception is critical, but also that in many cases VR can be used as a flexible alternative to flat-panel displays and a reasonable proxy for real environments.

2.2 Introduction

In recent years there have been rapid advances in virtual reality (VR) technology (Greengard, 2019; Xiong et al., 2021; Zhan et al., 2020), and VR has found a wide range of recreational and professional uses, including gaming (Linowes, 2020), cinema (Marantz, 2016), training (Xie et al., 2021), physical rehabilitation (Elor & Kurniawan, 2020), psychological therapy (Ong et al., 2022), and vision research (Hibbard, 2023; Scarfe & Glennerster, 2015, 2019). In many applications, key goals for VR design include *immersion*, the ability of the VR environment to provide rich, interactive stimuli, and *presence*, the viewer’s subjective sense of being fluidly engaged in the environment (Berkman & Akan, 2019; Sanchez-Vives & Slater, 2005; Slater et al., 2009). In some applications, an additional goal is *realism*, meaning that people and things appear to

the viewer in VR just as they would appear in the real world (Jung & Lindeman, 2021). Realism is important, for example, in some kinds of training, where the participant learns skills that must transfer to the real world, and it is also important in vision research, where VR is often used to show precisely controlled, computer-generated stimuli that are stand-ins for real objects and scenes. Interestingly, the available evidence suggests that these three goals are only loosely related, e.g., a high degree of realism may not be a prerequisite for strong presence (Jung & Lindeman, 2021; Slater, 2018)

Here we evaluate a specific but fundamental aspect of realism in VR, namely the extent to which viewers perceive black, white, and grey surface colours in virtual environments with the same accuracy as in real environments. We report achromatic colour matching experiments with a physical apparatus that uses real lights and surfaces, a VR environment designed to simulate the physical apparatus, and a flat-panel monitor that shows a rendered image of the physical apparatus. Our goal is to quantify and compare perception of achromatic surface colour in these three environments, in order to evaluate the extent to which VR and flat-panel displays can substitute for real lights and surfaces in applications where realism is important.

Lightness perception

Visual processing begins with stimulation of the 2D array of photoreceptors in the retina, but the retinal image provides only indirect and ambiguous information about useful properties of things we see (Belhumeur et al., 1999; Murray, 2021). For example, it is often useful to know the *surface colour* of things in the environment: this apple is green, that fox is light grey, and so on. The retinal image, though, depends both on the intrinsic surface colours of objects (i.e., the proportion of light they reflect at different wavelengths), and also on the spectrum and intensity of illumination in the environment. In order to perceive surface colour accurately, the human visual system must disentangle the effects of surface colour and lighting. This is a difficult computational problem that is still not well understood (Barron & Malik, 2015; Li et al., 2020), but remarkably, the

human visual system usually accomplishes it with a high degree of accuracy (Brainard & Maloney, 2011; Murray, 2021). This ability to perceive surface colour accurately in a wide range of lighting conditions and environments is called *colour constancy*.

In work on surface colour perception and colour constancy, researchers sometimes simplify the problem under study by limiting stimuli to achromatic colours: black, white, and shades of grey. In this case, the surface property of interest is *reflectance*, the proportion of incident light in the visible wavelength region that is reflected by a surface². In principle, reflectance can range from 0.0 to 1.0, but in everyday materials, black surfaces have a reflectance around 0.03, white surfaces have a reflectance around 0.9, and grey surfaces have values in between.

Reflectance is a physical property. The corresponding perceptual property, *lightness*, is defined³ as perceived reflectance (Gilchrist, 2006). For example, in a printed copy of the snake illusion (Figure 2.1), the four diamonds have the same physical reflectance, but very different lightness, as the top two diamonds appear much lighter than the bottom two. Lightness perception is a rich domain with a long history in experimental psychology; for reviews, see Gilchrist, 2006 and Murray, 2021.

Lightness perception raises many of the same problems as colour perception more generally. When we view a grey surface, the retinal light intensity depends on both the surface’s reflectance and the amount of light incident on the surface. For a matte (i.e., Lambertian) surface, the luminance (l) seen by an observer is proportional to the surface’s reflectance (r) and to the illuminance (i) incident on the surface.

$$l = \frac{ir}{\pi} \quad (2.1)$$

(The factor of π follows from the definitions of SI units for luminance and illuminance.)

²A more precise definition of reflectance takes into account the spectral distribution of the illumination and the spectral sensitivity of the human visual system. We will skirt details of photometry and colourimetry when they are not central to the discussion. For more information, see McCluney, 1994 and Fairchild, 2013.

³Alternatively, lightness is sometimes defined in terms of the perceived intensity of light reflected by a surface, relative to the light reflected by a white surface (Fairchild, 2013, p.88). We will follow the definition in the main text.

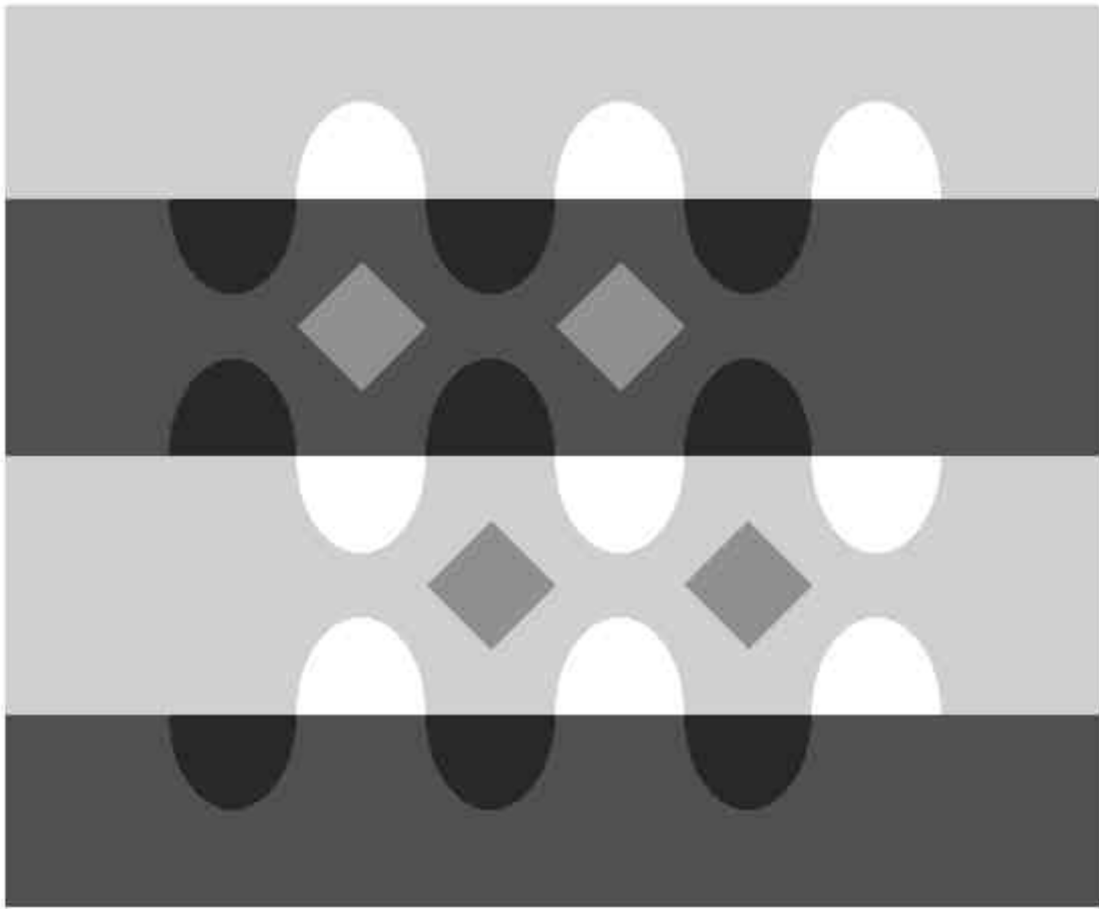


Figure 2.1: The snake illusion (Adelson, [2000](#)). The four diamonds are physically identical, but the top two appear lighter than the bottom two.

Clearly, many combinations of illuminance i and reflectance r can produce any given luminance l . Nevertheless, the human visual system is able to use contextual information and prior knowledge to estimate surface reflectance from such ambiguous luminance measurements. This ability is called *lightness constancy*.

Asymmetric lightness matching

A common experimental method for studying lightness perception is asymmetric lightness matching. In this design, the observer is shown a grey reference patch under one illuminant, and an adjustable match patch under another illuminant⁴ (Figure 2.2(a)). The observer’s task is to adjust the reflectance of the match patch so that it appears to be the same as the reflectance of the reference patch. An observer who has perfect lightness constancy will make this match setting accurately, even though the two patches are shown under different illuminants. One who has poor lightness constancy, and for example matches the luminance of the two patches instead of the reflectance, will make inaccurate match settings. Thus the observer’s match settings can be used to evaluate their lightness constancy under various experimental conditions.

One metric for quantifying lightness constancy is the *Brunswik ratio* (Brunswik, 1928). Suppose that in an asymmetric lightness matching task, the correct reflectance match setting is r_1 (indicating perfect constancy), and the reflectance setting that would result from luminance matching is r_0 (indicating a complete lack of constancy). If the observer’s actual match setting is r_m , then their Brunswik ratio is

$$\beta = \frac{r_m - r_0}{r_1 - r_0} \quad (2.2)$$

This ratio indicates where the match setting r_m lies between the two theoretically defined matches r_0 and r_1 . An observer who matches luminance has $\beta = 0$, and one who has perfect lightness constancy has $\beta = 1$. Matches resulting from partial lightness constancy

⁴This is ‘asymmetric’ matching because the reference and match stimuli are seen under different illuminants.

have $0 < \beta < 1$.

The *Thouless ratio* is a modification of the Brunswik ratio that takes into account the perceptual scaling of lightness (Thouless, 1931). Many studies have shown that lightness is a compressive function of reflectance, e.g., the perceived difference between reflectances 0.10 and 0.15 is much larger than the perceived difference between 0.60 and 0.65. In the Munsell colour system and CIELAB colour space, for example, which are attempts at perceptually uniform spaces where perceived colour differences are proportional to colour coordinate differences, the lightness dimension is approximately proportional to the cube root (a compressive function) of the corresponding physical dimension (Fairchild, 2013, pp. 80, 100). Similarly, the Weber-Fechner law states that a perceived quantity is a logarithmic function of the corresponding physical quantity (Fechner, 1948) – again, a compressive mapping. The Thouless ratio uses a logarithmic transform of reflectance, and is defined as

$$\tau = \frac{\log r_m - \log r_0}{\log r_1 - \log r_0} \quad (2.3)$$

Like the Brunswik ratio, the Thouless ratio indicates where the observer’s match setting lies between two theoretically defined reference points, corresponding to perfect constancy and no constancy. It is also similar to the ‘colour constancy index’ that indicates where an observer’s colour match setting lies relative to two reference points in an approximately perceptually uniform colour space (Arend et al., 1991; Brainard, 1998). The values of Thouless ratios found in lightness matching experiments depend strongly on stimulus properties such as the complexity of the scene, but typical values in rich, naturalistic scenes are around 0.8 (K. Y. Patel et al., 2018).

In an asymmetric lightness matching task such as the one we report below, the observer views a reference patch with a fixed reflectance and illumination, and a match patch under illumination that varies from trial to trial. The observer adjusts the reflectance of the match patch until it appears to be the same as the reflectance of the

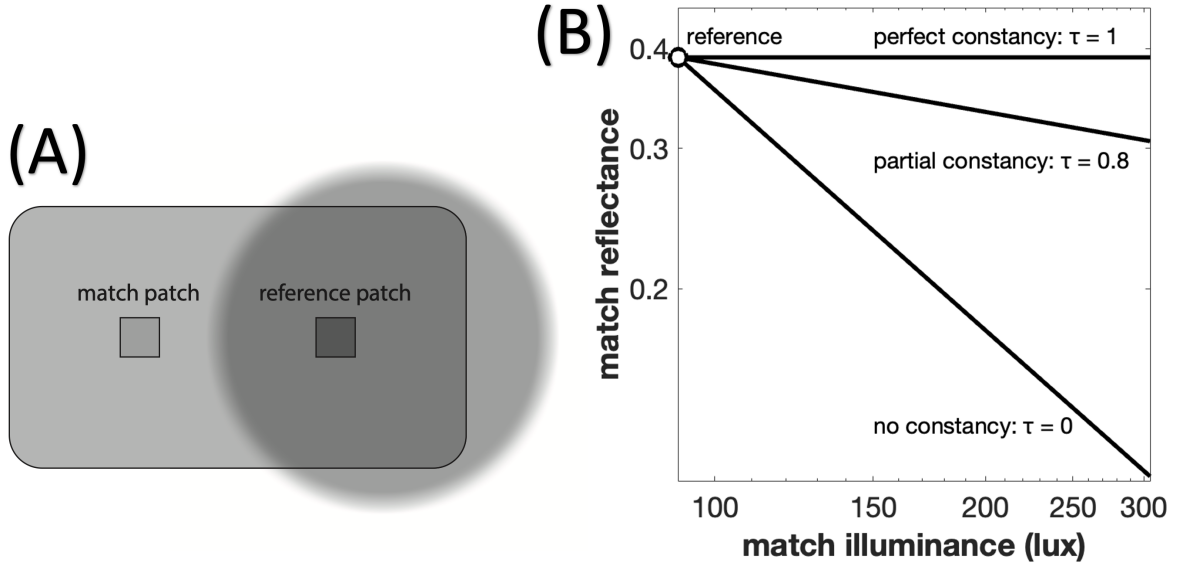


Figure 2.2: A typical asymmetric lightness matching experiment. (a) The reference patch has a fixed reflectance and illumination level. The match patch has an adjustable reflectance, and is shown under a different illumination level. The observer adjusts the match patch so that it appears to have the same reflectance as the reference patch. (b) Lightness matches for three different Thouless ratios. At high Thouless ratios (good constancy), the match reflectance is not strongly affected by the illuminance at the match patch. At low Thouless ratios (poor constancy), the match reflectance decreases substantially as illuminance increases.

reference patch. The match setting r_m can be plotted as a function of the illuminance i_m at the match patch (Figure 2.2(b)). In Appendix A we show that for an observer with a fixed Thouless ratio τ , this plot is a straight line on log-log axes, with slope $\tau - 1$. Thus such a plot gives a straightforward visual representation of the observer’s degree of lightness constancy. Two useful benchmarks are the line for perfect constancy ($\tau = 1$), with slope zero, and the line for luminance matching ($\tau = 0$), with slope -1.

Previous work

Experiments on visual perception often use computer-generated stimuli in order to maintain precise control over key image features. In fact, some experimental designs would be difficult to implement at all without computer-generated images, such as those using inconsistent lighting cues (Ostrovsky et al., 2005; Wilder et al., 2019) or inconsistent depth cues (Landy et al., 1995), or using scenes rendered without interreflections (Bloj

et al., 1999). VR extends this control further, and allows the experimenter to create immersive and interactive but well-controlled virtual environments.

But if we simulate a visual property such as surface reflectance in a computer-generated scene, does it generate the same visual percept as in a real scene? Clearly the answer will depend on the details of rendering and display methods. Previous work suggests, though, that even when care is taken to render stimuli realistically, perception of real and virtual environments can differ substantially, and realistic perception of virtual stimuli should always be a carefully tested conclusion, not an assumption. Some previous work shows that observers use qualitatively similar strategies in real and virtual environments, but with important quantitative differences. For example, Kimura et al. (2017) examined how observers determine their own orientation within real and VR environments, and found that although they rely on geometric cues and familiar landmarks in both cases, they rely more strongly on landmarks in VR. Rzepka et al. (2023) also found that when judging size, observers integrate familiar size cues differently in reality and in VR. Hartle and Wilcox, 2022 found that observers make qualitatively similar depth judgements in real and virtual scenes, but fail to achieve depth constancy in VR due to conflict between vergence and accommodation cues. Creem-Regehr et al. (2022) review work showing that observers underestimate egocentric distance in virtual environments compared to real environments.

Work on lightness and colour constancy has also revealed differences between perception of real and virtual environments. Bloj et al., 2004 used real objects and lights to examine whether observers compensate for local lighting conditions when estimating the reflectance of surface patches at various 3D orientations, and Boyaci et al., 2004 independently reported a similar study using computer-generated stimuli viewed through a stereoscope. These two studies reported qualitatively similar results, and showed that observers take into account the 3D distribution of lighting when estimating reflectance of surface patches at various 3D orientations. However, Morgenstern et al., 2014 reanalyzed their results in a way that made them quantitatively comparable, and showed that

lightness constancy was much better in Bloj et al.’s task, which used real stimuli, than in Boyaci et al.’s, which used computer-generated stimuli. This comparison had the caveat that the scenes in these two independently designed studies were broadly similar but different in many details. K. Y. Patel et al., [2018](#) also found substantial differences between lightness judgements in real and computer-generated scenes. On the other hand, Blakeslee et al., [2008](#) examined lightness judgements in real scenes and in virtual reality, using 2D Mondrian stimuli, and found similar results in both cases. Similarly, Gil Rodriguez et al., [2022](#) found good colour constancy in a complex scene shown in a VR headset, though without measuring constancy in a corresponding real scene for direct comparison. Radonjic et al. ([2016](#)) found that illumination discrimination thresholds were practically identical in a stereoscope and in a physical apparatus.

None of these previous studies compared measures of lightness constancy in real scenes and in rich VR environments. Here we carried out this comparison. We measured performance on a lightness matching task in three separate conditions, using (a) real lights and surfaces, (b) a VR headset, and (c) a flat-panel monitor, with key stimulus properties carefully matched across conditions. The real or simulated apparatus was a simple 2D panel with a single shadow boundary down the middle. We chose this configuration in order to examine a relatively simple lightness matching task (e.g., no requirement to estimate 3D surface orientation), and to test whether lightness constancy can be as good in VR as in a real environment under favourable conditions. We examined achromatic colour perception as a special case of colour constancy, where observers must overcome the ambiguity of greyscale retinal images in order to judge surface reflectances of objects being viewed in moderately complex, realistic scenes.

2.3 Methods

Participants

There were twelve participants, all of whom were unaware of the purpose of the experiment. Ten were female, two were male, and ages ranged from 19 to 35 years. 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 colour vision. All procedures were approved by the Office of Research Ethics at York University.

Stimuli

Physical environment. The observer sat in a $3.00\text{ m} \times 1.75\text{ m}$ room and viewed the experimental apparatus on a table (Figure 2.3). The front of the apparatus was an achromatic paper panel, $61\text{ cm horizontal} \times 31\text{ cm vertical}$, at a viewing distance of 71 cm (Figures 2.3 and 2.4). The panel showed a printed image with a background reflectance of 0.35 (i.e., 35%) and randomly placed circles and rectangles with reflectances ranging from 0.08 to 0.75. The panel was attached to a plexiglass backing which was supported by a metal frame. Centered in the panel were two circular apertures of diameter 2.5 cm , which were 6.5 cm apart, centre-to-centre. Immediately behind each aperture, flat against the back of the panel, was a disk of diameter 25 cm . Each disk had a circular metal backing with a circular paper printout attached to its surface. The paper showed an annulus whose reflectance ranged continuously from 0.06 to 0.80. A computer-controlled servo motor rotated each disk to adjust the part of the annulus (and hence the reflectance) seen through each aperture. Only a small area of the annulus was visible through the aperture at any time, and the range of reflectances visible due to the continuous reflectance gradient on the annulus was never greater than 6.3%, e.g., if the reflectance at one point of the aperture boundary was 0.2000, then the reflectance at opposite point was no greater than $0.2000 \times 1.063 = 0.0216$. As a result, the reflectance

gradient was not perceptible. The same poster printer and type of paper were used to create the printouts on the panel and the disks.

We measured the reflectance of parts of the apparatus 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 (see below). We calculated the reflectance of the target location as $r = 0.99 L/L_{99}$.

The apparatus and room were illuminated by overhead LED lights and a data projector. The LED lights were seven computer-controlled smart bulbs (Philips Hue, White and colour Ambiance E26, 1100 lumen; Signify N.V., Eindhoven, Netherlands) in an enclosed, spherical cloth lampshade of diameter 55 cm, located above, behind, and to the left of the observer (Figure 2.3). The colour and brightness of the LED lights were adjusted so that the 99% reflectance standard, placed at the centre of the panel apparatus, had an xy chromaticity of (0.30, 0.32) and a luminance of 37 cd/m² as measured by a spectrophotometer (SpectraScan PR-655; JADAK, North Syracuse, NY).

The data projector was located above, behind, and to the right of the observer, and directed towards the panel apparatus on the table (model CP-EX252N; Hitachi Ltd, Tokyo, Japan; Figure 2.3). We used the photometer and reflectance standard to measure the mapping from the greyscale RGB value displayed on the projector to the illuminance at the two apertures in the panel apparatus. The xy chromaticity of the projected light at the centre of the panel apparatus was (0.30, 0.32). We displayed a two-tone image on the data projector in order to create different uniform illumination intensities on the left and right halves of the panel apparatus (Figure 2.4(a)). (Specific illuminances are given below under Procedure.) The resulting illumination boundary extended beyond the panel apparatus, and created a shadow-like boundary on the table and surrounding furniture as well. We defocused the projector so that no pixelation was

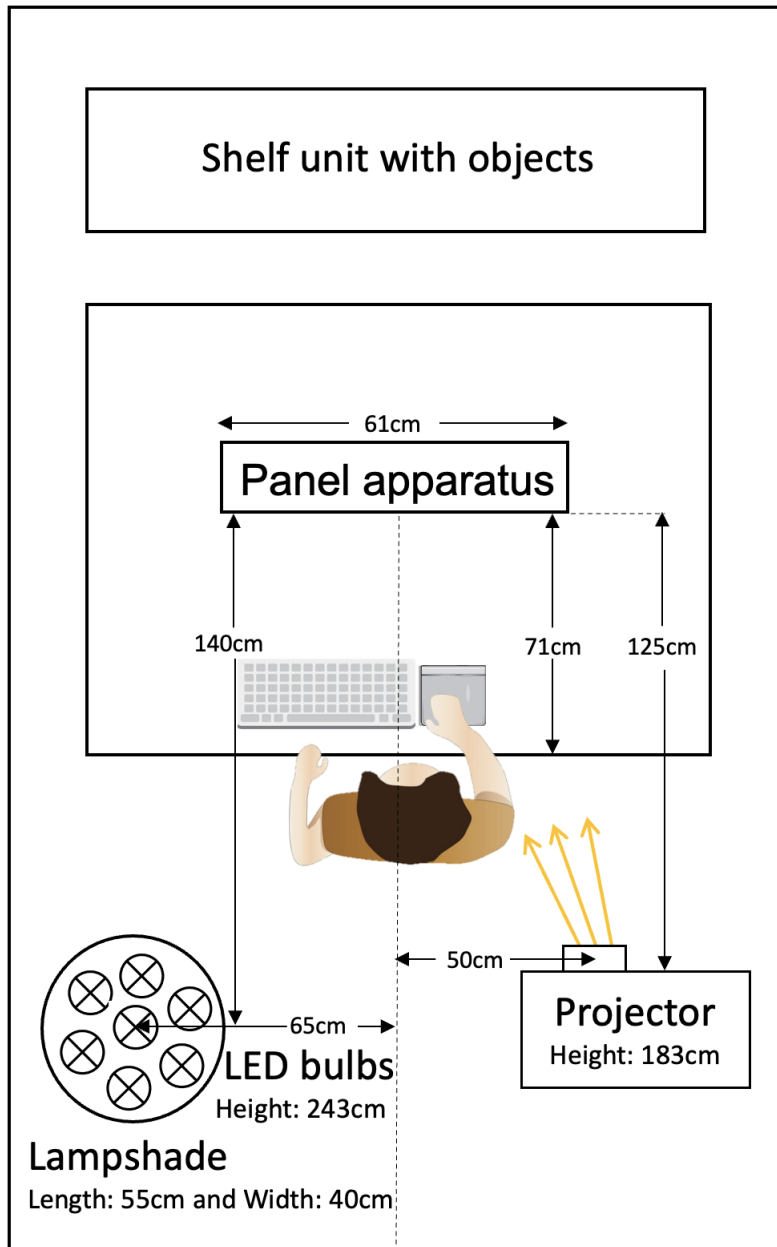


Figure 2.3: A schematic top-down view of the testing room and apparatus in the physical environment.

