Human Photoreceptor Signals

Allie Charlotte Hexley

A thesis submitted for the degree of

Doctor of Philosophy



Pembroke College

Department of Experimental Psychology

University of Oxford

Michaelmas 2021

Abstract

The first stage in the human response to light is the sensing of light by the five classes of photoreceptors in the retina: the three cone classes (short- [S], medium- [M], and long- [L] wavelength sensitive), the rods, and the melanopsin-containing intrinsically photosensitive retinal ganglion cells (ipRGCs). Signals from these five photoreceptor classes drive the full range of visual and non-visual light-mediated neural mechanisms, from the perception of movement and colour, to the pupillary light reflex and circadian entrainment. In traditional models of the function and interaction between these photoreceptors, cones are thought to be the primary drivers of daytime vision, rods the drivers of visual functions in low-light levels, while ipRGCs feed into the internal clock, driving the non-visual effects of light.

Such accounts may give the impression that understanding the retinal inputs to light mediated mechanisms is a fully solved problem, yet there are still many unanswered questions. For example, while cones and ipRGCs are known to both feed into control of the pupillary light reflex, little is understood about the post-receptoral pathways that combine cone and ipRGC input, and whether interactions between cones and ipRGCs feed into other light-mediated mechanisms. Models of retinal inputs to vision also fail to take into account the large degree of individual variation between observers, particularly in the case of colour vision. Furthermore, it is unclear how our visual system is able to solve the computational challenges of using the filtered and downsampled information conveyed by the photoreceptor signals as cues to guide our interactions with the external world.

This thesis presents a range of new methodologies, both in terms of custom-built equipment and experimental design, and novel analytical approaches, to study, characterize, and quantify retinal inputs in ways that have previously been unachievable. It presents new technical capabilities in stimulus generation and control to enable new tests of the photoreceptor inputs to vision. Overall, the primary focus of this thesis is the novel characterization of the biological and physiological factors that influence the first stages of vision and non-visual light mediated neural mechanisms.

List of Abbreviations

S Short-wavelength-sensitive (cone)

M Middle-wavelength-sensitive (cone)

L Long-wavelength-sensitive (cone)

ipRGC Intrinsically photosensitive retinal ganglion cell

RGC Retinal ganglion cell

CIE Commission Internationale de l'Eclairage/International Commission on Illumination

CIE06 CIE 170-1:2006 Technical Report

CIE S026 CIE S 026/E:2018 Technical Report

JND Just noticeable difference

LGN Lateral geniculate nucleus

SCN Suprachiasmatic nucleus

OPN Olivary pretectal nucleus

ERG Electroretinogram

CMFs Colour matching functions

MPHDR Multi-primary high dynamic range

HFP Heterochromatic flicker photometry

SNP Single-nucleotide polymorphism

L(S180) Carrier of serine at codon 180 of exon 3 of *OPN1LW*

L(A180) Carrier of alanine at codon 180 of exon 3 of *OPN1LW*

L(S180A) Serine to alanine substitution at codon 180 of exon 3 of *OPN1LW*

CVD Colour vision deficiency

LCR Locus control region

NGS Next generation sequencing

PCR Polymerase chain reaction

gnomAD Genome Aggregation Database

LED Light emitting diode

FWHM Full-width half maximum

HRR Hardy-Rand-Rittler (pseudoisochromatic test for colour vision)

ANOVA Analysis of variance

CNN Convolutional neural network

CCT Correlated colour temperature

D65 Daylight illuminant at 6500K

SOCS Standard object color spectra

MCC Matthew's correlation coefficient

CRT Cathode-ray tube

EOTF Electro-optical transfer function

LCD Liquid-crystal display

OLED Organic light emitting diodes

LUT Look-up table

FFT Fast Fourier transform

DPI Dots per inch

MTF Modulation transfer function

PSF Point spread function

OTF Optical transfer function

CSF Contrast sensitivity function

HDR High dynamic range

DLP Digital light processing

ANSI American National Standards Institute

IES Illuminating Engineering Society of North America

PSRM Photoreceptor signal reproduction metric

PSDM Photoreceptor signal distortion metric

PCDM Photoreceptor correlation distortion metric

VDU Visual display unit

AO-SLO Adaptive optics scanning laser ophthalmoscopy

AO-OCT Adaptive optics optical coherence tomography

Acknowledgements

This thesis was made possible thanks to funding received under the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 765911 (RealVision).