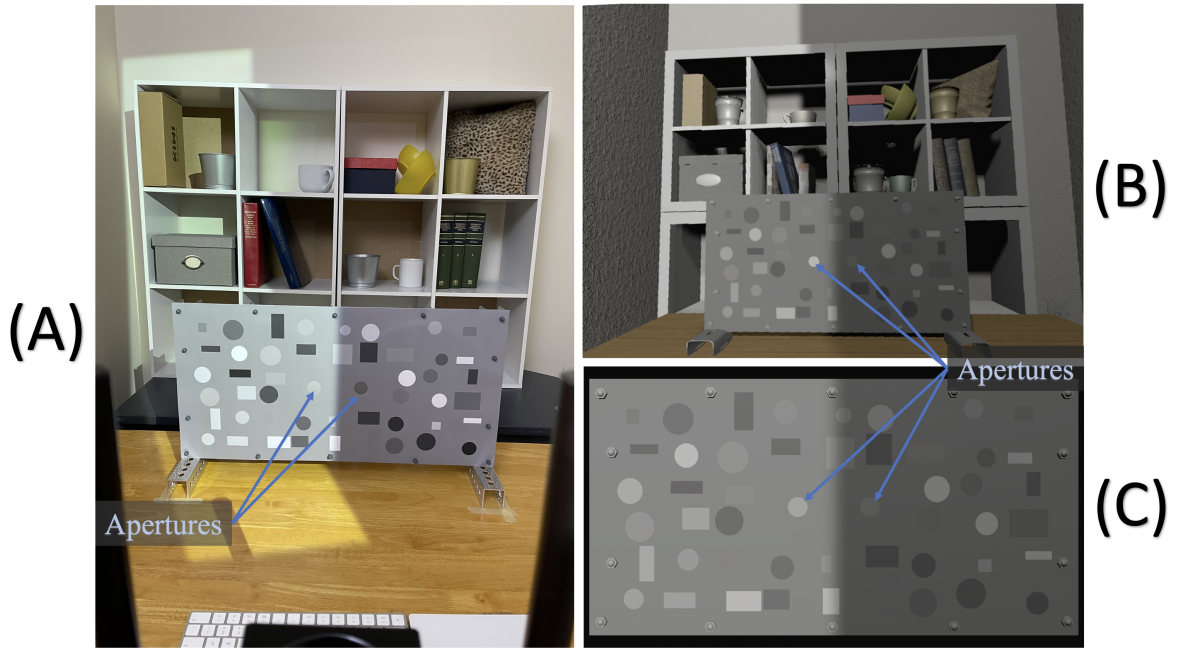


Figure 2.4: (a) Photograph of the physical room and apparatus, taken from the viewing position used by observers in the experiment. (b) A screen capture of the VR environment. (c) The panel apparatus rendered and displayed on a flat-panel monitor.

visible in the projected image, and to blur the lighting boundary, which as a result was penumbra-like and unlikely to be mistaken for a reflectance boundary.

Behind the table was a white bookshelf that displayed a range of objects, which we included to enrich the scene and provide visual information that observers could potentially use to estimate lighting conditions (Figure 2.4(a)).

VR environment. The observer viewed a scene in an Oculus Rift S headset (88° FOV, 1280×1440 pixel resolution per eye), driven by an NVIDIA GeForce GTX 1060 graphics card on a PC running Windows 10. The scene (Figure 2.4(b)) 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 2019.3.3f1). The virtual panel apparatus was carefully matched to the physical apparatus: the simulated size, position, and reflectance pattern of the virtual panel were the same as in the physical apparatus, as were the simulated size and positions of the two apertures. The panel apparatus was 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

rigorously. The virtual room contained walls, furniture, and objects that we assigned approximately the same simulated size, position, chromaticity, and material properties as their physical counterparts. Size and position were replicated by assigning the virtual objects simulated sizes that matched the measured sizes of the physical objects. Chromaticity and material properties were replicated by adjusting the virtual material properties until the virtual objects appeared approximately the same as 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.

Located behind the observer, the virtual lighting consisted of a point light source and a directional light source (Unity light types Point and Directional, with the background 'skybox' set to black). A virtual occluding surface behind the observer blocked the directional light from illuminating half of the scene, which produced a lighting boundary down the middle of the panel apparatus and surrounding furniture, as in the physical environment. The occluding surface could be moved to the left or right, depending on whether the brighter side of the illumination boundary was to be on the left or right side of the apparatus. Due to the limited luminance range of the VR headset, luminances were lower in the VR environment than in the physical environment. (Specific values are given under Procedure.) However, luminances on the virtual panel apparatus were proportional to the corresponding luminances on the physical apparatus. 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 luminances in the physical scene, to within approximately 10% error.

We calibrated the VR headset to ensure that the luminances displayed were proportional to rendered achromatic RGB values (i.e., the three integers in the range 0-255 that represent each pixel). Unity's Post-Processing Stack includes a feature called the "LUT Texture" that passes rendered RGB values through a nonlinearity specified by the user. We used the LS-110 photometer to measure the nonlinear mapping from

achromatic RGB value to physical luminance without the LUT Texture mechanism. We then used the LUT Texture to pass rendered achromatic RGB values through a compensating nonlinearity, with the result that physical luminance was proportional to the rendered achromatic RGB values. We describe this calibration procedure more thoroughly in a separate publication (Murray et al., 2022).

Flat-panel environment. Observers viewed the same virtual panel apparatus as in the VR environment, but displayed on a flat-panel LCD monitor (Dell UltraSharp 27" U2719D; Dell, Round Rock, TX; Figure 2.4(c)) in a dark room instead of a VR headset. The physical monitor was placed at a viewing distance of 71 cm, where its active area (59.5 cm $\times$ 43.5 cm) subtended the same horizontal visual angle as the panels in the physical and VR environments. The stimulus was rendered by the same Unity program used in the VR environment, with a virtual camera positioned so that the panel apparatus filled the full width of the monitor. We measured the monitor’s gamma function using the LS-110 photometer, and linearized the mapping from achromatic RGB to luminance using the same post-processing method as in the VR environment. The luminances of the various regions of the panel were proportional to the corresponding luminances in the physical and VR conditions.

Procedure

Before starting the experiment, the observer made a few practice lightness 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 paper visible through the apertures (i.e., lightness), 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 instructions from a prepared script. After the instructions and practice trials, the observer ran in the physical, VR, and flat-panel conditions in a randomly chosen order. Each observer

completed the experiment on a single day, with breaks of at least 20 minutes between the three conditions.

Physical condition. The observer sat at a table supporting the experimental apparatus, and their head position was stabilized by a chinrest. The panel apparatus described above was approximately at eye level.

On each trial, one aperture (left or right) was randomly chosen as the reference aperture, and the other was the match aperture. The illuminance on the half of the apparatus containing the reference aperture was set to 116 lx using the data projector. The illuminance on the half containing the match aperture was randomly set to one of five values: 145, 181, 227, 283, or 354 lx, also using the data projector. The reflectance at the reference aperture was randomly set to one of three values: 0.18, 0.39, or 0.55. The reflectance at the match aperture was initialized to a random value within the range displayable on the apparatus (0.06 to 0.75). A red dot, 2.5 cm in diameter, was shown 10 cm below the chosen match aperture for one second using the data projector, to indicate which was the match aperture on the current trial. A short beep then indicated the start of the trial. The observer adjusted the reflectance visible through the match aperture by swiping up or down on a trackpad. A quick stroke made a large adjustment in the match reflectance, and a slow stroke made a fine adjustment. The observer was instructed to adjust the material visible through the match aperture so that it appeared to be the same shade of grey paper as the material visible through the reference aperture. The observer 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 three reference reflectances and five match illuminances (given above). Each reflectance-illuminance combination was repeated five times with the left aperture as the reference, and five times with the right as the reference, for a total of $3 \times 5 \times (5+5) = 150$ trials.

VR condition. This condition was largely the same as the physical condition,

except that the observer viewed a virtual scene in a VR headset. Furthermore, 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. If participants moved their head too far back, they could see the green cube again, prompting them to readjust their head position until the cube disappeared. This arrangement encouraged the observer to keep their head at a fixed viewing position, while still allowing small head movements, much as with the chinrest in the physical condition.

When the observer's head was located at the position of the green cube, the virtual camera was a simulated distance of approximately 71 cm from the panel apparatus, which was the same viewing distance as in the physical condition. The simulated reference reflectances were also the same as in the physical condition. The reference and match apertures in the apparatus were assigned uniform reflectances, instead of replicating the small and imperceptible reflectance gradient at each aperture in the physical apparatus. The reference and match illuminances were lower because of the limited luminance range of the headset, but they were proportional to the values in the physical condition. The simulated reference illuminance was 51 lx, and the simulated match illuminances were 63, 79, 99, 125, and 155 lx. The observer used two joysticks on VR controllers to adjust the match reflectance. The joystick on the left controller changed the reflectance quickly, and the one on the right controller changed it more slowly. The observer pressed a button on the right controller to indicate that they had completed the reflectance match.

Flat-panel condition. This condition was largely the same as the physical condition, except that the observer viewed a rendering of the panel apparatus on a flat-panel LCD monitor. Only the panel apparatus, and no other element of the scene, was shown on the monitor, as the goal of this condition was to imitate a typical psychophysical experiment where only the primary stimulus elements are displayed,

e.g., Allred et al. (2012), Anderson and Winawer (2005), and Toscani et al. (2016). The monitor was placed on a table in a dark and otherwise empty testing room. The bookcase and objects that were behind the testing apparatus in the physical condition were not present in this condition. Head position was stabilized by a chinrest. The simulated reference reflectances were the same as in the other conditions. The simulated illuminances were higher than but proportional to those in the other two conditions: the simulated reference illuminance was 286 lx, and simulated match illuminances were 358, 447, 559, 699 and 874 lx. The observer adjusted the match reflectance using two joysticks on a gamepad, in a manner similar to the VR condition, and pressed a button on the gamepad to indicate that they had completed the match.

2.4 Results

Figure 2.5 shows two typical observers' match settings as a function of illuminance at the match patch. (Results for the remaining observers are shown in Appendix B.) The straight lines are sum-of-squares fits, constrained to pass through the points representing the reference stimuli, which are indicated by open circles. As explained above, the observer's Thouless ratio is given by one plus the slope of the fitted line, and this Thouless ratio is shown at the right of each line. In all three environment conditions, observers' match settings were well fit by straight lines, and match reflectance was only moderately affected by illuminance. (Code that produces this figure, as well as all other figures and analyses reported in the Results and Discussion sections, is provided as Supporting Information.)

Figure 2.6(a) shows Thouless ratios for all twelve observers in all conditions. Each set of three small blue dots connected by straight lines represents an individual observer, and the larger red dots represent mean Thouless ratios across observers. The mean Thouless ratio was 0.87 in the physical condition, 0.83 in the VR condition, and 0.79 in the flat-panel condition. Table 2.1 reports mean Thouless ratios in all conditions,

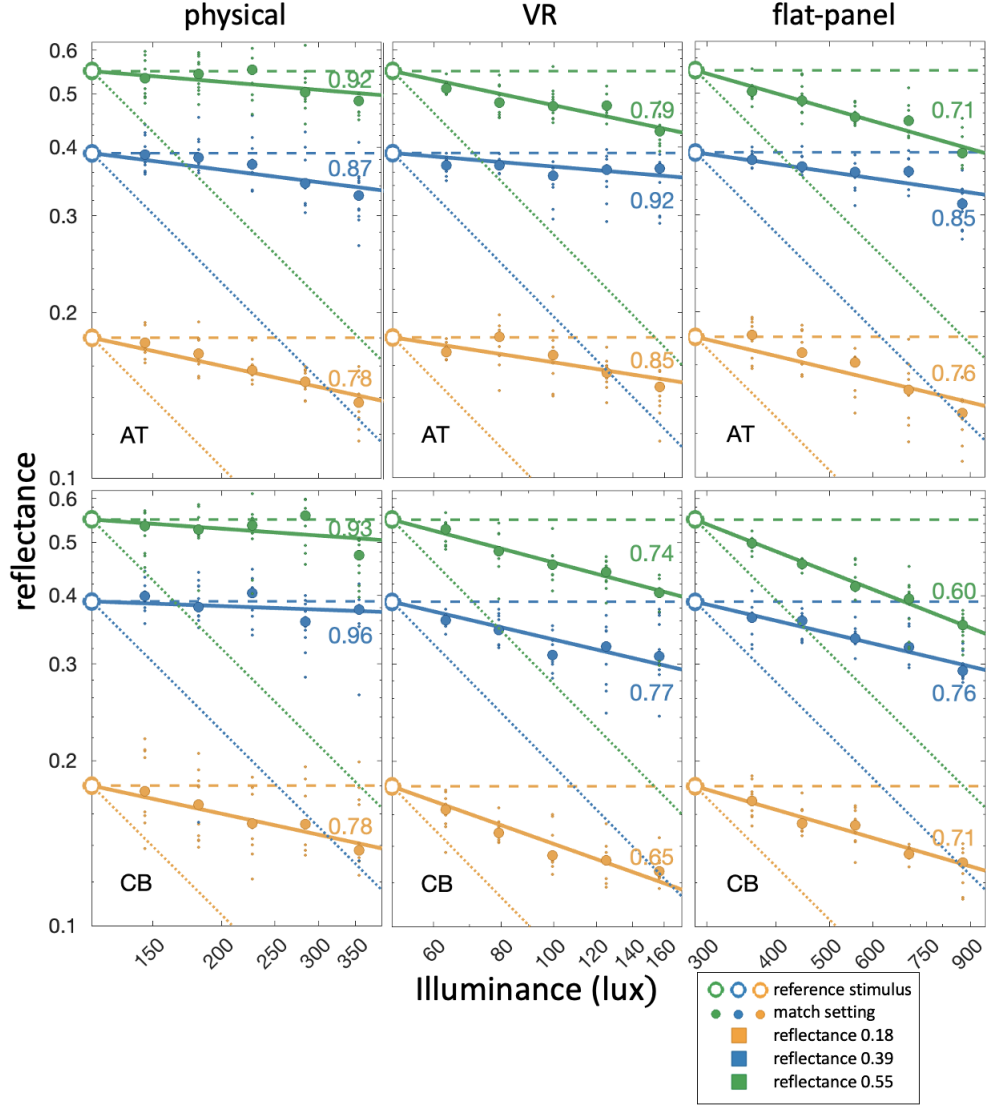


Figure 2.5: Results for two typical observers in the lightness matching task. Small dots represent reflectance matches on individual trials, and large filled dots represent the median match across trials. Straight lines are sum-of-squares fits, constrained to pass through the open circles that represent the reference stimuli. The three colours in each panel show data for the three reference reflectances. The horizontal dashed lines with slope zero represent perfect constancy ($\tau=1$), and the diagonal dotted lines with slope -1 represent luminance matching ($\tau=0$).

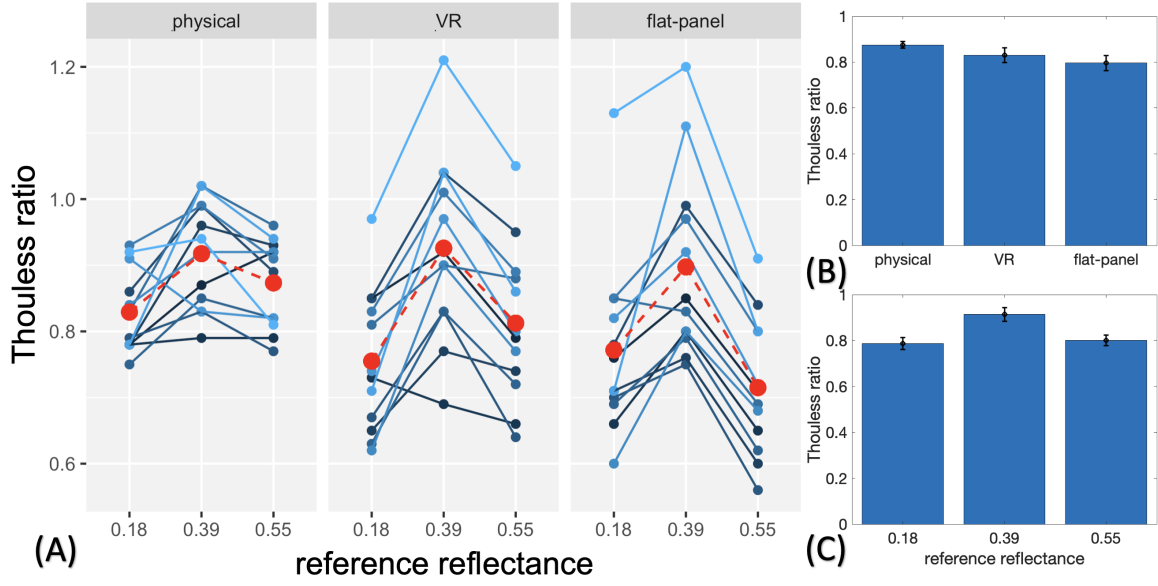


Figure 2.6: (a) Thouless ratios as a function of environment and reference reflectance. Each triad of blue dots connected by straight lines represents a single observer, and red dots represent means across observers. (b) Mean Thouless ratios by environment. (c) Mean Thouless ratios by reference reflectance.

as well as bootstrapped standard errors and 95% confidence intervals.

Figure 2.6(a) suggests that variability across observers was larger in the VR and flat-panel conditions than in the physical condition. A Levene's test for homogeneity of variance confirmed unequal variance across the three conditions ($F(2, 105) = 3.487$, $p < 0.05$). Post-hoc pairwise comparisons showed that there was a significant difference in variability between the physical and flat-panel conditions ($F(1, 70) = 5.38$, $p < 0.05$),

Table 2.1: Mean Thouless ratios by condition, with bootstrapped standard errors and confidence intervals. The mean for each environment is averaged over reference reflectances and participants, and the mean for each reference reflectance is averaged over environments and participants.

		mean	standard error	95% confidence interval
Environment	physical	0.873	0.015	0.844 - 0.901
	VR	0.831	0.032	0.772 - 0.900
	flat-panel	0.795	0.033	0.736 - 0.864
Reference reflectance	0.18	0.785	0.026	0.738 - 0.839
	0.39	0.914	0.030	0.857 - 0.974
	0.55	0.800	0.023	0.755 - 0.844

and between the physical and VR conditions ($F(1, 70) = 7.49, p < 0.05$) conditions, but not between the VR and flat-panel conditions ($F(1, 70) = 0.01, p > 0.05$).

Figure 2.6(a) also suggests that the high variability of Thouless ratios in the VR and flat-panel conditions was largely driven by individual differences, with observers tending to consistently show high or low Thouless ratios across reference reflectances and display types. To investigate this hypothesis, we plotted Thouless ratios in the flat-panel condition against the corresponding Thouless ratios in the VR condition, for each participant and reference reflectance (Figure 2.7). There was a clear linear relationship between Thouless ratios for the two display conditions. A correlation test using Bonferroni corrected significance level of $p = 0.05/3 = 0.017$ showed significant Pearson correlations between Thouless ratios in the VR and flat-panel conditions for all three reference reflectances: reference reflectance 0.18 ($r(34) = 0.85, p < .017, 95\%$ CI [0.54, 0.96]), reference reflectance 0.39 ($r(34) = 0.91, p < .017, 95\%$ CI [0.71, 0.98]), and reference reflectance 0.55 ($r(34) = 0.93, p < .017, 95\%$ CI [0.76, 0.98]).

As the Levene’s test indicated non-homogeneous variances, we compared mean Thouless ratios across conditions using bootstrap tests in order to avoid making strong distributional assumptions. We created 10^6 bootstrapped replications of the experiment, where each replication was based on twelve observers randomly selected with replacement from the twelve observers in the experiment. For each bootstrapped replication we calculated the mean Thouless ratio across observers for each of the three environments (physical, VR, flat-panel) and three reference reflectances (0.18, 0.39, 0.55). (See Appendix C for further details of the bootstrapped significance tests.) A two-tailed comparison of the bootstrapped means showed no significant difference between the physical and VR conditions ($p = 0.12$). However, the physical condition had a significantly higher mean than the flat-panel condition ($p < 0.05$), and similarly, the VR condition had a significantly higher mean than the flat-panel condition ($p < 0.001$). Furthermore, the mean for reference reflectance 0.39 was significantly greater than the means for reference reflectances 0.18 and 0.55 ($p < 0.001$ in both cases), but

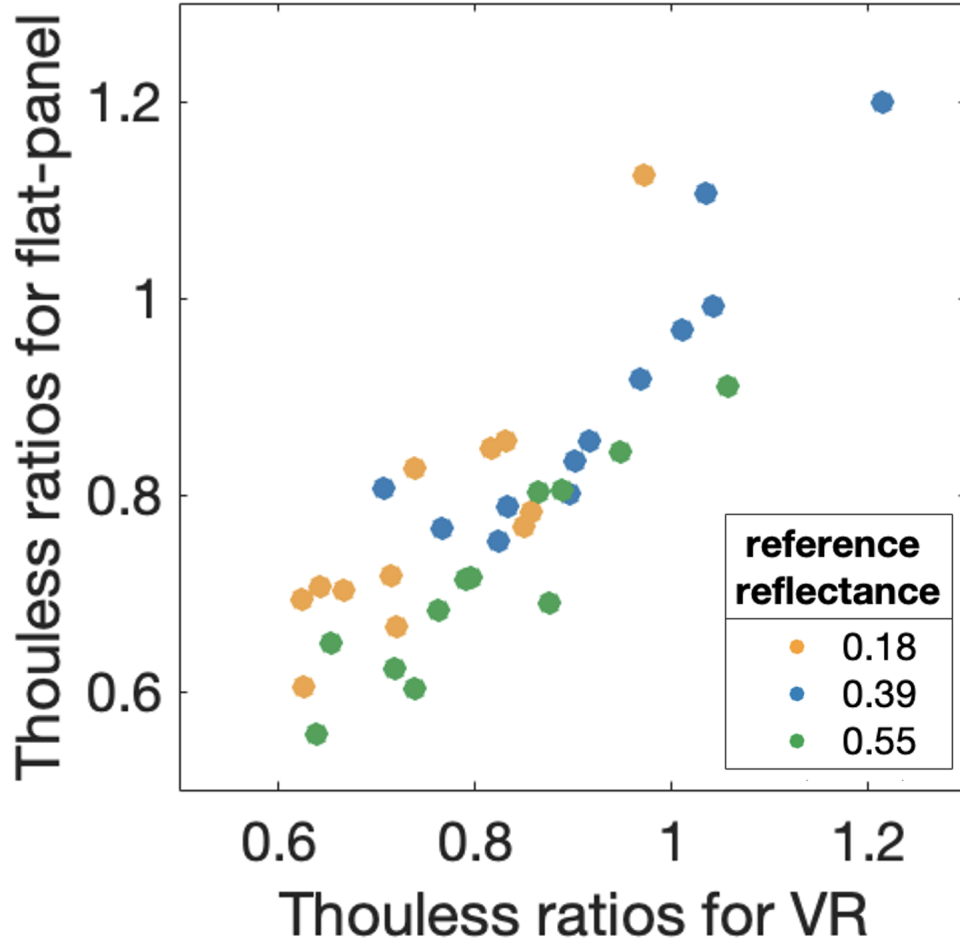


Figure 2.7: Thouless ratios in the VR and flat-panel conditions are strongly correlated.

the means for the latter two reference reflectances were not significantly different ($p = 0.41$). These p values were not corrected for multiple comparisons. However, for all but one of the comparisons that achieved significance, the results were highly significant ($p < 0.001$), and even a conservative correction, such as a Bonferroni correction (criterion $p = 0.05/6 = 0.008$), would still indicate significant differences. The one exception is the comparison of the mean Thouless ratio in the physical and flat-panel conditions ($p = 0.0137$), which a Bonferroni correction would render non-significant.

Figure 2.8 plots Thouless ratios against median response times, for each observer in each condition. The response method was slightly different in the three environments: a trackpad was used in the physical condition, and two different models of joysticks were used in the VR and flat-panel conditions. As a result, it is difficult to confidently

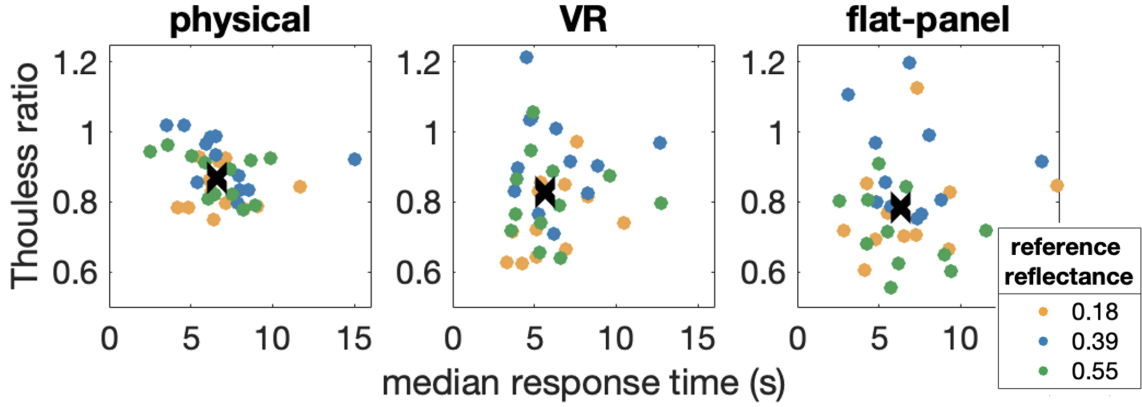


Figure 2.8: Thouless ratio versus median response time. The black 'x' indicates the median response time and Thouless ratio for each environment. The standard error for the medians was approximately the same size as the marker itself, and so are not displayed in this plot.

compare response times across conditions, but nevertheless we note that response times were broadly similar: median response time was 6.6 s in the physical condition, 5.7 s in the VR condition, and 6.3 s in the flat-panel condition. Furthermore, a correlation test showed no significant Pearson correlation between Thouless ratio and median response time in the physical condition ($r(34) = -0.26$, $p > 0.05$, 95% CI [-0.54, 0.08]), the VR condition ($r(34) = 0.09$, $p > 0.05$, 95% CI [-0.24, 0.41]), or the flat-panel condition ($r(34) = -0.04$, $p > 0.05$, 95% CI [-0.36, 0.29]).

2.5 Discussion

The goal of this study was to compare the levels of lightness constancy that observers achieve in real environments, in virtual environments, and on flat-panel displays. We found that in a simple 2D lightness matching task, observers had good constancy in all three environments, with Thouless ratios typically in the range 0.7 to 0.9. Constancy was not significantly different in the physical and VR environments. It was significantly better in the physical environment and VR environments than on the flat-panel monitor, but even there the effect size was not large, with mean Thouless ratios of 0.87 (physical) and 0.83 (VR) versus 0.79 (flat-panel).

The Thouless ratios reported in all three environments indicate good lightness constancy, compared to previous experiments with physical stimuli. For example, Gilchrist (2006, p. 31) re-examined data from Katz’s pioneering lightness constancy experiments, and found Thouless ratios between 0.35 and 0.75. Gilchrist pointed out that Katz’s apparatus provided somewhat impoverished lighting cues, resulting in lower Thouless ratios than in later studies. K. Y. Patel et al., 2018 reported Thouless ratios with a mean of 0.82 in a condition with a physical lightness matching apparatus. Similar values are also typically reported in colour constancy experiments that use a Thouless-like index to quantify constancy in rich, physical environments, such as Kraft and Brainard, 1999, who reported a mean constancy index of 0.83 in a condition where a wide range of colour and lighting cues were available.

Variability in virtual conditions. An unexpected finding in the present experiment was that inter-subject variability in lightness constancy was significantly greater with VR and flat-panel displays than with the physical apparatus. We can only speculate on the reasons for this difference, but it may have occurred because the virtual environments did not provide completely realistic or consistent cues to lighting and surface properties. For example, the Lambertian material model that we used for the simulated apparatus is only an approximate representation of real matte materials (Guarnera et al., 2016), and furthermore commercial VR displays provide conflicting cues to depth and surface reflectance. An environment with inconsistent cues may amplify the effects of individual differences in how observers integrate multiple cues to judge surface properties (Westerman & Cribbin, 1998), or individual differences in observers’ Bayesian priors for scene properties (W. J. Adams et al., 2004). We cannot be certain about this interpretation, though, and the higher variability of performance in virtual environments calls for further study.

A role for contrast polarity. Another unexpected finding was that lightness constancy was significantly better for one reference reflectance (0.39) than for the other two (0.18, 0.55) (Figure 2.6(c)). This difference was substantial: the Thouless ratio was

0.91 in the former condition, versus 0.79 and 0.80 in the latter two, which is larger than the difference between the real and flat-panel conditions discussed earlier (see Table 2.1). This effect might be explained by the fact that the reference reflectance 0.39 was just slightly higher than the reflectance of the background panel, which was 0.35. Human lightness matches cannot be completely explained in terms of contrast matching (Kraft & Brainard, 1999), but in this case, where the luminance of the reference patch was slightly higher than its surround, contrast polarity may have provided observers with a useful heuristic: they could use the constraint that the luminance of the match patch should also be slightly higher than its surround.

To illustrate this heuristic, here we calculate the match setting that observers would have made in the physical condition with reference reflectance 0.39, if their Thouless ratios had been similar to those for the other two reference reflectances. The grey background of the apparatus had reflectance 0.35, so a reference reflectance of 0.39 produced a positive-contrast reference patch. In the physical condition, the mean Thouless ratio for reference reflectances 0.18 and 0.55 was 0.80. Equation (4) shows that with reference reflectance 0.39, reference illuminance 116 lux, and match illuminance 354 lux (the highest value used in the experiment), the Thouless ratio 0.80 predicts a match reflectance of 0.31, which is a negative-contrast match setting on a background of reflectance 0.35. In fact, observers mostly made positive-contrast match settings in this condition, with a median match reflectance of 0.37. This simple calculation shows that if observers had the same Thouless ratio for reference reflectance 0.39 as for the other two reference reflectances, then under high match illuminances, they would have had to make a negative-contrast match to a positive-contrast reference stimulus. We speculate that observers were unwilling to make such a match, and instead gave the match patch a luminance slightly higher than its surround, resulting in a higher Thouless ratio. The other two reference reflectances were much higher or lower than the background reflectance, so in those cases the heuristic did not provide a useful constraint on observers' match settings.

We explored this hypothesis further by developing a simple model where an observer’s match setting was determined by both an illumination discounting stage and a penalty on match settings that produced opposite contrast polarities at the reference and match patches. We did not numerically optimize the fit of this model to our data; rather, our goal was to show qualitatively that the contrast-polarity heuristic can produce inflated Thouless ratios for low-positive-contrast stimuli, as observed in our experiment with human observers. The Supporting Information includes a MATLAB implementation of the model.

The first component of the model used equation (A.1) to calculate an initial match reflectance r_{m*} that was not subject to the penalty on opposite-polarity matches described below. This calculation assumed a Thouless ratio $\tau = 0.8$, which was the average value for human observers for reference reflectances 0.18 and 0.55. τ was therefore a free parameter of the model. To make this step more realistic as a model of human behaviour, the model added a random perturbation to the estimate. The random component was a sample from a normal distribution with mean zero and standard deviation $\sigma_1 = 0.03$, which was the median standard deviation of human observers’ match settings.

The second component of the model chose a match reflectance r_m by minimizing an objective function with two terms. The first term f_1 was the squared difference between r_m and the match reflectance r_{m*} described in the previous paragraph:

$$f_1(r_m) = (r_m - r_{m*})^2 \quad (2.4)$$

The second term f_2 penalized match settings r_m that gave the match patch a different local contrast polarity than the reference patch. That is, on trials where the reference reflectance r_{ref} was higher than the background reflectance r_{bg} , the model penalized match reflectance settings r_m that were lower than the background reflectance r_{bg} , and

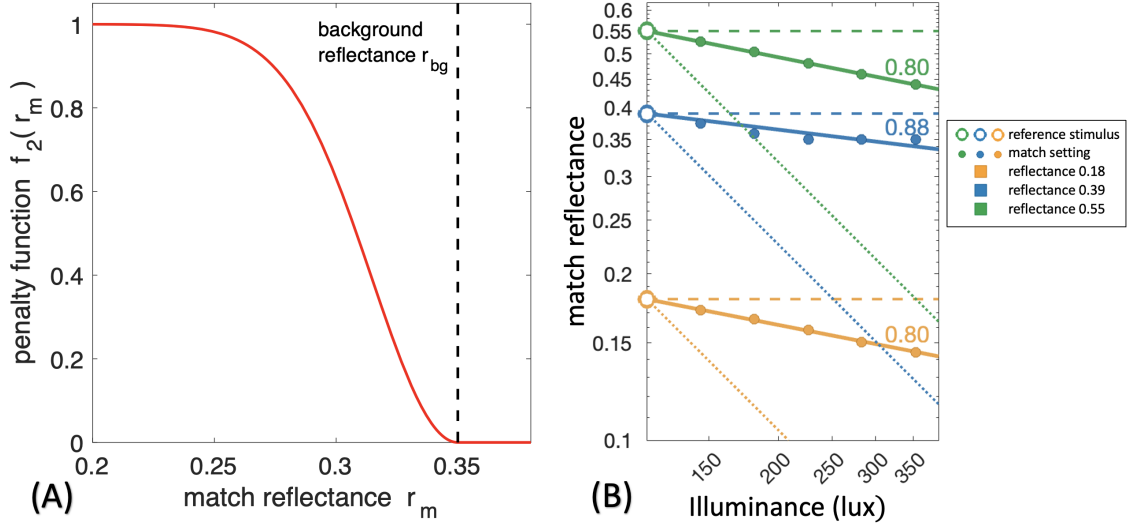


Figure 2.9: (a) The penalty function f_2 . (b) The model's median match settings and the corresponding Thouless ratios.

vice versa. The penalty term was

$$f_2(r_m) = F(\text{sgn}(r_{ref} - r_{bg})(r_{bg} - r_m), \alpha, \beta) \quad (2.5)$$

Here $F(x, \alpha, \beta)$ is the Weibull cumulative distribution function, and $\text{sgn}(x)$ is the signum function:

$$\text{sgn}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ +1 & x > 0 \end{cases} \quad (2.6)$$

We plot the penalty function f_2 in Figure 2.9(a). α is a scale parameter that determines the match reflectance at which the penalty function saturates; we used $\alpha = 0.05$. β is a shape parameter; we used $\beta = 2$. $F(x, \alpha, \beta)$ is nonzero only when $x > 0$, so the sgn factor causes the penalty function to penalize positive values of $r_m - r_{bg}$ when $r_{ref} < r_{bg}$, and to penalize negative values when $r_{ref} > r_{bg}$.

The complete objective function was

$$f(r_m) = f_1(r_m) + w f_2(r_m) \quad (2.7)$$

The weight w determined the maximum value of the contrast-polarity penalty, and we used $w = 0.1$. The model’s match setting was the reflectance that minimized the objective function f . This value varied from trial to trial, because f_1 depended on the initial match reflectance r_{m*} , which was stochastic as explained above.

We computed 1000 match reflectance settings (simulating 1000 trials) for each combination of reference reflectance and match illuminance used in the experiment with human observers. Figure 2.9(b) shows the results. The model had a Thouless ratio of 0.8 for reference reflectances 0.18 and 0.55; the penalty term had little effect in these cases, where the reference reflectances were far above or below the background reflectance of 0.35. The model had a Thouless ratio of 0.88 for the reference reflectance 0.39; here the penalty term favoured positive-contrast match settings, resulting in a higher Thouless ratio. Note that the penalty term was not a hard constraint, and 2.9(b) shows that even with reference reflectance 0.39, the model sometimes produced match settings lower than the background reflectance (0.35), as was the case with human observers as well. These Thouless ratios are practically identical to those reported above for human observers, and overall these results show that the contrast-polarity heuristic is a plausible explanation of the higher Thouless ratios that we observed for the reference reflectance (0.39) that was close to the background reflectance.

Comparison to previous studies. Blakeslee et al., 2008 also examined lightness judgements in real and VR environments. The goal of their study, however, was to examine a range of simultaneous contrast effects, not to measure lightness constancy, or to systematically compare performance in real and VR environments. Accordingly, their stimuli were simple, classic simultaneous contrast figures, and did not provide rich lighting cues. As a result, their observers reported that lightness judgements were effortful, conscious calculations, rather than spontaneous judgements of perceived surface reflectance. For these reasons, Blakeslee et al.’s results are not immediately comparable to ours, but it is still worth noting that their observers made similar lightness matches in real and VR environments, consistent with our findings.

How consistent is our main finding, namely that there was no significant difference in lightness constancy between real and VR environments, with the previous studies reviewed in the introduction? There may seem to be a discrepancy between our findings and Morgenstern et al., 2014, which found substantial differences between real and virtual environments. However, there are several possible reasons for these divergent findings. First, as noted earlier, Morgenstern et al. compared lightness constancy between two previous studies that were designed and run independently, and had many stimulus differences (Bloj et al., 2004; Boyaci et al., 2003). As a result, Morgenstern et al.’s comparison is not as reliable as a comparison between real and VR tasks as the present study.

Second, Boyaci et al., 2003, which provided data for the virtual part of Morgenstern et al.’s comparison, used a stationary, custom-built stereoscopic display, not a VR headset. It may be that the immersive, interactive VR environment in the present study contributed to good lightness constancy.

Third, and we believe most importantly, the task in our study was fundamentally different from the one in Morgenstern et al. In our study, the task was to match the reflectance of two patches on a single frontoparallel surface, surrounded by the same background reflectance, and separated by a single shadow boundary. In the studies examined by Morgenstern et al., the task was to view an isolated reference patch in a complex scene, rotated to a new 3D orientation on each trial, and to choose a lightness match from a separate frontoparallel palette of match patches in a different part of the scene. It seems plausible that the latter task was intrinsically more demanding, as it required observers to compensate for the intensity of incident light at the 3D orientation of the reference patch (which changed from trial to trial), and to compare the resulting reflectance estimate to a palette at a different orientation and location. It may be that observers achieve good lightness constancy in VR environments on simple tasks like ours, but that more demanding tasks like those examined by Morgenstern et al. reveal shortcomings of virtual environments. This is a promising avenue for future work.

Conclusion. In conclusion, we found similar levels of lightness constancy in a simple 2D lightness matching task in real and VR environments, and only slightly poorer constancy on a flat-panel display. Inter-observer variance, however, was substantially greater with VR and flat-panel displays than with the physical apparatus. This discrepancy should be considered when developing applications where realistic performance is critical, but overall our results show that VR can often be a flexible alternative to flat-panel displays, and a reasonable proxy for real environments.

Author contributions

All authors contributed to the conceptualization, design of methodology, and revision of this published paper. 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 programming and statistical analysis, revised the drafts of the manuscript, led the calibrations detailed in a separate publication (Murray et al., [2022](#)) and drafted the original section on the heuristic model explaining the impact of contrast polarity, developed in response to a reviewer’s request. Richard F. Murray also primarily constructed the physical apparatus using the robotic components, while Khushbu Y. Patel set up and matched each element of the stimulus background in Unity. Emma Widenmann and Jay Y. Patel assisted in data collection.

Appendices

Appendix A

We assume that all surfaces are Lambertian. Let the reference stimulus have reflectance r_1 and illuminance i_1 , so following equation (1) for a Lambertian surface, the luminance is $\ell_1 = r_1 i_1 / \pi$. Let the observer's match setting at the test stimulus be reflectance r_m , under illuminance i_m , and so have luminance $\ell_m = r_m i_m / \pi$. If the observer has no lightness constancy and simply matches the luminance of the reference and test stimuli, then their match setting r_0 satisfies $r_0 i_m = r_1 i_1$, or $r_0 = r_1 i_1 / i_m$. If we substitute this expression for r_0 into equation (2.3) and solve for the match setting $\log r_m$, we find

$$\log r_m = (\tau - 1)(\log i_m - \log i_1) + \log r_1 \quad (\text{A.1})$$

which is an affine function of $\log i_m$ with slope $m = \tau - 1$.

Appendix B

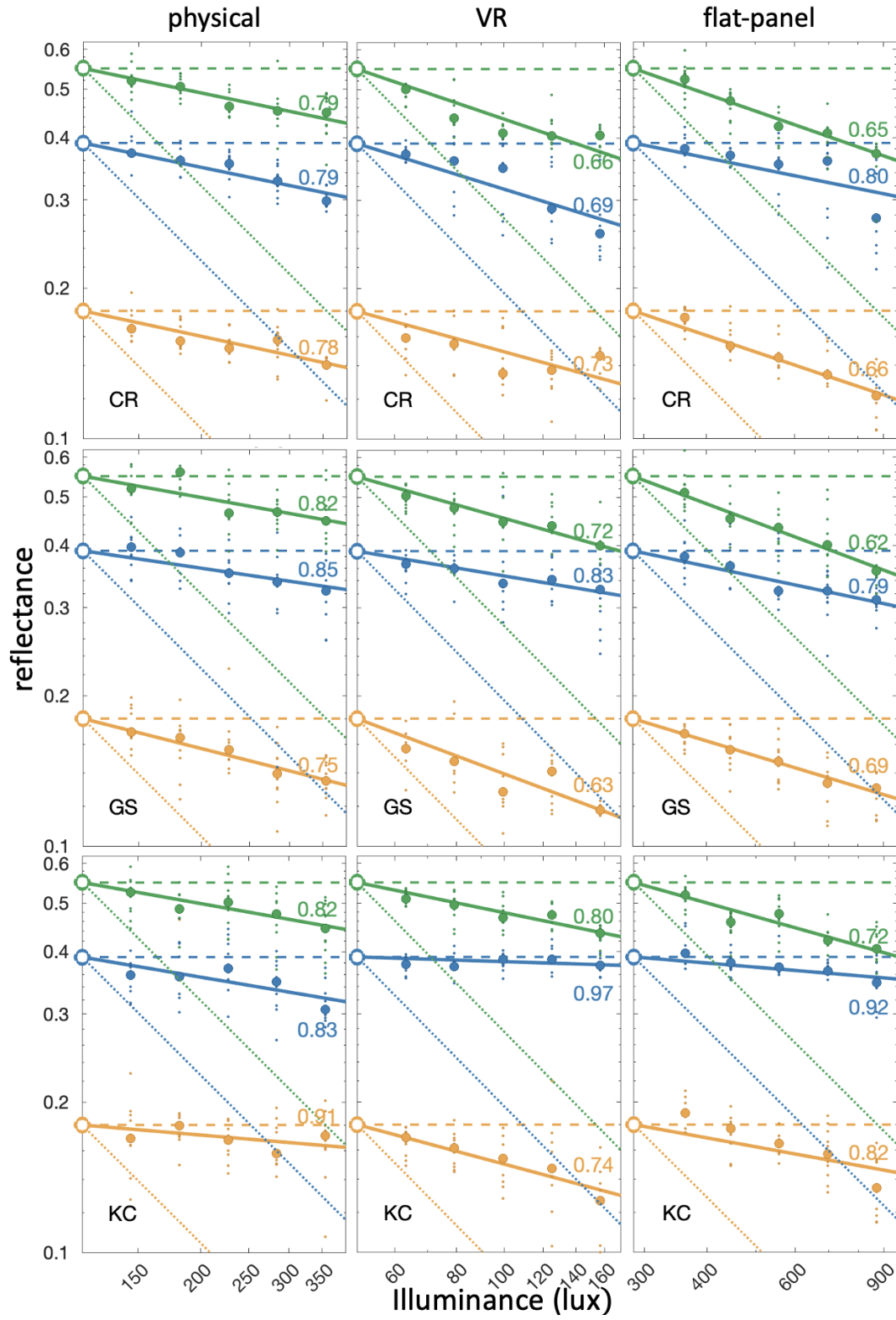


Figure B.1: Results for additional observers in the lightness matching task. See caption of Figure 2.5 for details.

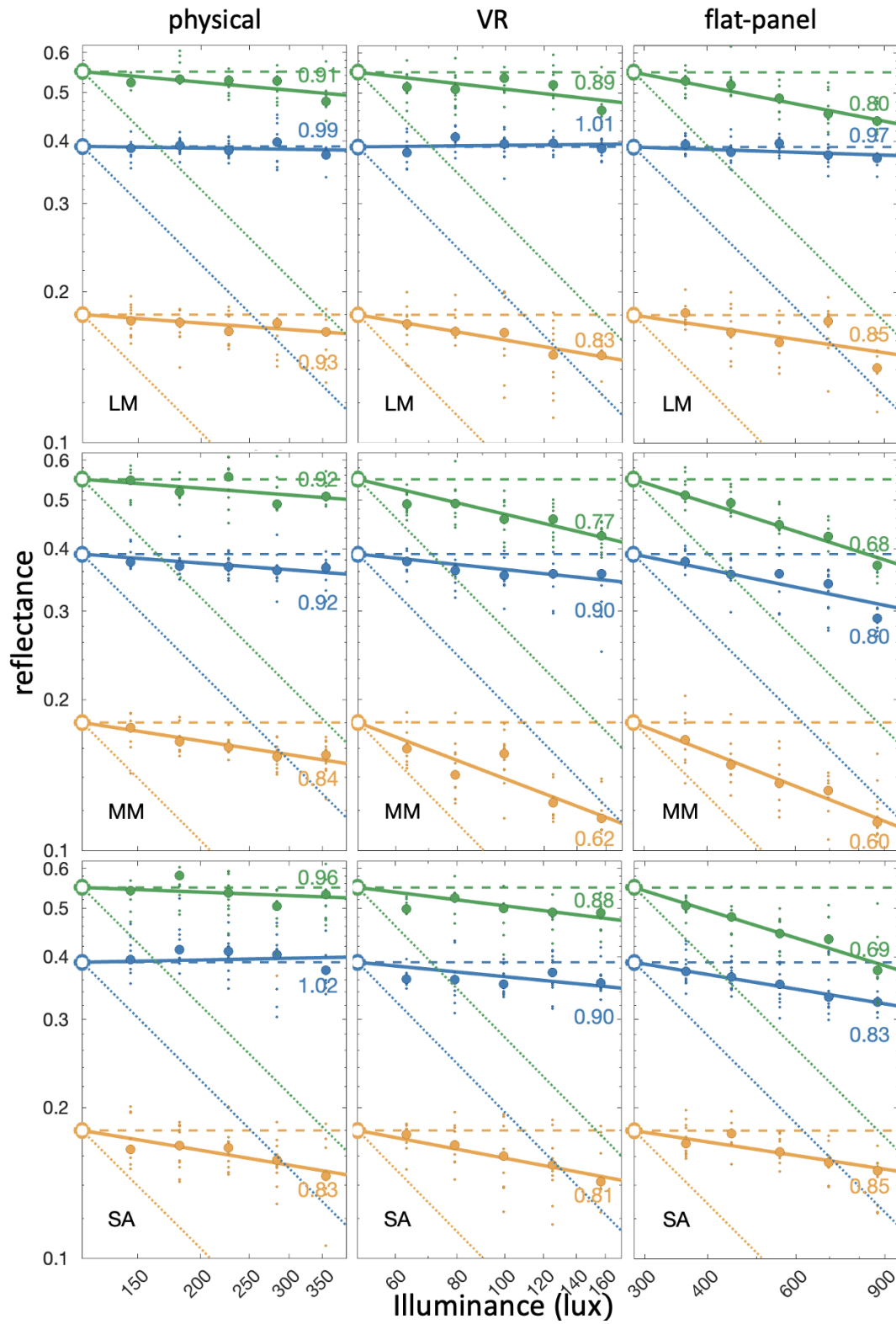


Figure B.2: Results for additional observers in the lightness matching task. See caption of Figure 2.5 for details.

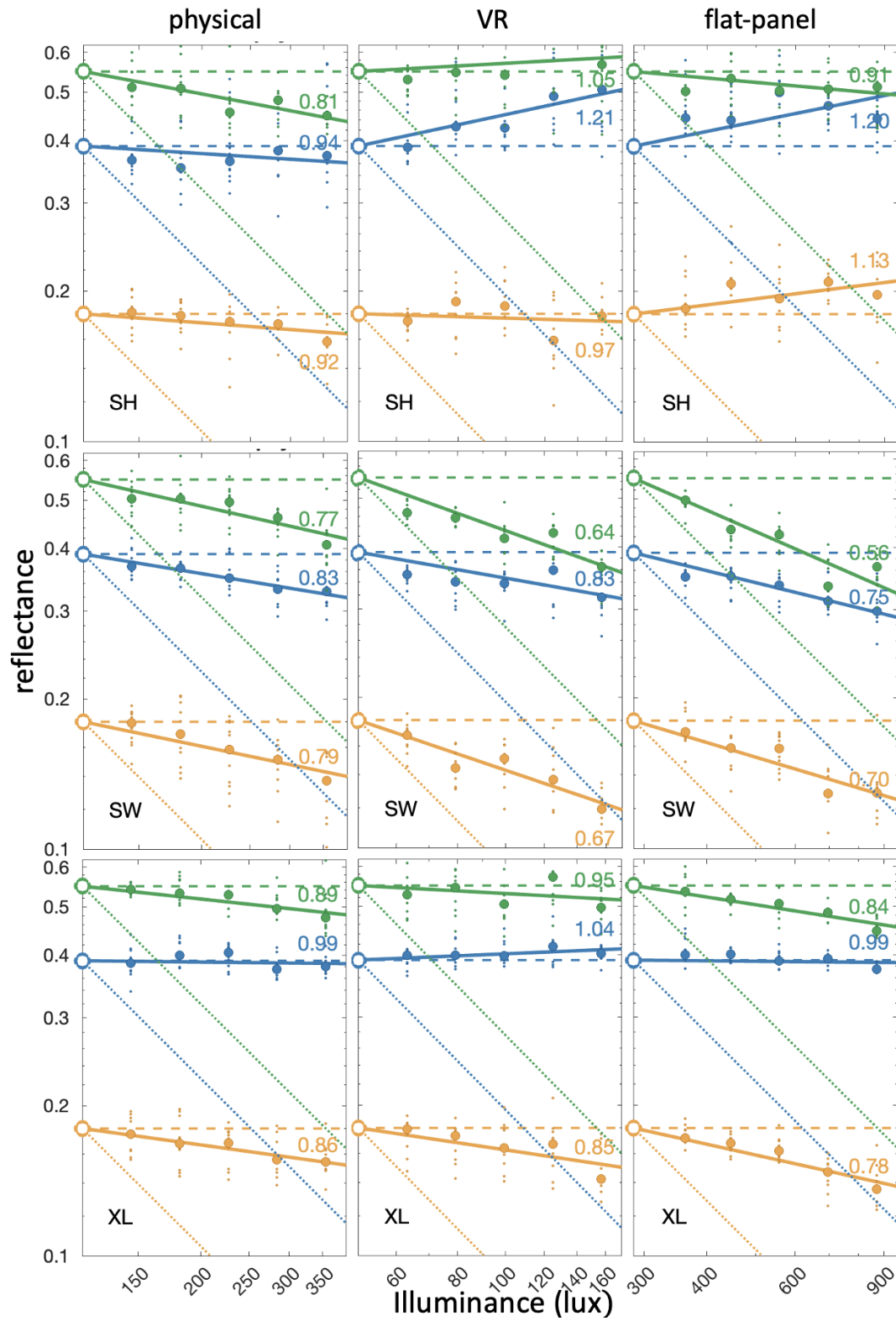


Figure B.3: Results for additional observers in the lightness matching task. See caption of Figure 2.5 for details.

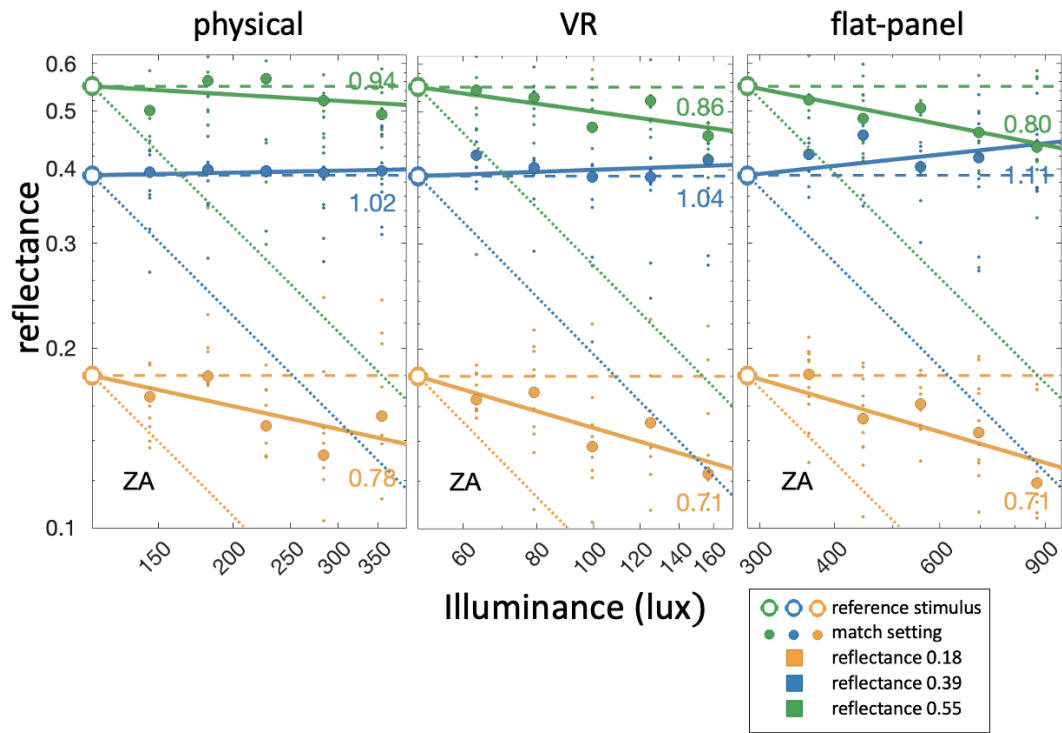


Figure B.4: Results for an additional observer in the lightness matching task. See caption of Figure 2.5 for details.

Appendix C

Here we provide details of the bootstrapped significance tests of mean Thouless ratios.

Figure 2.6 shows Thouless ratios for each display method, reference reflectance, and observer. This data can be represented as a $3 \times 3 \times 12$ matrix T_{ijk} , where each entry is the Thouless ratio for display method i (an integer from 1 to 3), reference reflectance j (also an integer from 1 to 3), and observer k (an integer from 1 to 12). The red dots in Figure 2.6 show mean Thouless ratios across observers, which can be represented as a 3×3 matrix M_{ij} , where each element is the average of T_{ijk} for k from 1 to 12.

On each bootstrap iteration, we simulate a repetition of the experiment by creating a new $3 \times 3 \times 12$ matrix $T_{ijk}^{(b)}$ of Thouless ratios. We generate this matrix by choosing twelve observers with replacement from the twelve observers in the experiment. That is, each 3×3 slice of $T_{ijk}^{(b)}$ for a given value of k is a randomly chosen 3×3 slice of the original data T_{ijk} , representing the Thouless ratios of a single observer. (“With replacement” means that the same observer may appear more than once in the resampled matrix $T_{ijk}^{(b)}$.) We then calculate the bootstrapped 3×3 matrix of means $M_{ij}^{(b)}$ by taking the average of $T_{ijk}^{(b)}$ over k from 1 to 12.

We repeat this sampling procedure $B = 10^6$ times, producing B simulated repetitions of the experiment, represented as $T_{ijk}^{(b)}$ and $M_{ij}^{(b)}$, where b ranges from 1 to B .

This resampled data forms the basis of the bootstrapped significance tests. For example, to test whether the mean Thouless ratio in condition $i = 1$, $j = 1$ (say,

the physical condition and reference reflectance 0.18) is significantly greater than the Thouless ratio in condition $i = 1$, $j = 2$ (say, the physical condition and reference reflectance 0.39), we find the proportion of bootstrapped samples for which $M_{11}^{(b)}$ is greater than $M_{12}^{(b)}$. If this is true for at least 95% of the samples, then we conclude that the first mean is significantly greater than the second mean at a significance level of $p < 0.05$.

For additional details of bootstrapping methods, see Efron and Tibshirani, [1994](#). We provide MATLAB code that implements these significance tests as Supporting Information.

Appendix D

In addition to the summary provided in our main manuscript, here is a clearer, step-by-step guide on how we calibrated the VR headset, detailed in our earlier work (Murray et al., [2022](#)):

1. **Check the Project Settings:** Ensure that Unity is set to Gamma mode, which is often the default. If not, switch to Gamma mode.
2. **Measure Initial Luminance:** Use an LS-110 photometer to measure the luminance output for various shades of gray from black to white. Initially, these measurements might display a curve, indicating nonlinearity, similar to Figure [D.1\(a\)](#).
3. **Create a LUT Correction File:** Based on the initial luminance measurements, create a Look-Up Table (LUT) texture file. This file will help correct the display to ensure that the luminance displayed is proportional to the rendered achromatic framebuffer (RGB) value.
4. **Configure Post-Processing Stack:** Apply the LUT texture file in Unity's settings under the Color Grading section of the Post-Processing Stack to correct the display.
5. **Confirm Calibration Accuracy:** Re-measure the luminance output by the headset across the same range of achromatic RGB values. The new measurements

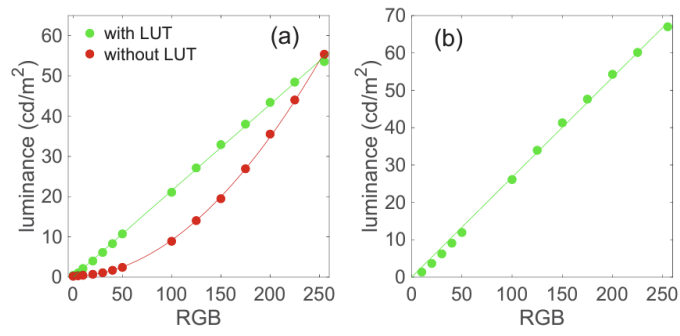


Figure 3. (a) Luminance versus achromatic RGB, with and without a linearizing LUT texture. The fit to the data without a LUT is a gamma function. The fit to the data with a LUT is a straight line constrained to pass through the origin. (b) Linearization confirmed with measurements from a pinhole camera.

Figure D.1: Taken from Murray et al., [2022](#)

should form a straight line, indicating that the display's luminance now accurately corresponds to the intended gray levels, similar to Figure D.1(b).

Appendix E

At the request of the evaluating committee, here we report the effect sizes of the most important comparisons of mean Thouless ratios across conditions. The comparison between the mean Thouless ratios in the Physical and VR conditions had a Cohen's $d = 1.78$, which is a large effect size, although it was not statistically significant ($p = 0.12$). In contrast, the differences in mean Thouless ratios between the Physical and Flat-Panel conditions not only had a very large effect size ($d = 3.14$) but were also statistically significant ($p = 0.01$), indicating a robust effect that is highly unlikely to be due to chance. Furthermore, the comparison between VR and Flat-panel displays yielded a large effect size ($d = 1.07$) and a significant difference p -value ($p < 0.001$), indicating strong and reliable differences between these conditions.

Chapter 3

An Equivalent Illuminant Analysis of Lightness Constancy with Physical Objects and in Virtual Reality

This chapter has been submitted to Behavior Research Methods and is currently under review (BR-Org-24-359):

Patel, K. Y., Wilcox, L. M., Maloney, L. T., Ehinger, K. A., Patel, J. Y., Singh, S., & Murray, R. F. (submitted:BR-Org-24-359). An equivalent illuminant analysis of lightness constancy with physical objects and in virtual reality. *Behavior Research Methods*.

3.1 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 fronto-parallel 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: All-Cue (replicating the physical apparatus), Reduced-Depth (no disparity or parallax), Shadowless (no cast shadows), and 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 cues to lighting conditions. 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 reasons for this finding.

3.2 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 et al., 2022; Rensink & Cavanagh, 2004), cast shadows (H. Adams et al., 2021), image contrast (Boyaci et al., 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 et al., 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*)⁵. In complex, realistic environments, viewers are able to

⁵For more precise definitions of these photometric terms, see McCluney, 1994.

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 (K. Y. 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 et al., 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 et al., 2022).

Lightness and color perception are computationally difficult tasks that require viewers to disentangle the effects of lighting and surface reflectance. Estimating re-

flectance requires taking into account factors such as surface orientation, material properties, and the lighting distribution in a scene (Boyaci et al., 2006; Doerschner et al., 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 et al., 2008 found similar levels of lightness constancy using simple 2D stimuli in VR and in a physical apparatus constructed with paper and lights. K. Y. 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 Rodriguez et al., 2022 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, Radonjic 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. K. Y. Patel et al., 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).

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:

$$L = \frac{\rho}{\pi} [D \max(\mathbf{n} \cdot \mathbf{s}, 0) + A] \quad (3.1)$$

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:

$$\rho = \frac{\pi L}{D \max(\mathbf{n} \cdot \mathbf{s}, 0) + A} \quad (3.2)$$

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 (3.2) 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 et al., 1999, 2004; Boyaci et al., 2004, 2006; Boyaci et al., 2003; Doerschner et

al., 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 fronto-parallel plane. (This fronto-parallel 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.

Shadows can have a strong effect on lightness (Adelson, 1993; Boyaci et al., 2006), as can binocular disparity and motion parallax (Kitazaki et al., 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 (3.2) 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.

3.3 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 $\times$ 1.75 m room and viewed the apparatus on a table, from a viewing distance of 74 cm (Figures 3.1 and 3.2). The equipment had two components: the reference apparatus and the match apparatus.

The reference apparatus was located on the left side of the table (Figure 3.1). It consisted of a reference patch and several geometric objects attached to a wooden frame. The reference patch was a 3 $\times$ 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

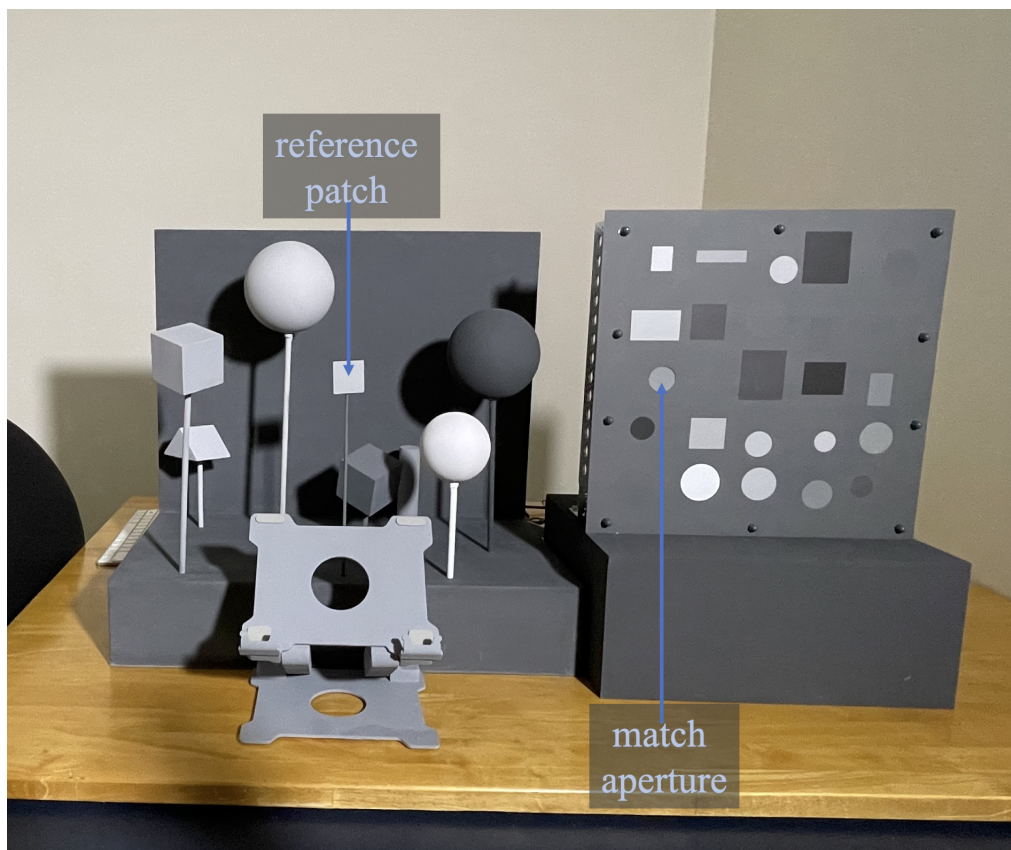


Figure 3.1: The physical apparatus seen from participants' viewing position.

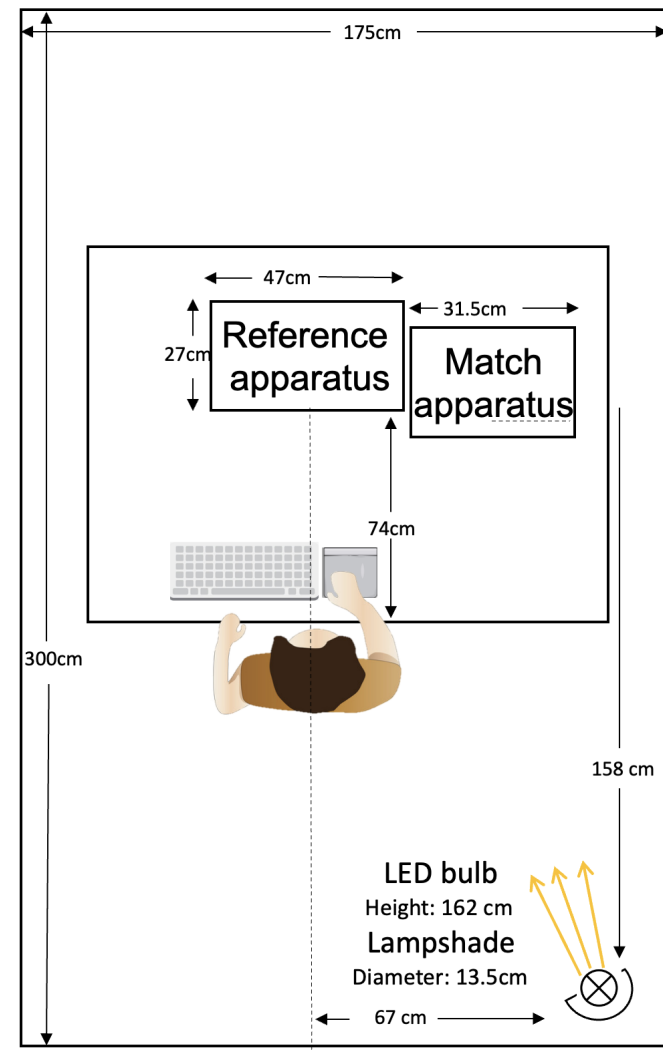


Figure 3.2: A schematic top-down view of the physical room and apparatus. The match apparatus was placed slightly in front of the reference apparatus, so that the reference and match patches were equally distant from the participant.

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.

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 3.1). 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 3.2). 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 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 3.1). 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 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 3.1. 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 3.3) 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. Chromaticity and material properties were replicated by adjusting the virtual material properties until the virtual objects appeared approximately the same as 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.

Due to the limited luminance range of the headset, luminances were lower in the VR environment than in the physical environment (specific values are given

under Procedure.) However, luminances on the reference and match apparatuses 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 et al., 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 3.3 (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. In the Shadowless condition, no cast shadows were rendered (Figure 3.3 (b)). In the Reduced-Context condition, only the reference patch and the front panel of the matching apparatus were visible (Figure 3.3 (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

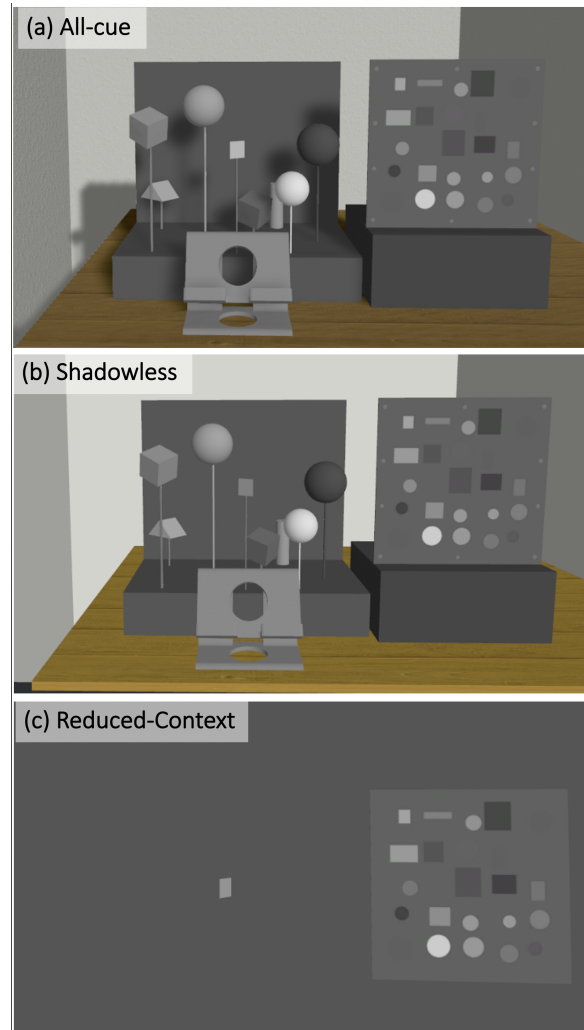


Figure 3.3: Screen captures of VR scenes from the participant’s viewpoint. (a) The All-cue condition, modelled after the physical testing room. The Reduced-Depth condition appeared the same, but was configured for zero disparity and no motion parallax. (b) The Shadowless condition had no shadows. (c) The Reduced-Context condition, which showed only the reference patch and the match apparatus.

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 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 (3.2) 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 (3.2):

$$\rho_h = \frac{\pi L k}{D_h \max(\mathbf{n}_h \cdot \mathbf{s}_h, 0) + A_h} \quad (3.3)$$

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 (3.1) and (3.2), 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. The perceived surface normal can 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 $\mathbf{s}_h$:

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

Here we assume the participant is located on the $+x$ -axis, looking towards the origin, and the $+z$ -axis is upwards. The dot product in equation (3.3) 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 (3.3) 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,

$$\rho_h = \frac{\pi L k}{D_h \max(\cos \phi_s \cos(\psi_n - \psi_s), 0) + A_h} \tag{3.4}$$

$$= \frac{L}{\alpha(\max(\cos(\psi_n - \psi_s), 0) + \beta)} \tag{3.5}$$

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 $-90^\circ \leq \phi_s \leq 90^\circ$, to move this factor outside the max function. The main parameters of interest in equation (3.5) 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 (3.5) 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 ϕ_s 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:

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

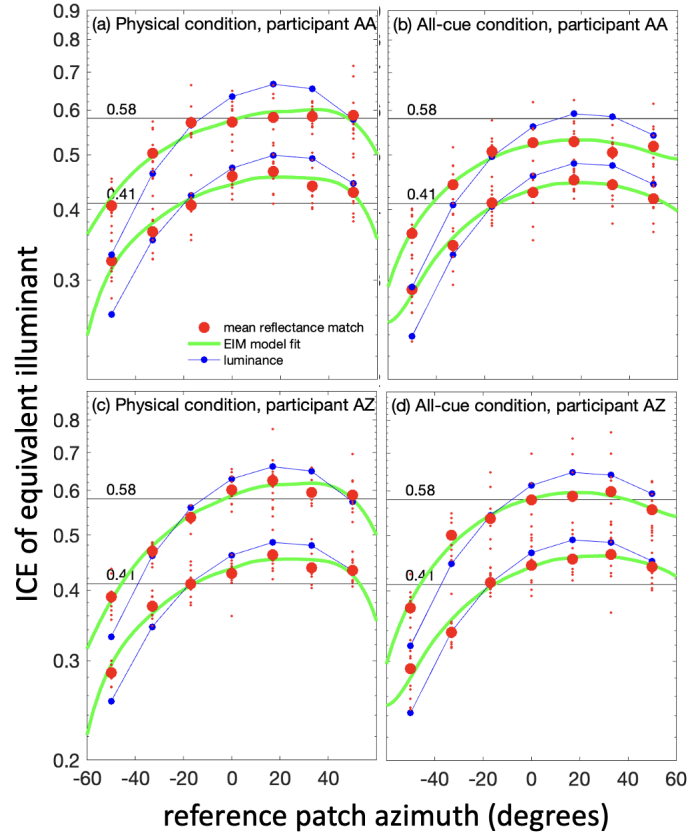


Figure 3.4: Reflectance matches as a function of the azimuth of the reference patch for two typical participants in the Physical and All-Cue conditions. The small red dots are reflectance matches on individual trials, and the large red dots are means. The green fitted lines show the EIM model fit to 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.

Results

Lightness matches

Figure 3.4 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 3.5 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-

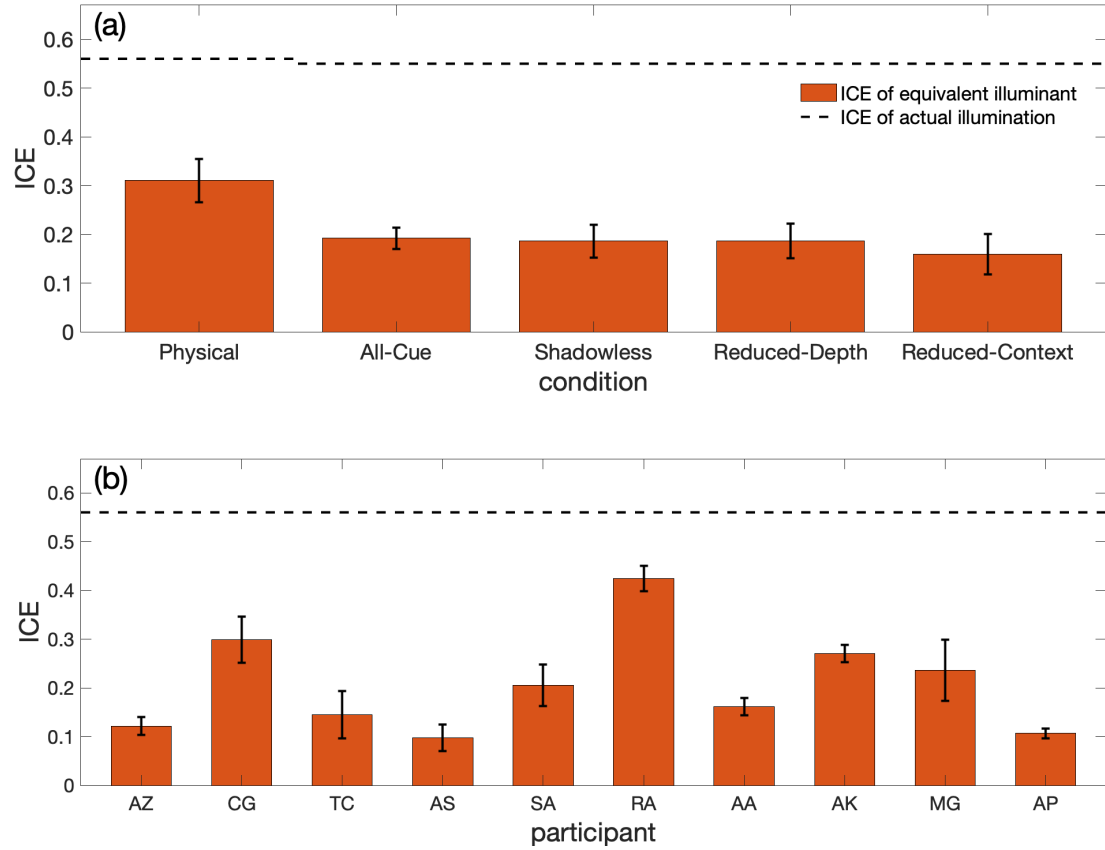


Figure 3.5: Mean value of the fitted ICE parameter, with error bars showing standard errors. (a) Mean for each condition, averaged over participants. The black dashed line shows the actual ICE values for lighting in the physical (0.56) and VR scenes (0.55). (b) Mean for each participant, averaged over conditions. The black dashed line shows ICE=0.55, the lower of the lighting diffuseness values in the physical and VR conditions.

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 comparisons 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 ($p > 0.05$).

Orientation matches

Figure 3.6 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 3.7 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,

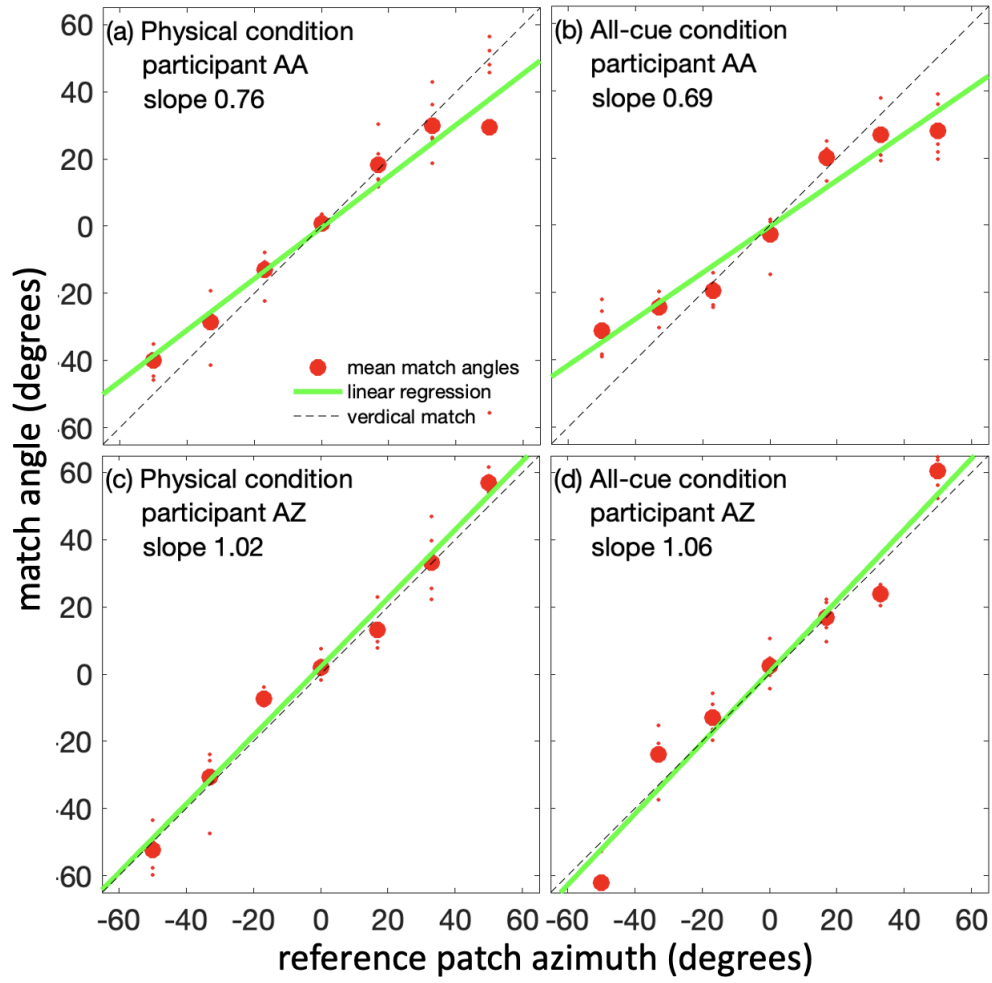


Figure 3.6: Orientation matches for two typical participants in the Physical and All-Cue conditions. The small red dots show matches on individual trials, and the large red dots are means. The green lines show a least-squares regression of match angle against reference patch angle. The black dashed line represents veridical orientation matches for reference.

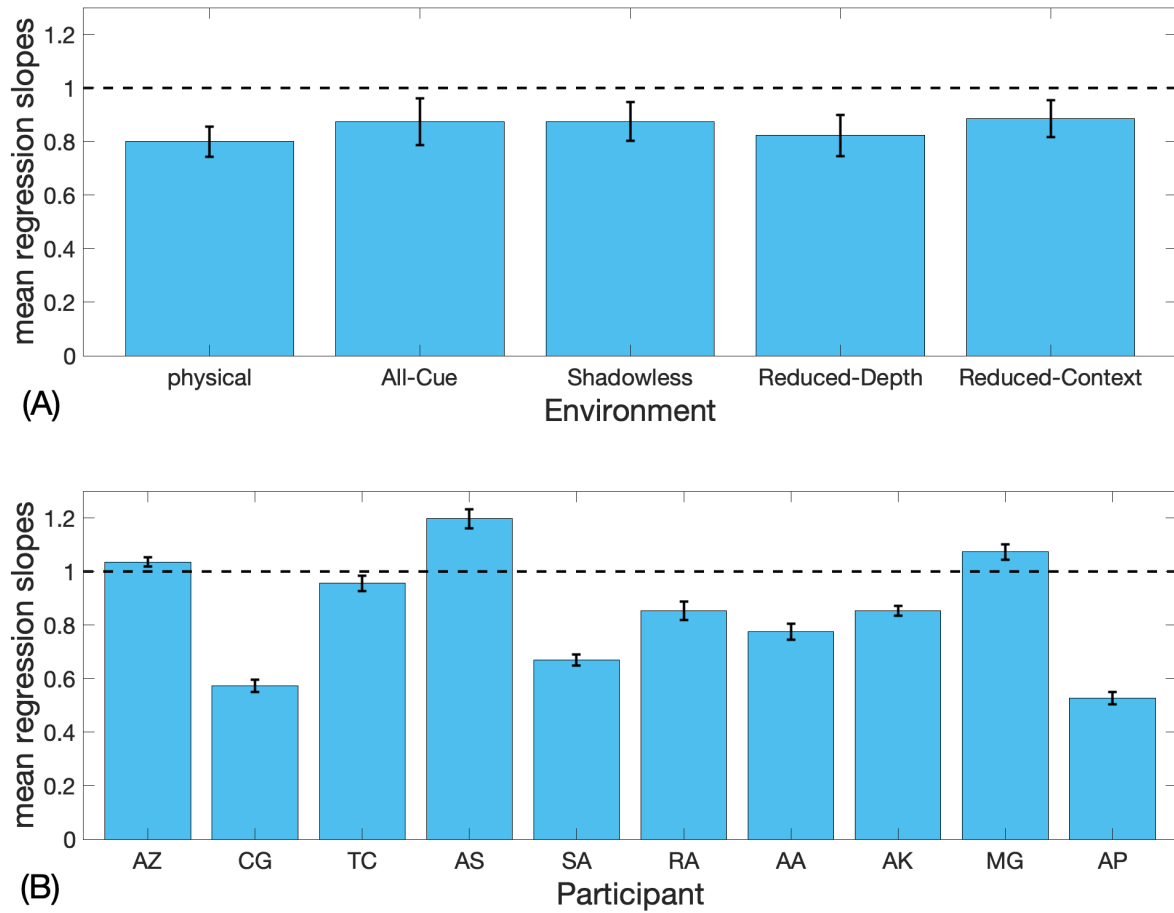


Figure 3.7: Average regression slopes in the orientation matching task, (a) for each condition and (b) for each participant. Error bars show standard errors. The black dashed line shows a value of 1.0, corresponding to veridical orientation matches.

2015).

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 et al., 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 (3.5) 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 3.8a 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 3.8b 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 3.5a), 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, if implicit slant judgements were also biased, like those shown here, this may be one reason why lightness judgements showed imperfect constancy.

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

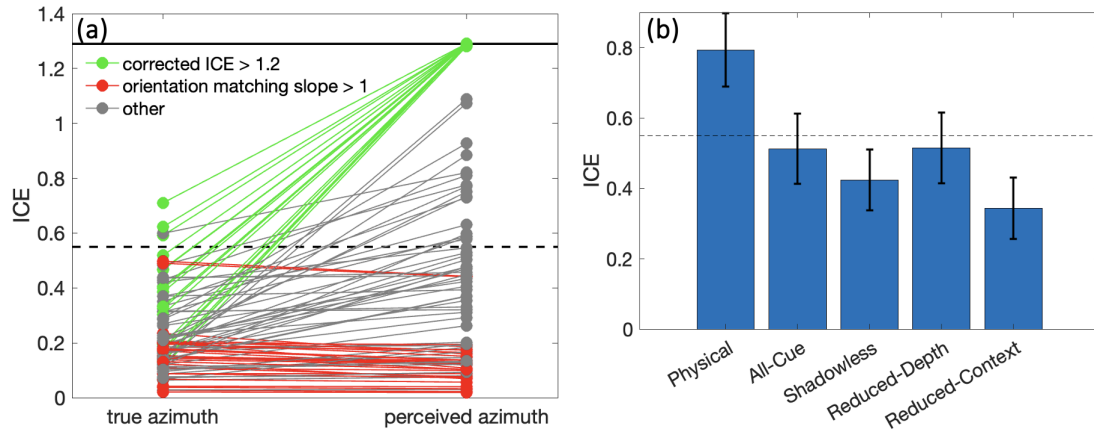


Figure 3.8: (a) ICE estimates based on actual and perceived orientation of the reference patch, for all participants in all conditions. (b) ICE estimates based on perceived orientation, averaged by condition.

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 3.5).

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.

To test these hypotheses, 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.

3.4 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.

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. In each condition participants completed the same lightness matching and orientation matching tasks as in Experiment 1.

Results and discussion

Figure 3.9 shows mean ICE in each condition, calculated using the true azimuth of the reference patch. We reproduce the ICE values from Experiment 1 (Figure 3.5a) 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 F.

Figure 3.9 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

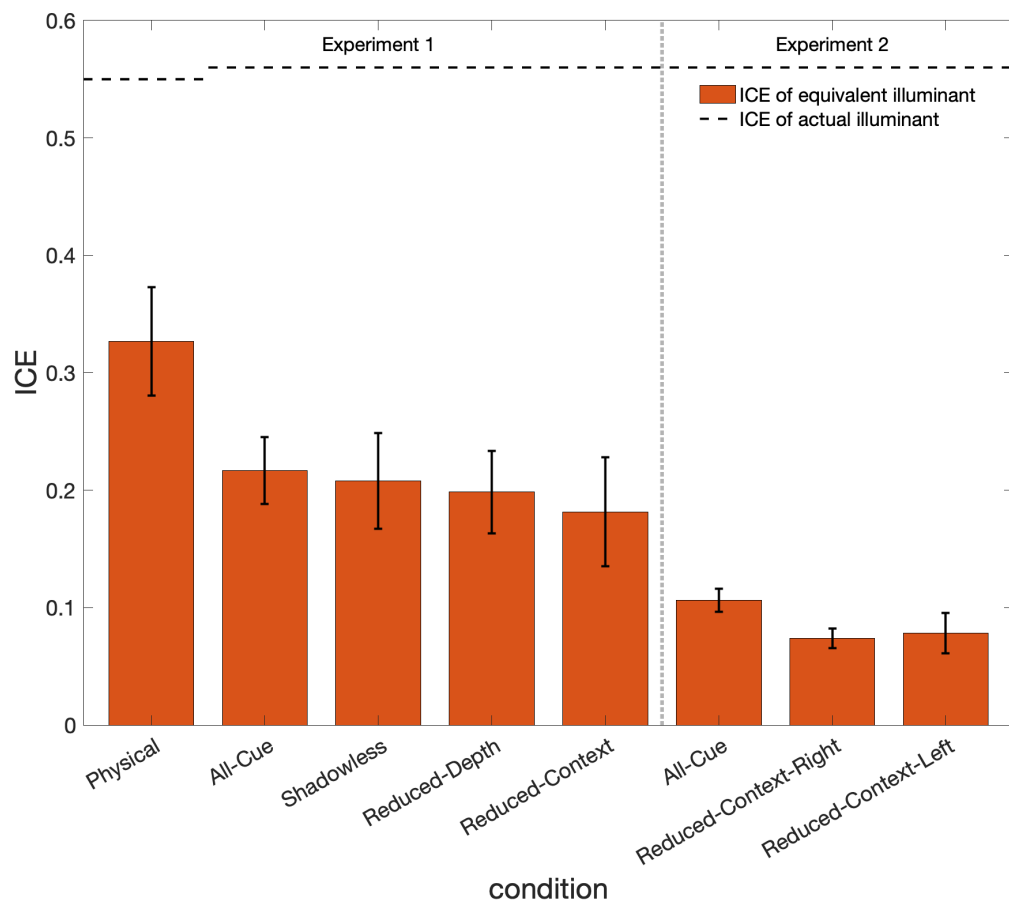


Figure 3.9: Mean ICE in Experiments 1 and 2, averaged by condition. The black dashed line indicates the ICE of the actual lighting in the Physical (0.55) and VR conditions (0.56).

intensity that guides the participant’s lightness matches (equation (3.5)). An independent samples t -test showed that the mean value of this parameter was significantly different in the Reduced-Context-Left ($M = -25^\circ$) and Reduced-Context-Right ($M = 21^\circ$) 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.

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, though not a great deal, as their levels of constancy in the three conditions were not markedly different.

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.

3.5 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

$$\tau = \frac{\log r_m - \log r_0}{\log r_1 - \log r_0} \quad (3.6)$$

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 G 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 3.10(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

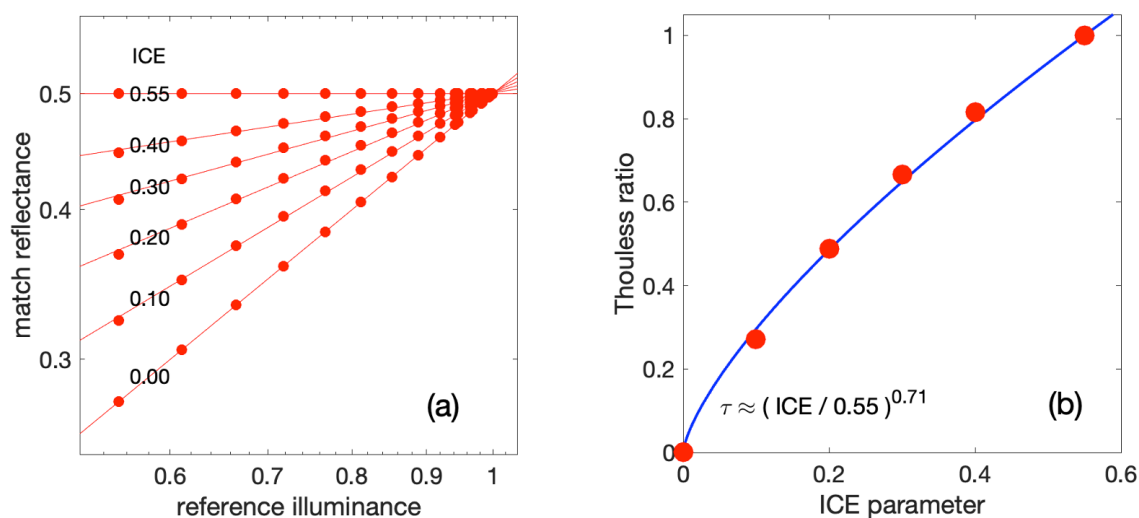


Figure 3.10: Simulated EIM observers. (a) Match reflectance as a function of reference illuminance for model observers with several ICE parameter values. (b) Thouless ratio as a function of ICE parameter value. The blue line shows the fitted power function.

assign to a Lambertian reference patch at a range of orientations. In the Supporting Information we provide code for this simulation.

Figure 3.10(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 ($\text{ICE} = 0.55$) 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 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 3.10(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

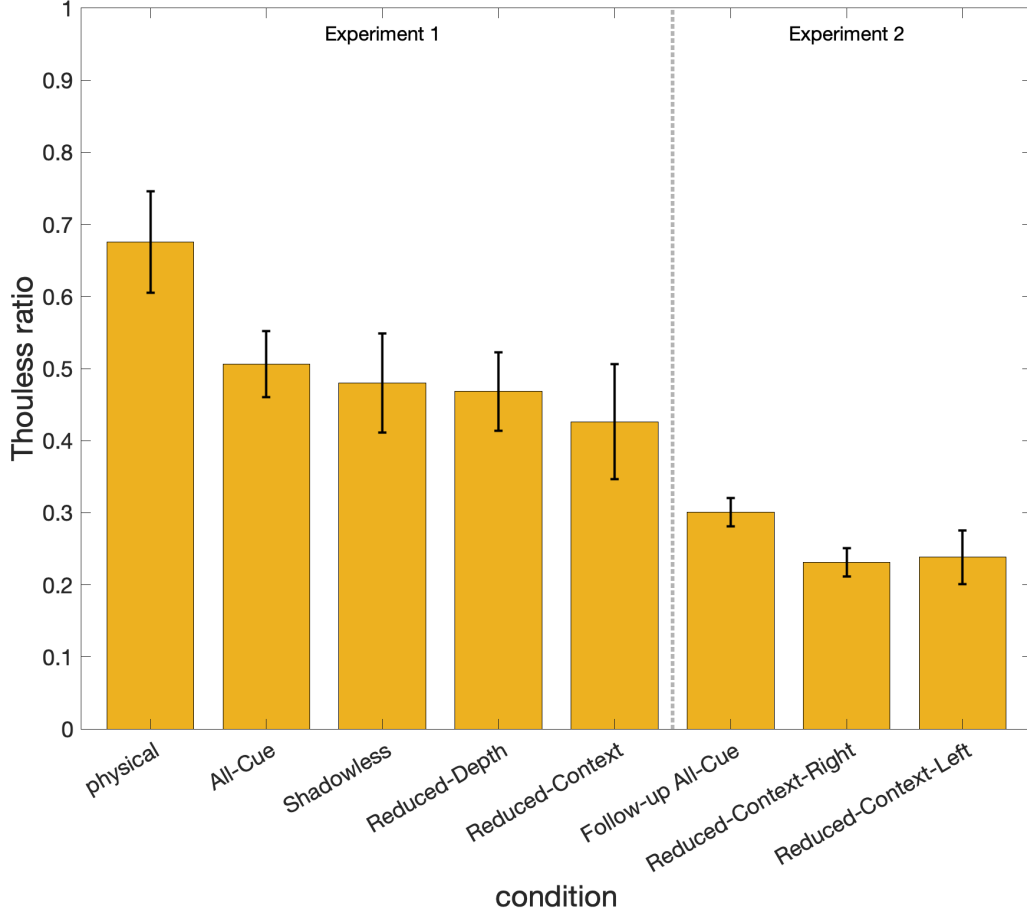


Figure 3.11: Mean Thouless ratios in Experiments 1 and 2, averaged by condition.

$$\tau = (\text{ICE}/0.55)^{0.71}.$$

We can use this result to transform ICE values in the present experiments into Thouless ratios. Figure 3.11 shows the data from the summary in Figure 3.9, 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., K. Y. Patel et al., 2018, 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 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 3.1: 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; K. Y. 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 complexity 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.

There are also several possible reasons why lightness constancy was better in the Physical condition than in the VR conditions. 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 perhaps vice versa, and these features may have helped viewers to disentangle lighting and surface properties. Similarly, 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. K. Y. 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.

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 3.12 illustrates this effect. The figure shows results in the Reduced-Context condition from a simulated model observer. The model does not estimate lighting conditions or surface orientation, but simply passes the stimulus 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 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, 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.

Concluding remarks. These experiments reveal interesting and important differences between perception in real and virtual environments. VR technology has made

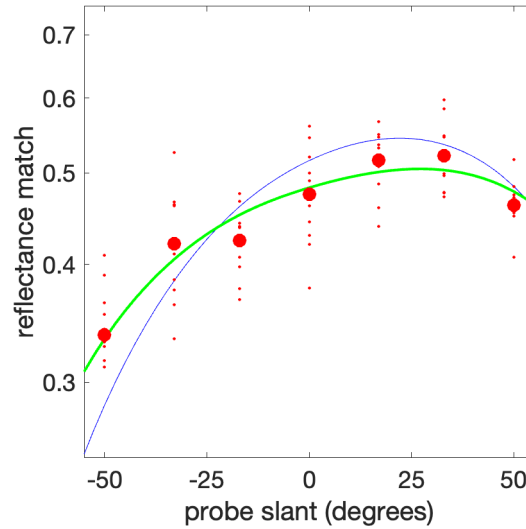


Figure 3.12: Results from a model observer in the Reduced-Context condition. The model’s match reflectance is simply a stochastic, nonlinear function of the reference luminance. The model mimics a degree of lightness constancy ($\text{ICE} = 0.17$, Thouless ratio = 0.43), despite being unable to estimate 3D orientation or lighting conditions. As in figure 3.4, the large red dots are mean reflectance matches, the green fitted curve is the EIM model fit and the blue curve is the stimulus luminance, scaled to make a east-squares fit to the mean reflectance matches.

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.

3.6 Open Practices Statement

The data and supporting analysis code for all experiments are included on OSF at <https://doi.org/10.17605/OSF.IO/3GXRA>.

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.

Appendices

Appendix F

ICE parameters in Experiments 1 and 2

This table reports fitted ICE values for individual participants in Experiments 1 and 2, excluding participants who had mean absolute error greater than 10° in the orientation matching task. For Experiment 1, which had a repeated-measures design, all values in each row are from a single participant. For Experiment 2, which had an independent measures design, each value is from a different participant.

Experiment 1					Experiment 2		
Physical	All-Cue	Shadow -less	Reduced- Depth	Reduced- Context	All-Cue	Reduced- Context -Right	Reduced- Context -Left
0.2276	0.1626	0.1904	0.1649	0.1162	0.1004	0.0794	0.0536
0.3235	0.3774	0.4030	0.3097	0.3417	0.1262	0.0482	0.0261
0.1515	0.1266	0.1069	0.0978	0.1086	0.0701	0.1167	0.1221
0.1573	0.1619	0.0441	0.1328	0.0345	0.1366	0.0989	0.0402
0.1646	0.1430	0.1530	0.1429	0.0525	0.1270	0.0667	0.0478
0.4982	0.2887	0.3115	0.2146	0.2744	0.1223	0.0986	0.1987
0.5105	0.2022	0.1890	0.1742	0.1733	0.0906	0.0618	0.0740
0.4940	0.3550	0.4165	0.4638	0.4939	0.1287	0.0536	0.0380
0.3959	0.1865	0.1784	0.1729	0.1578	0.1192	0.0338	0.1239
0.3463	0.1641	0.0874	0.1110	0.0630	0.0419	0.0820	0.0595

Appendix G

Calculation of Thouless Ratio

We assume that all surfaces are Lambertian. Let the reference patch have reflectance r_f and illuminance i_f . As a Lambertian surface, its luminance is $\ell_f = r_f i_f / \pi$. Let the viewer's match setting at the match patch be reflectance r_m , under illuminance i_m , and so have luminance $\ell_m = r_m i_m / \pi$. In the definition of the Thouless ratio in equation (3.6), r_1 is the match setting that results from perfect constancy, which in this case is $r_1 = r_f$. If the viewer has no lightness constancy and simply matches the luminance of the reference and match stimuli, then their match setting r_0 satisfies $r_0 i_m = r_f i_f$, or $r_0 = r_f i_f / i_m$. If we substitute these expressions for r_1 and r_0 into equation (3.6) and solve for the match setting $\log r_m$, we find

$$\log r_m = (1 - \tau)(\log i_f - \log i_m) + \log r_f \quad (\text{G.1})$$

which is an affine function of $\log i_f$ with slope $1 - \tau$.

Appendix H

At the request of the evaluating committee, here we report effect sizes (Cohen’s d) for the most important comparisons of mean ICE values across conditions. We selected specific comparisons based on their relevance to our research questions.

One important comparison was between the Physical and Follow-up All-Cue conditions, which assesses the effectiveness of virtual environments in simulating real-world settings. Here the difference between mean ICE values had a Cohen’s d of 0.31, indicating a small to moderate effect size.

Another important comparison was between the Follow-up All-Cue condition and the Reduced-Context-Left and Reduced-Context-Right conditions, which allowed us to evaluate the impact of contextual cues on lightness constancy. The comparison between Follow-up All-Cue and Reduced-Context-Right conditions had $d = 0.26$, indicating a small effect size. The comparison between Follow-up All-Cue and Reduced-Context-Left conditions had $d = 0.08$, indicating a very small effect size. These results support our conclusion that, in this experiment, contextual cues were not highly effective in supporting realistic levels of lightness constancy.

Chapter 4

Evaluating lightness constancy in a high-definition VR environment

This chapter will be submitted to the Color and Imaging Conference as proceedings paper.

4.1 Abstract

As VR technology has advanced, its use in performance-critical fields such as medical training and vision research has grown, driving the need for increasingly realistic VR environments. In previous work (K. Y. Patel et al., [submitted:BR-Org-24-359](#)), we evaluated lightness constancy in a task where viewers matched the reflectance of surfaces at different 3D orientations, and we found significantly poorer lightness constancy in VR than in a physical apparatus. Poor constancy in VR may have been due to oversimplified rendering of scenes, e.g., largely achromatic Lambertian surfaces. Motivated by these findings, in the present study we evaluated lightness constancy in more realistic VR scenes, rendered with a broader array of materials, colors, textures, and specular highlights, as well as more realistic shadows. We tested two conditions: a Full-Context condition, where these lighting and material cues were available, and a Reduced-Context condition, where they were not. Participants had significantly better lightness constancy in the Full-Context condition than in the Reduced-Context condition, meaning that they exploited these additional cues. However, lightness constancy was still quite poor in absolute terms, despite the availability of rich lighting and material cues. The reasons for this failure of constancy are unclear from previous literature, and this finding suggests a promising research problem with both fundamental interest and practical applications. We speculate on possible reasons for poor constancy in current VR environments.

4.2 Introduction

Virtual reality (VR) technology has advanced significantly in recent years, finding applications in diverse fields such as gaming (Zyda, [2005](#)), training simulations (Oyelere et al., [2020](#)), therapeutic interventions (Dennis & Patterson, [2020](#)), and vision research (Hibbard, [2023](#)). A key goal of VR development has been enhancing realism, particularly in the accurate rendering of light and its interaction with surfaces, which is important

for ensuring good performance in many applications.

Achieving realism in VR is a challenge due to limitations in current technology, such as low image resolution, limited field of view, and a constrained dynamic range for luminance and color. These constraints, coupled with the necessity for real-time rendering, often lead to compromises in physical accuracy (Pharr et al., 2023), and results in a discrepancy between perception in real and virtual environments. This perceptual gap affects user performance in VR, as shown by discrepancies in tasks such as navigation (Kimura et al., 2017), size perception (Rzepka et al., 2023), depth judgment (Hartle & Wilcox, 2022), and estimating egocentric distance (Creem-Regehr et al., 2022). Thus realism in VR is an ongoing aspiration, requiring continuous evaluation of relevant scene properties to achieve realistic experiences.

Color constancy is one important factor for realism in VR, as it prevents unnatural color shifts between different lighting conditions. *Reflectance*, or *albedo*, is the proportion of visible light reflected by a surface, and *lightness* is defined as perceived reflectance. Lightness constancy is a special case of color constancy, and refers to the ability to perceive reflectance, i.e., achromatic surface color, accurately across varying lighting conditions and environments. Achieving lightness constancy requires consideration of surface orientation, material properties, and scene lighting (Boyaci et al., 2006; Doerschner et al., 2004; Flock & Freedberg, 1970; Kraft & Brainard, 1999).

Some previous studies have examined lightness constancy in relatively simple tasks that required matching the perceived reflectance of two patches on a flat surface, on opposite sides of a shadow boundary (Brainard et al., 1997; Katz, 1935). We have found that in this task, viewers achieve similar levels of lightness constancy in real and VR environments, though with greater inter-subject variability in VR (K. Y. Patel et al., 2024).

In a similar but more demanding task, viewers match the lightness of patches at different 3D orientations relative to a light source, and on separate, disconnected objects (Bloj et al., 2004; Boyaci et al., 2004; Morgenstern et al., 2014; K. Y. Patel

et al., 2018). We have found that in this task, lightness constancy is much weaker in VR than in real environments (K. Y. Patel et al., [submitted:BR-Org-24-359](#)), which is consistent with an earlier study that compared results from a range of tasks using real and virtual environments (Morgenstern et al., 2014).

Equivalent illumination models

Equivalent illumination models (EIMs), which are an example of ‘inverse optics’ modelling (Pizlo, 2001), have been a useful tool for understanding color and lightness constancy (Boyaci et al., 2003; Brainard, 1998; Brainard & Maloney, 2011). The EIM we use here represents achromatic lighting as the sum of a directional light source with illuminance D , in a direction given by unit vector $\mathbf{s}$, and an ambient source with illuminance A in all directions. For a surface with unit normal $\mathbf{n}$, the total illuminance from these light sources is

$$I = D \max(\mathbf{n} \cdot \mathbf{s}, 0) + A \quad (4.1)$$

Here $\max(x, y)$ is the greater of x and y , and $\cdot$ is the vector dot product. A Lambertian surface with reflectance ρ under this illumination produces image luminance given by

$$L = \frac{\rho}{\pi} (D \max(\mathbf{n} \cdot \mathbf{s}, 0) + A) \quad (4.2)$$

The division by π takes into account the diffusion of light over a hemisphere, which is characteristic of matte surfaces (McCluney, 1994).

Solving this equation for ρ gives a model of how a viewer may estimate surface reflectance from luminance and lighting parameters:

$$\rho = \frac{\pi L}{D \max(\mathbf{n} \cdot \mathbf{s}, 0) + A} \quad (4.3)$$

An equivalent illuminant model assumes that the viewer recovers the luminance L correctly, but may have mistaken estimates of the lighting parameters D , A , and $\mathbf{s}$, and

may also misperceive the surface orientation $\mathbf{n}$. An EIM analysis uses this model to describe the viewer’s lightness matching behavior in an interpretable way by inferring the lighting parameters (and possibly also the surface orientation bias) that best account for the reflectance matches ρ . That is, we infer an ‘equivalent illuminant’ that appears to guide the viewer’s lightness matching behavior.

EIM analyses have typically found that viewers estimate the lighting direction $\mathbf{s}$ quite accurately, and have only a small fronto-parallel bias for the surface orientation $\mathbf{n}$, but that they often dramatically over-estimate the ambient-to-directional ratio A/D , resulting in a characteristic pattern of failure of lightness constancy that tends towards luminance matching (Bloj et al., 2004; Boyaci et al., 2004; K. Y. Patel et al., submitted:BR-Org-24-359). Brainard and Maloney (2011) and K. Y. Patel et al. (2024) review EIMs and their application to modelling color and lightness constancy.

Present study

In a previous study (K. Y. Patel et al., submitted:BR-Org-24-359), we found weaker lightness constancy in VR compared to a physical apparatus, in a task where viewers matched the lightness of patches at different 3D orientations. One reason for the failure of constancy in VR may have been that, in order to make it easier to match the physical and VR conditions, we used a simplified rendering model in VR scenes, which consisted primarily of greyscale Lambertian surfaces. However, Boyaci et al., 2006 found that realistic features such as specular highlights were important for good lightness constancy in virtual environments.

In the present study, we aim to create a VR environment that offers more realistic lighting and material cues, to test whether this improves constancy compared to our previous findings. We use Unity’s High Definition Render Pipeline (HDRP), in order to render more realistic materials and shadows, and we include a wider range of material types, colors, and textures. We also incorporate a wider range of surrounding and background objects, created with realistic materials to enhance overall realism and

provide additional lighting cues. We evaluate viewers’ reflectance matching behavior and degree of lightness constancy using an EIM analysis.

We implement two sets of experimental conditions: Full-Context and Reduced-Context. The Full-Context conditions show scenes with the wide range of material and lighting cues described above. The Reduced-Context conditions show test patches with just minimal context, and no surrounding objects that could provide lighting cues to support lightness matching. The Reduced-Context conditions are control conditions where we expect viewers to largely match luminance. Comparing the Full-Context and Reduced-Context conditions allows us to test whether viewers exploit the lighting cues available in the Full-Context conditions, and comparing absolute levels of lightness constancy to those in our previous study (K. Y. Patel et al., [submitted:BR-Org-24-359](#)) allows us to test whether viewers benefit from the more realistic lighting and material cues in the present experiment. In each condition, viewers also perform an orientation matching task, so that we can assess any effect of biases in orientation perception on their lightness matching behaviour.

4.3 Methods

Participants

There were 21 naive participants. Eleven were female, nine were male, and ages ranged from 18 to 34 years. All reported normal or corrected-to-normal visual acuity and no known anomalies in color vision. All procedures were approved by the Office of Research Ethics at York University.

Stimuli

Stimuli were shown in a Meta Quest Pro headset (FOV $106^\circ \times 96^\circ$, resolution 1920×1800 pixels per eye), driven by an NVIDIA GeForce GTX 3090 Ti graphics card on a PC running Windows 10. Scenes were rendered in Unity 2022.2 using the High

Definition Render Pipeline (HDRP), version 14 (Unity Technologies, 2022).

Lightness task. Participants viewed a virtual scene that showed equipment on a table (Figure 4.1), from a simulated viewing distance of 74 cm. The equipment had two components: the reference apparatus and the match apparatus.

The reference apparatus was on the left. In the Full-Context conditions, it included a reference patch surrounded by objects of various colors and materials (Figure 4.1(a)). The reference patch was a 3×3 cm grey square, whose orientation and reflectance varied from trial to trial (details under Procedure).

The match apparatus was on the right. It was a 31.5 cm square, covered with randomly placed grey circles and rectangles, with reflectances ranging from 0.07 to 0.71, on a grey background of reflectance 0.20. A 3×3 cm match patch, whose reflectance could be adjusted by the participant, was on the left side of the panel.

The scene surrounding the reference and match apparatus included objects such as tables, chairs, trees, rocks, a globe, and small plants, all within an area enclosed by walls.

The Reduced-Context conditions were the same as the Full-Context conditions, except that only the reference patch and match apparatus were visible (Figure 4.1(b)). The luminance of the ambient surround was the same as that of the grey panel behind the reference patch in the Full-Context conditions.

Simulated lighting consisted of a directional source and an ambient source. The directional source was located behind the participant at an elevation of 22° . In the Full-Context-Right and Reduced-Context-Right conditions it was 22° to the right of directly behind, and in the Full-Context-Left and Reduced-Context-Left conditions it was 22° to the left. The directional source had 7.7 times the illuminance of the ambient source.

We calibrated the VR headset so that physical luminance was proportional to rendered achromatic RGB value (0-255). We followed the procedure in (Murray et al., 2022), except that instead of the LUT file required by Unity’s Built-in Render Pipeline,

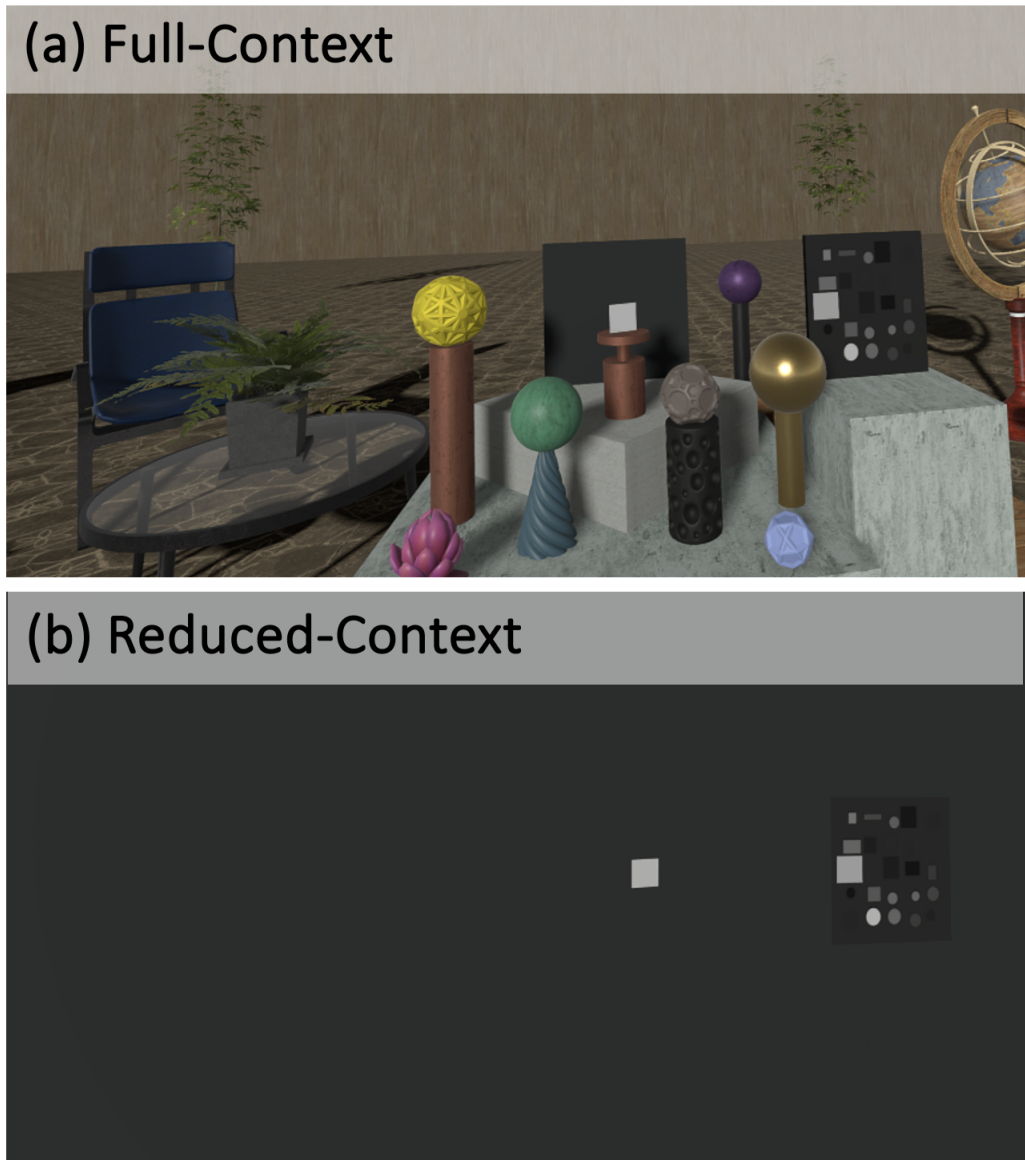


Figure 4.1: VR scenes from the participant's viewpoint. (a) The Full-Context condition. (b) The Reduced-Context condition, which showed only the reference patch and the match apparatus.

we used a .cube file and the tonemapping feature of the HDRP.

Orientation task. In the orientation matching task, a small panel was added to the scene, in front of the reference apparatus. The panel showed a fixed horizontal line, and an adjustable line that the participant could rotate using a thumbstick. Both lines were 7 cm long, and were white on a grey background.

Procedure

There were four conditions: Full-Context-Left, Full-Context-Right, Reduced-Context-Left, and Reduced-Context-Right. Each participant completed one condition. Participants completed the lightness task first, and then the orientation task.

On each trial of the lightness task, the reflectance of the reference patch was randomly set to 0.41 or 0.58, and the orientation was randomly set to -50° , -33° , -17° , 0° , 17° , 33° , or 50° . Here 0° indicates an orientation facing the participant, and positive angles indicate that the patch was rotated to the right. The participant used a thumbstick to adjust the reflectance of the match patch so that it appeared to be the same as the reflectance of the reference patch. Response time was unlimited. Each combination of reflectance and orientation was repeated ten times, for a total of 140 trials.

On each trial of the orientation task, the reference patch (always with reflectance 0.58) was randomly set to one of the seven orientations used in the lightness task. The participant adjusted the line on the orientation-matching panel so that it matched the orientation that the reference patch would have if viewed from above. Each orientation was repeated five times, for a total of 35 trials.

Analysis

We fitted the EIM in equation (4.3) to each participant's reflectance matches, as follows. The reference patch surface normal $\mathbf{n}$ and the lighting direction $\mathbf{s}$ can be

written in spherical coordinates with azimuth ψ and elevation ϕ :

$$\mathbf{n} = (\cos \phi_n \cos \psi_n, \cos \phi_n \sin \psi_n, \sin \phi_n) \quad (4.4)$$

$$\mathbf{s} = (\cos \phi_s \cos \psi_s, \cos \phi_s \sin \psi_s, \sin \phi_s) \quad (4.5)$$

The reference patch was upright, so $\phi_n = 0$. Following Bloj et al., 2004; Boyaci et al., 2004; K. Y. Patel et al., submitted:BR-Orig-24-359, we incorporated a free multiplicative parameter k on the right-hand side of equation (4.3), to allow the participant's reflectance matches to be off by an arbitrary scale factor. In spherical coordinates, after some simplification, equation (4.3) then becomes

$$\rho = \frac{L}{\alpha(\cos(\psi_n - \psi_s) + \beta)} \quad (4.6)$$

where

$$\alpha = \frac{D \cos \phi_s}{\pi k}, \beta = \frac{A}{D \cos \phi_s} \quad (4.7)$$

We made a least-squares fit of equation (4.6) to participants' reflectance match settings. In this analysis, L was set to the luminance of the reference patch on each trial, and ϕ_s was set to the elevation (22°) of the light source. α , β , ψ_s , and k were free parameters. Assuming that the participant perceived the lighting elevation ϕ_s correctly, we can use the fitted value of β to find the ambient-to-directional lighting ratio that is consistent with their responses:

$$A/D = \beta \cos \phi_s \quad (4.8)$$

Morgenstern et al. (2014) define illuminance contrast energy (ICE) as the coefficient of variation (standard deviation divided by mean) of illuminance over all 3D surface orientations at a given point in space. ICE is therefore a measure of the directionality or diffuseness of lighting (Xia et al., 2017), ranging from 0 for completely ambient light to 1.29 for a distant point source. Morgenstern et al. show that an ambient-to-directional

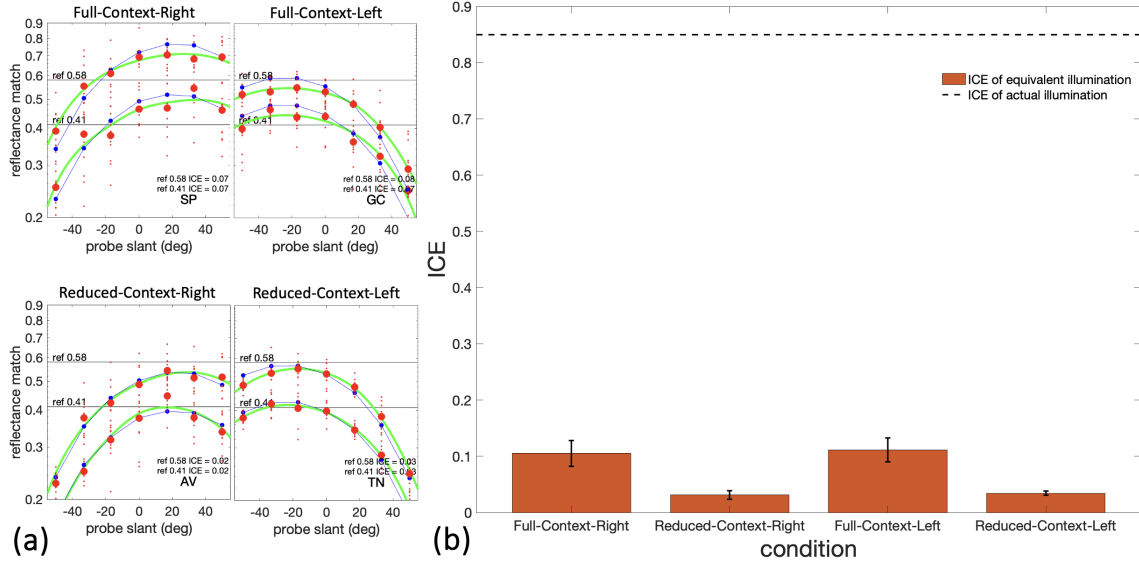


Figure 4.2: (a) Reflectance matches as a function of the orientation of the reference patch for a typical participant in each condition. Small red dots are reflectance matches on individual trials, and large red dots are means. Green fitted lines show least-squares fits of the EIM model. Horizontal black lines indicate the reflectance of the reference patches. Blue curves show stimulus luminance, scaled to make a least-squares fit to mean reflectance matches. (b) Mean values of the fitted ICE parameter, averaged over participants. Error bars show standard errors. The black dashed line shows the true ICE value of the simulated lighting in the VR scene.

lighting ratio A/D corresponds to an ICE value of

$$\lambda = \frac{\sqrt{5/48}}{(A/D) + 0.25} \quad (4.9)$$

The simulated lighting in the present experiment had $A/D = 1/7.7$, corresponding to $\text{ICE} = 0.85$.

Following the criterion used in K. Y. Patel et al., [submitted:BR-Org-24-359](#), we excluded from the analysis one participant who had a mean absolute error greater than 10° in the orientation matching task.

4.4 Results

Lightness matches

Figure 4.2(a) shows reflectance matches as a function of reference patch orientation for a typical participant in each of the four conditions. If participants had perfect constancy, their reflectance matches would fall along the horizontal black lines, indicating no effect of reference patch orientation relative to the light source. If participants had no constancy and simply matched luminance, their reflectance settings would follow the inverted-U pattern of the blue data points, which show the stimulus luminance, scaled to fit the mean reflectance matches in a sum-of-squares manner. Participants' reflectance matches fell between these two theoretical extremes, as is the case in most studies of lightness constancy. The green lines show sum-of-squares fits of the EIM model in equation (4.6) to mean reflectance matches.

Figure 4.2(b) shows mean ICE values in each condition. Independent-samples t -tests showed no significant differences between Full-Context-Right ($M = 0.11$, $SD = 0.05$) and Full-Context-Left ($M = 0.11$, $SD = 0.05$; $t(8) = 0.20$, $p = 0.84$) conditions, or between Reduced-Context-Right ($M = 0.04$, $SD = 0.02$) and Reduced-Context-Left ($M = 0.04$, $SD = 0.01$; $t(8) = 0.29$, $p = 0.78$). Consequently, we pooled data within the Full-Context conditions ($M = 0.11$, $SD = 0.04$) and Reduced-Context conditions ($M = 0.04$, $SD = 0.02$) for subsequent analyses.

Mean ICE values in pooled Full-Context and Reduced-Context conditions were both significantly less than the veridical value of 0.85 and significantly greater than 0, as determined by individual one-sample t -tests, with all p values below the Bonferroni-corrected alpha level of 0.025. Thus participants' reflectance matches were consistent with an overestimate of lighting diffuseness, as in most previous studies (Brainard & Maloney, 2011; Morgenstern et al., 2014; K. Y. Patel et al., submitted:BR-Org-24-359).

An independent samples t -test showed that mean ICE was significantly higher in the pooled Full-Context conditions than in the Reduced-Context conditions ($t(18) = 4.60$,

$p < 0.05$). Thus participants benefited from the additional lighting and material cues in the Full-Context conditions, and had better lightness constancy in those conditions than in the Reduced-Context conditions.

Orientation matches

Participants' orientation matches were proportional to the simulated azimuth of the reference patch, but were biased toward fronto-parallel. The average slope of a linear regression of match azimuth against true azimuth was 0.80, significantly less than the value of 1 that would indicate veridical matches ($t(19) = -4.55$, $p < 0.05$). This underestimate of slant is consistent with previous studies (Bloj et al., 2004; Boyaci et al., 2003; K. Y. Patel et al., [submitted:BR-Org-24-359](#); Saunders & Chen, 2015). Furthermore, an ANOVA showed no significant differences in regression slopes across the four conditions ($F(12, 3) = 0.30$, $p = 0.83$), indicating a consistent bias toward fronto-parallel orientations.

4.5 Discussion

In an EIM analysis of lightness matching, a low ICE value, meaning an overestimate of lighting diffuseness in the equivalent illuminant, indicates imperfect lightness constancy. An overestimate of diffuseness means that the participant behaves as if a change in reference patch orientation should change the reference patch luminance relatively little, and so the participant mistakenly attributes the change in luminance to a change in reflectance. We found that mean ICE was closer to the veridical value in the Full-Context conditions than in the Reduced-Context conditions, indicating that participants exploited the lighting cues available in the former conditions to support lightness constancy. However, another clear and important result was that all mean ICE values were much lower than the true ICE of the simulated lighting (Figure 4.2), pointing to generally poor constancy in all conditions. This aligns with previous re-

search using similar stimuli and tasks (Morgenstern et al., 2014; K. Y. Patel et al., submitted:BR-Org-24-359).

Lightness constancy can also be quantified using the Thouless ratio, a measure that facilitates comparisons of constancy across studies Thouless, 1931. The Thouless ratio is defined as

$$\tau = \frac{\log r_m - \log r_0}{\log r_1 - \log r_0} \quad (4.10)$$

where r_m is the participant’s reflectance match setting, r_1 is the setting that would result from perfect constancy, and r_0 is the setting that would result from luminance matching. The Thouless ratio ranges from 0 to 1, where $\tau = 1$ indicates perfect constancy, and $\tau = 0$ indicates luminance matching, i.e., no constancy.

In previous work, we showed how to convert equivalent-illuminant ICE parameters to Thouless ratios (K. Y. Patel et al., submitted:BR-Org-24-359). The methods reported there show that with a true ICE value of 0.85, as in the present experiment, the conversion is approximately given by

$$\tau = (\text{ICE}/0.85)^{0.65} \quad (4.11)$$

Figure 4.3 shows the mean ICE values from Figure 4.2b, pooled across Left and Right conditions, and converted to Thouless ratios. Thouless ratios were low, and even in the Full-Context condition the mean was just 0.22, indicating very weak constancy. In complex physical scenes with rich lighting and material cues, where lightness is evaluated across a shadow boundary, we found that Thouless ratios are typically around 0.8 to 0.9 (K. Y. Patel et al., 2024). In a previous study with a similar task to the present experiment, that required matching reflectance across 3D orientations, we measured a Thouless ratio of 0.68 in a physical apparatus. In the same study, in VR conditions with less realistic rendering than in the present experiment, and in a design where each participant ran in just one condition as reported here, Thouless ratios ranged from 0.35 to 0.12. Thus despite enriching the VR scene with additional lighting and material cues

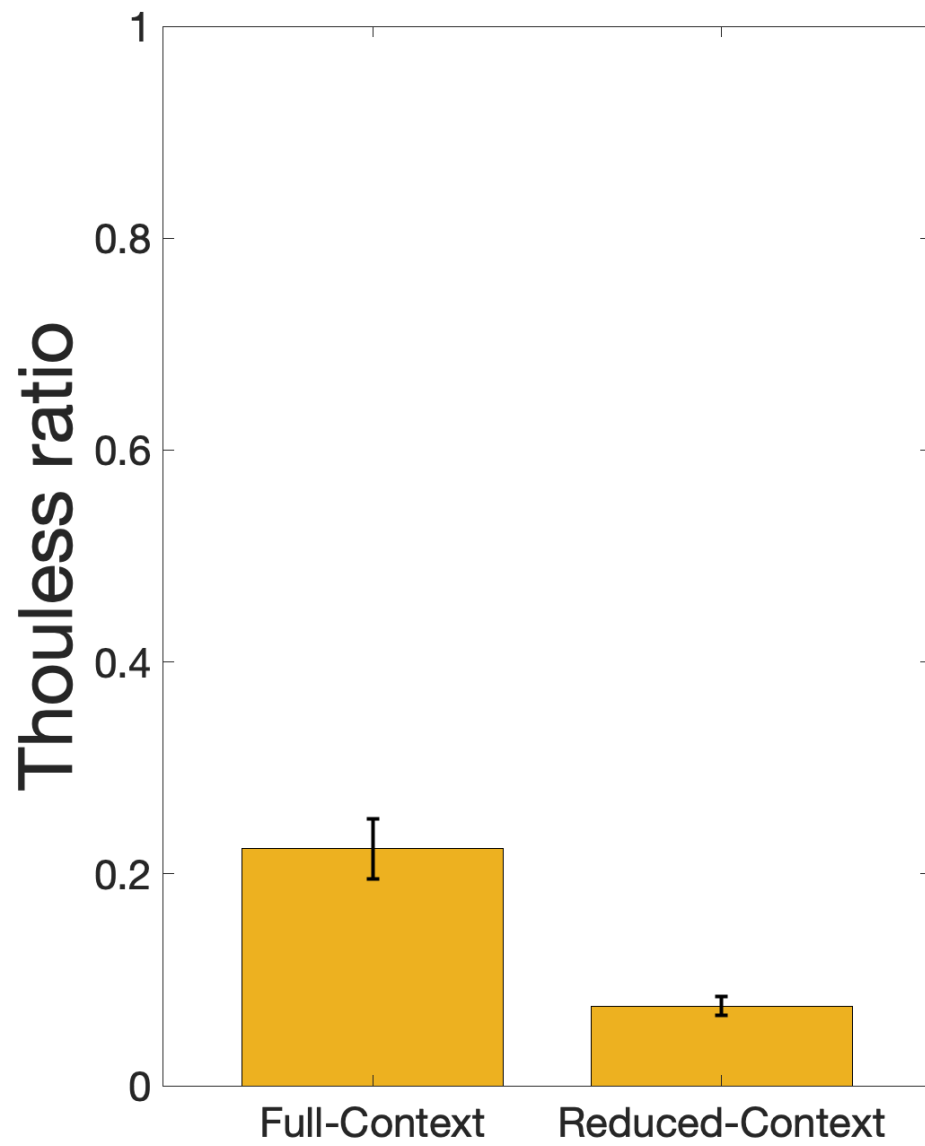


Figure 4.3: Mean Thouless ratios in Full- and Reduced-Context conditions. Error bars show standard errors.

in the present study, lightness constancy did not improve compared to our previous study.

Errors in perception of surface orientation are unlikely to explain this result. As reported above, the mean slope in regressions of perceived vs. true azimuth was 0.80, which is similar to values measured in physical and VR scenes in previous studies (Bloj et al., 2004; Boyaci et al., 2003; Saunders & Chen, 2015).

This substantial failure of lightness constancy even in complex and carefully rendered VR scenes is surprising. It is unclear from previous literature why constancy should be weaker in such scenes than with physical stimuli (K. Y. Patel et al., [submitted:BR-
Org-24-359](#)). One possibility is that the explicit orientation judgements we used to test participants' orientation perception are different from implicit judgements that guide lightness perception. Another possibility is that fine surface textures (both 2D and 3D) and subsurface scattering that were not rendered in our VR scenes play an important role in perception of realistic materials and lightness. Similarly, pixellation is often visible in VR, and this departure from realism may interfere with perception.

In the present experiment, the test patch was an isolated square supported on a thin pole. Previous work has shown that viewers are highly insensitive to lighting inconsistencies across scenes (Ostrovsky et al., 2005; Wilder et al., 2019). It may be that lightness constancy is substantially better in scenes where reference and test patches are integrated into solid, perhaps polyhedral objects, as viewers appear to be better at integrating lighting cues across single objects than across distinct parts of complex scenes.

Conclusion

These results highlight significant perceptual differences in lightness constancy between real and virtual environments. Despite advances in VR technology that allow for the creation of complex, interactive environments, our findings show that lightness matching performance in VR only marginally surpasses simple luminance matching. It

remains unclear why lightness constancy is substantially weaker in VR. Investigating these perceptual discrepancies remains a critical avenue for enhancing VR applications and deepening our understanding of visual processing in realistic settings.

Author contributions

All authors contributed to the conceptualization, design of methodology, and review 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, revised the drafts of the conference paper and the code used in the experiment, and led the calibrations that will be detailed in a separate publication. Laurie M. Wilcox provided guidance during the development of the VR scene. Suyash Singh assisted in data collection and programming of the experiments.

Appendices

Appendix I

At the request of the evaluating committee, here we report effect sizes in the analysis of mean ICE values. The main comparison in this paper was between the mean ICE values in Full-Context ($M = 0.11$, $SD = 0.04$) and Reduced-Context ($M = 0.04$, $SD = 0.02$) conditions. The effect size was given by Cohen's $d = 2.2$, indicating a very large effect size. This indicates that, even though lightness constancy was much poorer in the Full-Context condition than in typical physical scenes, it was nevertheless markedly higher than in the Reduced-Context condition, and was supported by the lighting and material cues provided in the virtual scenes.

Chapter 5

Conclusion

This dissertation advances our understanding of lightness constancy, a fundamental aspect of human visual perception that enables consistent perception of object reflectance despite changes in illumination. Figure 5.1 presents Thouless ratios across all studies in the dissertation. The results demonstrate uneven success in achieving realistic lightness constancy in VR. While it is possible to achieve realistic lightness constancy in VR for tasks like discerning lightness across a shadow boundary, we struggle with more complex tasks that require accurate integration of the surfaces' orientation. This limitation is particularly significant for applications that rely on precise visual representations, such as medical training simulations and vision research studies (Hibbard, 2023). In these fields, realistically representing visual environments is essential for advancing our understanding and enhancement of human perception and interaction. The findings underscore the importance of ongoing visual perception research to enhance VR technologies for practical applications.

5.1 How Realistic is Lightness Constancy in VR?

In summary, as shown in Figure 5.1, the results of our studies reveal that even with robust lighting and shape cues, lightness constancy is significantly better in physical environments than in VR. For lightness judgments across a shadow boundary in rich

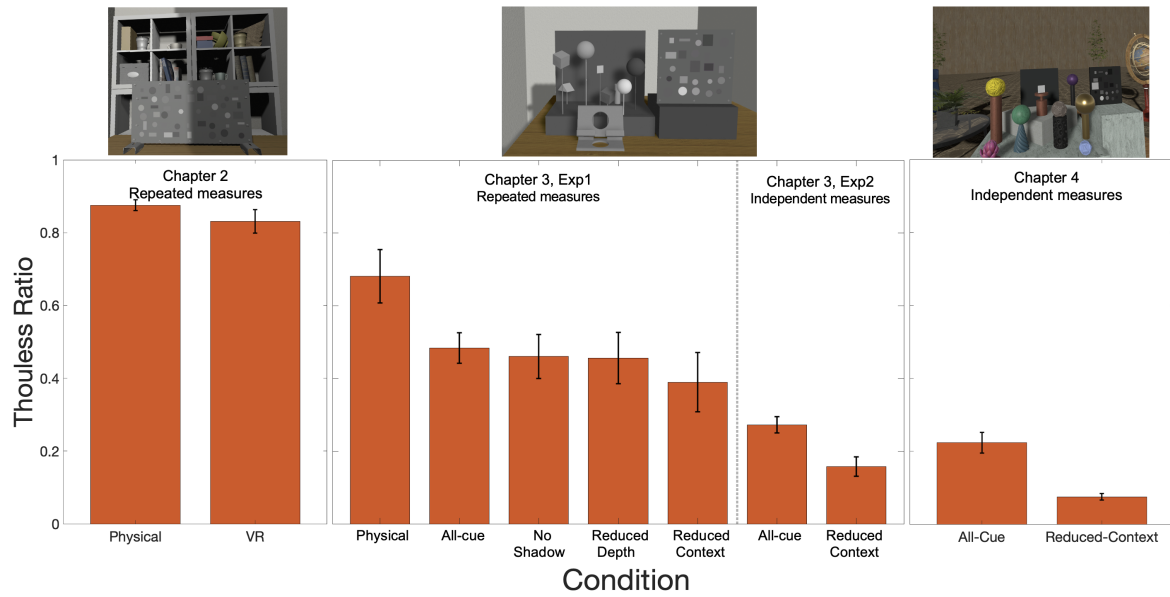


Figure 5.1: Thouless ratios from all studies in the dissertation. For Chapters 3 and 4, the 'left' and 'right' conditions are averaged for the relevant All-Cue and Reduced-Context conditions.

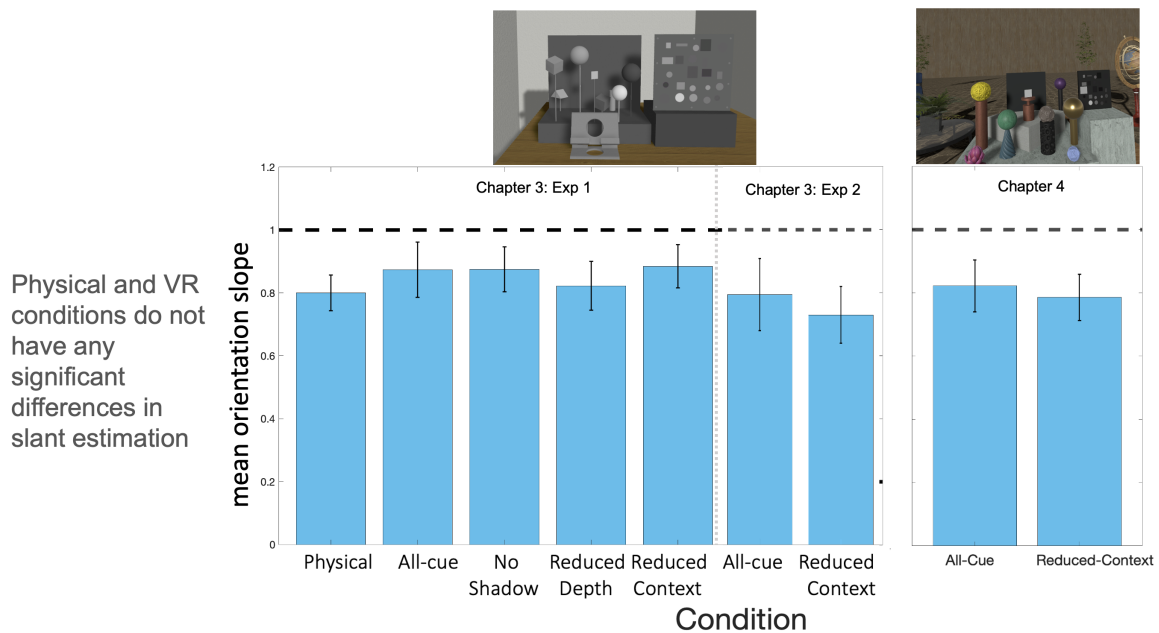


Figure 5.2: Integrated overview of orientation slopes across conditions from Chapters 3 and 4 in the dissertation. Again, the 'left' and 'right' conditions are averaged for the relevant All-Cue and Reduced-Context conditions.

scenes like in Chapter 2, we see that the participants are able to achieve fairly high levels of lightness constancy in both physical and VR scenes. Although the average levels of lightness constancy were similar between both physical and VR conditions, the VR conditions showed significantly higher variability. This indicates that for a subset of participants, lightness constancy was not as high in VR as in physical environments.

In Chapters 3 and 4, participants exhibited moderate levels of lightness constancy when judging lightness across 3D orientation in the physical environment. However, the levels of lightness constancy observed when lightness was judged across 3D orientation in the physical environment were lower than those observed when lightness was judged across a shadow boundary, as seen in Figure 5.1. In contrast, lightness constancy in VR was significantly less realistic when lightness was judged across 3D orientation compared to when lightness was judged across a shadow boundary. Participants made orientation matches for the slant estimation tasks in Chapters 3 and 4, where they adjusted a line to estimate the perceived orientation of the reference patch. Figure 5.2 presents the combined mean orientation slopes from these tasks. The average underestimation of slant, with a slope of about 0.8, is consistent with previous literature and across both VR and physical conditions (Bloj et al., 2004; Boyaci et al., 2003; Saunders & Chen, 2015). Our investigation into reasons for less realistic lightness constancy in VR revealed that participants can accurately estimate the orientation of the reference patch, indicating that other factors may have contributed to the diminished realism. The possible underlying factors, along with future research directions, are explored in the following section.

One of the primary reasons for the observed differences in lightness constancy between Chapters 2 and 3 may be attributed to variations in the lightness constancy tasks used in each chapter. Specifically, in Chapter 3, the test patch is attached to the rest of the scene only by a thin rod, which could influence perceptual integration differently compared to Chapter 2, where the test patch is part of a larger, more coherent object. This difference in presentation might affect the visual system's ability to integrate

the test patch with its surrounding context, potentially reducing lightness constancy. The isolation of the test patch might minimize contextual cues that aid in maintaining constancy by providing fewer relational signals about the lighting and surrounding colors, which are crucial for achieving stable lightness perception across varying illumination conditions. This speculation suggests that the integration of the test patch into the broader visual scene plays a significant role in how lightness constancy is achieved, pointing to the importance of object and scene context in visual perception studies. Further investigations could explore how different methods of presenting test patches impact lightness constancy, potentially guiding future experimental designs to better understand the underlying mechanisms of this perceptual phenomenon.

Overall, these results lead us to conclude that while VR can serve as a reasonable proxy for real-world scenarios in simpler visual tasks, it currently falls short in accurately replicating realistic lightness constancy in more demanding visual tasks. This discrepancy highlights an important area for future research and technological refinement, aiming to enhance the realism and effectiveness of VR environments for complex perceptual tasks.

5.2 Role of Immersion

The ability of VR to simulate realistic and complex three-dimensional spaces allows for dynamic interactions with the environment, where users can experience changes in lighting and shadows as they move. This capability is critical for studying lightness constancy because it offers a more accurate simulation of how we perceive objects under varying lighting conditions in daily life. In contrast to stationary displays, where perception is limited to the fixed perspectives and static light conditions, VR's interactive nature means that users can observe objects from multiple angles and distances.

From a psychological perspective, immersion can lead to greater attentional engagement with the visual task at hand, potentially enhancing the accuracy of visual information processing. This cognitive shift towards prioritizing visual accuracy may be

neurologically supported by increased activation in brain regions associated with visual and spatial awareness. Some neuroimaging studies provide evidence of enhanced cortical activation during VR tasks compared to tasks on traditional displays, suggesting that the immersive nature of VR significantly influences brain activity (Bohil et al., 2011; Kober et al., 2012).

5.3 Future Directions for VR

It remains unclear why lightness constancy is substantially weaker in VR for the task where the 3D orientation of the reference patch was varied. The disparity in performance raises intriguing questions about the fundamental and practical limitations of current VR technologies in simulating realistic visual perceptions. Fundamental limitations might include the inability of VR systems to accurately model the complex interactions of light with different materials, such as diffraction, refraction, or the subtleties of surface textures and how they affect light absorption and scattering. Whereas, practical limitations of VR technologies involve implementation challenges like resolution, refresh rates, and luminance range, which may hinder achieving realistic lightness constancy due to current technological and economic constraints. Understanding these perceptual challenges is essential for improving VR's effectiveness in replicating real-world visual experiences.

A notable limitation of the studies presented in this dissertation is that the task may have encouraged a strategy to mostly focus on the relevant patches rather than constantly scanning the whole scene. A promising alternative method involves prolonged adaptation during which participants searched for objects placed throughout the scene, which has been shown to improve color constancy (Rodriguez et al., 2024). This method encourages participants to look around and become more familiar with the lighting in the environment, potentially also enhancing their performance in lightness estimation tasks. This aspect of Rodriguez et al. (2024)'s approach may have improved participants'

understanding of local lighting conditions and geometry, offering a more comprehensive engagement with the environment compared to our task, where participants could focus solely on a reference and a match patch.

Another method uses a physically-based 360° imaging technique to render realistic lighting in VR. This technique involves capturing a high dynamic range image of the real test-room under controlled lighting conditions using a light probe. The captured image includes the complete light field of the environment, encompassing direct and ambient light intensities and directions, which is then used to render the virtual scene with realistic lighting. This approach aims to provide a more comprehensive simulation of real-world lighting conditions compared to traditional point and ambient lighting models. Traditional models, while computationally simpler, often fail to capture the subtle nuances and full dynamic range of natural lighting. This technique is claimed to enhance the realism and applicability of virtual environments for various applications (Abd-Alhamid et al., 2019). Implementing this methodology could potentially improve lightness constancy in VR by providing a more accurate representation of the light field.

In evaluating the field of view restrictions imposed by VR headsets, which typically include a characteristic black surround, it is important to understand how these limitations can alter visual perception compared to the open view available in the physical condition. Studies have demonstrated that a restricted FoV can adversely affect depth perception, spatial awareness, and may even exacerbate simulator sickness (Lin et al., 2002; Ruddle et al., 2011). To address this issue, a possible modification could involve adapting the physical condition to simulate a similar restricted field of view as experienced in VR. This adjustment aims to standardize the experimental conditions, thereby providing a more controlled and directly comparable assessment of visual perception across different setups than previously documented in this dissertation.

5.4 Concluding Statement

In conclusion, this dissertation enhances our understanding of human lightness perception and its variability across different environments. It underscores the ongoing challenges in achieving realism in VR, especially for tasks requiring lightness judgments when the target surfaces' orientation is varied. While VR is a promising tool, its current limitations highlight the need for continued research and development to ensure it can be reliably used in fields requiring realistic visual perception.

References

- Abd-Alhamid, F., Kent, M., Bennett, C., Calautit, J., & Wu, Y. (2019). Developing an innovative method for visual perception evaluation in a physical-based virtual environment. *Building and Environment*, 162, e106278.
- Adams, H., Stefanucci, J., Creem-Regehr, S., & Bodenheimer, B. (2022). Depth perception in augmented reality: The effects of display, shadow, and position. *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 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. (2000). Lightness perception and lightness illusions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 339–351). The MIT Press.
- Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, 262(5142), 2042–2044.
- Allred, S. R., Radonjić, A., Gilchrist, A. L., & Brainard, D. H. (2012). Lightness perception in high dynamic range images: Local and remote luminance effects. *Journal of Vision*, 12(2), 1–16.
- Anderson, B. L., & Winawer, J. (2005). Image segmentation and lightness perception. *Nature*, 434(7029), 79–83.

- Arend, L. E., & Goldstein, R. (1987). Simultaneous constancy, lightness, and brightness. *Journal of the Optical Society of America A*, 4(12), 2281–2285.
- Arend, L. E., Reeves, A., Schirillo, J., & Goldstein, R. (1991). Simultaneous color constancy: Papers with diverse munsell values. *Journal of the Optical Society of America A*, 8, 661–672.
- Arend, L. E., & Spehar, B. (1993). Lightness, brightness, and brightness contrast: 1. illuminance variation. *Perception & Psychophysics*, 54(4), 446–456.
- Barron, J. T., & Malik, J. (2015). Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8), 1670–1687.
- Baslamisli, A. S., Le, H.-A., & Gevers, T. (2018). Cnn based learning using reflection and retinex models for intrinsic image decomposition. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 6674–6683.
- Beck, J., Prazdny, K., & Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception & Psychophysics*, 35(5), 407–422.
- Belhumeur, P. N., Kriegman, D. J., & Yuille, A. L. (1999). The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1), 33–44.
- Berkman, M. I., & Akan, E. (2019). Presence and immersion in virtual reality. In N. Lee (Ed.), *Encyclopedia of computer graphics and games*. Springer.
- Blakeslee, B., Cope, D., & McCourt, M. E. (2016). The oriented difference of gaussians (odog) model of brightness perception: Overview and executable mathematical notebooks. *Behavior Research Methods*, 48, 306–312.
- 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.
- Blakeslee, B., & McCourt, M. E. (2003). A multiscale spatial filtering account of brightness phenomena. In *Levels of perception* (pp. 47–72). Springer.

- Blakeslee, B., & McCourt, M. E. (2008). Nearly instantaneous brightness induction. *Journal of Vision*, 8(2), 1–18.
- Blakeslee, B., & McCourt, M. E. (2012). When is spatial filtering enough? investigation of brightness and lightness perception in stimuli containing a visible illumination component. *Vision Research*, 60, 40–50.
- 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.
- Blakeslee, B., Reetz, D., & McCourt, M. E. (2009). Spatial filtering versus anchoring accounts of brightness/lightness perception in staircase and simultaneous brightness/lightness contrast stimuli. *Journal of Vision*, 9(3), 22–22.
- 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), 735–746.
- Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature reviews neuroscience*, 12(12), 752–762.
- Boletsis, C., Cedergren, J. E., & Kongsvik, S. (2017). Hci research in virtual reality: A discussion of problem-solving. *International Conference on Interfaces and Human Computer Interaction, IHCI 2017, Portugal*, 21–23.
- Bonneel, N., Kovacs, B., Paris, S., & Bala, K. (2017). Intrinsic decompositions for image editing. *Computer Graphics Forum*, 36(2), 593–609.
- Boyaci, H., Doerschner, K., & Maloney, L. (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. (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. *Journal of the Optical Society of America A*, 15(2), 307–325.
- Brainard, D. H., Brunt, W., & Speigle, J. (1997). Color constancy in the nearly natural image. 1. asymmetric matches. *Journal of the Optical Society of America A*, 14(9), 2091–2110.
- Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illumination models. *Journal of Vision*, 11(5), 1–18.
- Bressan, P. (2006). The place of white in a world of grays: A double-anchoring theory of lightness perception. *Psychological Review*, 113(3), 526.
- Bressan, P., & Actis-Grosso, R. (2001). Simultaneous lightness contrast with double increments. *Perception*, 30(7), 889–897.
- Brunswik, E. (1928). Zur Entwicklung der Albedowahrnehmung. *Zeitschrift für Psychologie*, 109, 40–115.
- Chalmers, A., & Ferko, A. (2008). Levels of realism: From virtual reality to real virtuality. *Proceedings of the 24th Spring Conference on Computer Graphics*, 19–25.
- Cipresso, P., Giglioli, I. A. C., Raya, M. A., & Riva, G. (2018). The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in psychology*, 2086.
- Cornsweet, T. N. (1970). *Visual perception*. New York: Academic Press.
- 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.

- Dakin, S. C., & Bex, P. J. (2003). Natural image statistics mediate brightness ‘filling in’. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1531), 2341–2348.
- Das, P., Karaoglu, S., & Gevers, T. (2022). Intrinsic image decomposition using physics-based cues and cnns. *Computer Vision and Image Understanding*, 223, 103538.
- Dennis, O. P., & Patterson, R. M. (2020). Medical virtual reality. *Journal of Hand Therapy*, 33(2), 243–245.
- 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.
- Doerschner, K., Boyaci, H., & Maloney, L. (2004). Human observers compensate for secondary illumination originating in nearby chromatic surfaces. *Journal of Vision*, 4(2), 3–3.
- Economou, E., Zdravkovic, S., & Gilchrist, A. (2007). Anchoring versus spatial filtering accounts of simultaneous lightness contrast. *Journal of Vision*, 7(12), 2–2.
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Chapman; Hall/CRC Press.
- Elor, A., & Kurniawan, S. (2020). The ultimate display for physical rehabilitation: A bridging review on immersive virtual reality. *Frontiers in Virtual Reality*, 1(585993), 1–17.
- Epstein, W. (1961). Phenomenal orientation and perceived achromatic color. *The Journal of Psychology*, 52(1), 51–53.
- Fairchild, M. D. (2013). *Color appearance models, third edition*. West Sussex, UK: John Wiley & Sons, Ltd.
- Fechner, G. T. (1948). *Elements of psychophysics*, 1860.
- Flock, H., & Freedberg, E. (1970). Perceived angle of incidence and achromatic surface color. *Perception & Psychophysics*, 8(4), 251–256.

- Gerhard, H. E., & Maloney, L. T. (2010a). Detection of light transformations and concomitant changes in surface albedo. *Journal of Vision*, *10*(9), 1–22.
- Gerhard, H. E., & Maloney, L. T. (2010b). Estimating changes in lighting direction in binocularly viewed three-dimensional scenes. *Journal of Vision*, *10*(9), 1–14.
- Gerschütz, B., Fechter, M., Schleich, B., & Wartzack, S. (2019). A review of requirements and approaches for realistic visual perception in virtual reality. *Proceedings of the Design Society: International Conference on Engineering Design*, *1*(1), 1893–1902.
- Gil Rodriguez, 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 Rodriguez, R., Toscani, M., Guarnera, D., Guarnera, G. C., Bayer, F., & Gegenfurtner, K. (2020). Color constancy in a virtual reality environment. *Journal of Vision*, *20*(11), 1226–1226.
- Gilchrist, A. (1977). Perceived lightness depends on perceived spatial arrangement. *Science*, *195*(4274), 185–187.
- Gilchrist, A. (1980). When does perceived lightness depend on perceived spatial arrangement? *Perception & Psychophysics*, *28*(6), 527–538.
- Gilchrist, A. (2006). *Seeing black and white*. New York: Oxford University Press.
- Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., Spehar, B., Annan, V., & Economou, E. (1999). An anchoring theory of ljournal of lightness perception. *Psychological Review*, *106*(4), 795–834.
- Greengard, S. (2019). *Virtual reality*. The MIT Press.
- Guarnera, D., Guarnera, G. C., Ghosh, A., Denk, C., & Glencross, M. (2016). Brdf representation and acquisition. *Computer Graphics Forum*, *35*, 625–650.
- Hartle, B., & Wilcox, L. M. (2022). Stereoscopic depth constancy for physical objects and their virtual counterparts. *Journal of Vision*, *22*(4), 1–19.

- He, J., & Wu, Y. (2022). Application of digital interactive display design based on computer technology in vr film. *Mobile Information Systems, 2022*, 1–7.
- Heinemann, E. G. (1955). Simultaneous brightness induction as a function of inducing- and test-field luminances. *Journal of Experimental Psychology, 50*(2), 89–96.
- Hibbard, P. B. (2023). Virtual reality for vision science. In *Topics in behavioral neurosciences* (pp. 1–29). Springer.
- Horn, B. K. (1974). Determining lightness from an image. *Computer Graphics and Image Processing, 3*(4), 277–299.
- Itoh, Y., Langlotz, T., Iwai, D., Kiyokawa, K., & Amano, T. (2019). Light attenuation display: Subtractive see-through near-eye display via spatial color filtering. *IEEE Transactions on Visualization and Computer Graphics, 25*(5), 1951–1960.
- Itoh, Y., Langlotz, T., Sutton, J., & Plopski, A. (2021). Towards indistinguishable augmented reality: A survey on optical see-through head-mounted displays. *ACM Computing Surveys (CSUR), 54*(6), 1–36.
- Jacobsen, A., & Gilchrist, A. (1988). The ratio principle holds over a million-to-one range of illumination. *Perception & Psychophysics, 43*, 1–6.
- Jung, S., & Lindeman, R. W. (2021). Does realism improve presence in VR? Suggesting a model and metric for VR experience evaluation. *Frontiers in Virtual Reality, 2*(693327), 1–7.
- Katz, D. (1935). *The world of touch*. London: Kegan Paul, Trench, Trubner; Co.
- Kim, M., Gold, J. M., & Murray, R. F. (2018). What image features guide lightness perception? *Journal of Vision, 18*(13), 1–1.
- 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.
- Kober, S. E., Kurzmam, J., & Neuper, C. (2012). Cortical correlate of spatial presence in 2d and 3d interactive virtual reality: An eeg study. *International Journal of Psychophysiology*, 83(3), 365–374.
- Koenderink, J. J., Pont, S. C., Van Doorn, A. J., Kappers, A. M., & Todd, J. T. (2007). The visual light field. *Perception*, 36(11), 1595–1610.
- 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. (1977). The retinex theory of color vision. *Scientific American*, 237(6), 108–129.
- Land & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America A*, 61(1), 1–11.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35(3), 389–412.
- Li, Z., Shafiei, M., Ramamoorthi, R., Sunkavalli, K., & Chandraker, M. (2020). Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2475–2484.
- Lin, J.-W., Duh, H. B.-L., Parker, D. E., Abi-Rached, H., & Furness, T. A. (2002). Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings IEEE Virtual Reality 2002*, 164–171.

- Linowes, J. (2020). *Unity 2020 virtual reality projects: Learn vr development by building immersive applications and games with unity 2019.4 and later versions*. Packt Publishing Ltd.
- Maloney, L. T. (1999). Physics-based approaches to modeling surface color perception. *Color vision: From genes to perception*, 387–416.
- Marantz, A. (2016). Studio 360. *The New Yorker*, 86–94.
- McCluney, R. (1994). *Introduction to radiometry and photometry*. Artech House, Inc.
- 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), 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.
- 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., Patel, K. Y., & Wiedenmann, E. S. (2022). Luminance calibration of virtual reality displays in unity. *Journal of Vision*, 22(13):1, 1–9.
- 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., Brainard, D. H., Flachot, A., & Patel, J. Y. (2023). An intrinsic image network evaluated as a model of human lightness perception. *Electronic Imaging*, 35(10), 1–6.

- Noguchi, K., & Masuda, N. (1971). Brightness changes in a complex field with changing illumination a re-examination of Jameson and Hurvich's study of brightness constancy. *Japanese Psychological Research*, 13(2), 60–69.
- Ong, T., Wilczewski, H., Soni, H., Nisbet, Q., Paige, S. R., Barrera, J., Welch, B. M., & Bunnell, B. (2022). The symbiosis of virtual reality exposure therapy and telemental health: A review. *Frontiers in Virtual Reality*, 3(848066), 1–11.
- Ostrovsky, Y., Cavanagh, P., & Sinha, P. (2005). Perceiving illumination inconsistencies in scenes. *Perception*, 34(11), 1301–1314.
- Oyelere, S. S., Bouali, N., Kaliisa, R., Obaido, G., Yunusa, A. A., & Jimoh, E. R. (2020). Exploring the trends of educational virtual reality games: A systematic review of empirical studies. *Smart Learning Environments*, 7, 1–22.
- Patel, J., Flachot, A., Vazquez-Corral, J., Brainard, D. H., Wallis, T. S., Brubaker, M. A., & Murray, R. F. (2023). A deep convolutional neural network trained to infer surface reflectance is deceived by mid-level lightness illusions. *Journal of Vision*, 23(9), 4817.
- Patel, K. Y., Munasinghe, A., & Murray, R. (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., Singh, S., & Murray, R. F. (submitted:BR-ORG-24-359). An equivalent illuminant analysis of lightness constancy with physical objects and in virtual reality. *Behavior Research Methods*.
- Patel, K. Y., Wilcox, L. M., Maloney, L. T., Ehinger, K. A., Patel, J. Y., Wiedenmann, E., & Murray, R. F. (2024). Lightness constancy in reality, in virtual reality, and on flat-panel displays. *Behavior Research Methods*, 1–19.
- Pessoa, L., Mingolla, E., & Arend, L. E. (1996). The perception of lightness in 3-d curved objects. *Perception & Psychophysics*, 58(8), 1293–1305.
- 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.
- Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 103778.
- Radonjic, A., & Gilchrist, A. L. (2014). Lightness perception in simple images: Testing the anchoring rules. *Journal of Vision*, *14*(13), 1–13.
- Radonjic, A., Pearce, B., Aston, S., Krieger, A., Dubin, H., Cottaris, N. P., Brainard, D. H., & Hurlbert, A. C. (2016). Illumination discrimination in real and simulated scenes. *Journal of Vision*, *16*(11), 1–18.
- Redding, G. M., & Lester, C. F. (1980). Achromatic color matching as a function of apparent target orientation, target and background luminance, and lightness or brightness instructions. *Perception & Psychophysics*, *27*(6), 557–563.
- 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.
- Rodriguez, R. G., 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.
- Rojas-Sánchez, M. A., Palos-Sánchez, P. R., & Folgado-Fernández, J. A. (2023). Systematic literature review and bibliometric analysis on virtual reality and education. *Education and Information Technologies*, *28*(1), 155–192.

- Rother, C., Kiefel, M., Zhang, L., Schölkopf, B., & Gehler, P. (2011). Recovering intrinsic images with a global sparsity prior on reflectance. *Advances in Neural Information Processing Systems*, 24.
- Rudd, M. E., & Zemach, I. K. (2005). The highest luminance anchoring rule in achromatic color perception: Some counterexamples and an alternative theory. *Journal of Vision*, 5(11), 983–1003.
- Ruddle, R. A., Volkova, E., & Bühlhoff, H. H. (2011). Walking improves your cognitive map in environments that are large-scale and large in extent. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 18(2), 1–20.
- 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.
- Sanchez-Vives, M. V., & Slater, M. (2005). From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4), 332–339.
- 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), 1–11.
- Scarfe, P., & Glennerster, A. (2019). The science behind virtual reality displays. *Annual Review of Vision Science*, 5, 529–547.
- 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.
- Shen, L., & Yeo, C. (2011). Intrinsic images decomposition using a local and global sparse representation of reflectance. *CVPR 2011*, 697–704.
- Shevell, S. K., & Kingdom, F. A. (2008). Color in complex scenes. *Annual Review of Psychology*, 59, 143–166.

- Shi, J., Dong, Y., Su, H., & Yu, S. X. (2017). Learning non-lambertian object intrinsics across shapenet categories. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 1685–1694.
- Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology*, 109(3), 431–433.
- Slater, M., Lotto, B., Arnold, M. M., & Sánchez-Vives, M. V. (2009). How we experience immersive virtual environments: The concept of presence and its measurement. *Anuario de Psicología*, 40, 193–210.
- Sutherland, I. E. (1965). The ultimate display. *Proceedings of the IFIP Congress*, 2(506–508), 506–508.
- Thouless, R. (1931). Phenomenal regression to the real object. *British Journal of Psychology*, 21(4), 339–359.
- Toscani, M., Zdravković, S., & Gegenfurtner, K. R. (2016). Lightness perception for surfaces moving through different illumination levels. *Journal of Vision*, 16(15), 1–18.
- Unity Technologies. (2020). Unity, version 2019.3.3f1. <https://docs.unity3d.com/2020.1/Documentation/Manual/UnityManual.html>
- Unity Technologies. (2022). Unity, version 2022.3 (lts). <https://docs.unity3d.com/2022.3/Documentation/Manual/UnityManual.html>
- Von Helmholtz, H. (1867). *Handbuch der physiologischen optik: Mit 213 in den text eingedruckten holzschnitten und 11 tafeln* (Vol. 9). Voss.
- Von Helmholtz, H. (1925). *Helmholtz's treatise on physiological optics*. Optical Society of America.
- Wallach, H. (1948). Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology*, 38(3), 310.
- Wallach, H. (1976). On perception. new york: Quadrangle. *New York Times Book Co.*
- Wang, Z., Liu, Y., & Lu, F. (2023). Discriminative feature encoding for intrinsic image decomposition. *Computational Visual Media*, 1–22.

- Westerman, S. J., & Cribbin, T. (1998). Individual differences in the use of depth cues: Implications for computer- and video-based tasks. *Acta Psychologica*, 99, 293–310.
- Wilder, J. D., Adams, W. J., & Murray, R. F. (2019). Shape from shading under inconsistent illumination. *Journal of Vision*, 19(6):2. <https://doi.org/10.1167/19.6.2>
- Wilson, C. J., & Soranzo, A. (2015). The use of virtual reality in psychology: A case study in visual perception. *Computational and mathematical methods in medicine*, 2015.
- Xia, L., Pont, S., & Heynderick, I. (2017). Separate and simultaneous adjustment of light qualities in a real scene. *i-Perception*, 8(1), 2041669516686089.
- Xia, L., Pont, S. C., & Heynderickx, I. (2014). The visual light field in real scenes. *i-Perception*, 5(7), 613–629.
- Xie, B., Alghofalli, R., Zhang, Y., Jiang, Y., Lobo, F. D., Li, C., Li, W., Huang, H., Akdere, M., Mousas, C., & Yu, L.-F. (2021). A review on virtual reality skill training applications. *Frontiers in Virtual Reality*, 2(645153), 1–19.
- Xiong, J., Hsiang, E.-L., He, Z., Zhan, T., & Wu, S.-T. (2021). Augmented reality and virtual reality displays: Emerging technologies and future perspectives. *Light: Science & Applications*, 10(216), 1–30.
- Zhan, T., Yin, K., Xiong, J., He, Z., & Wu, S.-T. (2020). Augmented reality and virtual reality displays: Perspectives and challenges. *iScience*, 23(101397), 1–13.
- Zhang, Y., Liu, H., Kang, S.-C., & Al-Hussein, M. (2020). Virtual reality applications for the built environment: Research trends and opportunities. *Automation in Construction*, 118, 103311.
- Zimmerman, T. G., Lanier, J., Blanchard, C., Bryson, S., & Harvill, Y. (1986). A hand gesture interface device. *ACM Sigchi Bulletin*, 18(4), 189–192.
- Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer*, 38(9), 25–32.