Firstly, I must say a massive thank you to my supervisor, Hannah Smithson, for her continued support of my work. I am particularly grateful for the many opportunities she has given me, and for always making time to help me navigate the academic career pathway as an early career researcher. I've thoroughly enjoyed our impassioned conversations about various projects and am looking forward to continuing working together in the future.

I'd also like to thank my co-supervisor Manuel Spitschan, for creating many invaluable opportunities for me at this early career stage, and for his patience and willingness to help with my many science and career questions, and Laura Young, for acting as a bonus mentor to me during these past three years. In fact I must send my thanks to the entirety of the Oxford Perception Lab, past and present, for being such wonderful colleagues and true pleasures to work with. In particular, I must thank Takuma Morimoto, for setting an excellent example of what an early career researcher should be and for many enthusiastic discussions about colour vision; and Sarah Regan, for being the first person to show me how to run a psychophysics experiment, and teaching me more than she realises she did. I give my specific thanks to the many collaborators I have had the joy to work on the projects contained in this thesis with in the relevant chapter acknowledgements.

I offer perhaps my biggest thanks to my friends and family, without who's support I would not even have begun the DPhil journey in the first place. To all of my Pembroke MCR friends, thank you for making the past three years some of the best of my life, despite over half of those being in a global pandemic. To Hannah and Ariana, thank you for helping me stay sane during the first lockdown (or perhaps I should more aptly say, for going a little bit crazy with me); and to Hannah and Cillian, thank you for the same during all of the turbulence since. To my brother, Adam, it's been a joy to live in the same city again during the past three years. To my cousin, Phil, thank you for being my first inspiration, role model, and mentor. And, to my parents, Colin and Lisa, thank you for always being my number one supporters, I hope I have and will continue to make you proud.

Finally, to Constantin: I love you, and I could write a whole other thesis on the thanks I owe to you.

Contents

1 Introduction **23**

1.1 Introduction . . . 24

1.2 Human photoreceptor signals . . . 25

1.3 Cones, rods, and ipRGCs . . . 26

1.4 Individual differences in human photoreceptor signals . . . 30

1.4.1 The standard observer . . . 30

1.4.2 Variability between individuals . . . 32

1.4.3 Quantifying variation: the CIE06 model . . . 34

1.4.4 Variability in rods and ipRGCs . . . 38

1.4.5 Individual differences in luminous efficiency . . . 39

1.4.6 Modelling individual differences . . . 39

1.5 Photoreceptor inputs to vision . . . 39

1.5.1 Photopic vision . . . 40

1.5.2 Scotopic vision . . . 42

1.5.3 Mesopic vision . . . 43

1.5.4 ipRGC inputs to vision . . . 44

1.5.5 Vision beyond photoreceptor signals . . . 46

1.6 Photoreceptor inputs to non-visual light mediated neural mechanisms . . . 47

1.6.1 Circadian rhythm . . . 47

1.6.2 Melatonin suppression . . . 48

1.6.3 Pupil size . . . 48

1.7 Methods of studying photoreceptor signals . . . 48

1.8 Colorimetry . . . 52

1.8.1 CIE 1931 XYZ . . . 52

1.8.2 CIE 1931 xy chromaticity diagram . . . 54
1.8.3 CIELUV . . . 54
1.8.4 CIELAB and ΔE . . . 55
1.8.5 MacLeod-Boynton Chromaticity Diagram . . . 56
1.9 Thesis overview . . . 57
1.9.1 Chapter 2: Individual differences in cone photoreceptor signals . . . 57
1.9.2 Chapter 3: Exploring the statistical regularities in photoreceptor signals from surfaces for illuminant perception with machine-learning . . . 57
1.9.3 Chapter 4: Online visual psychophysics: is it feasible to study photoreceptor signals with online experiments? . . . 58
1.9.4 Chapter 5: Demonstrating a multi-primary high-dynamic range display for photoreceptor isolating experiments . . . 58
1.9.5 Chapter 6: Novel metrics for characterizing the reproduction of photoreceptor signals . . . 59
1.10 Acknowledgements . . . 59

2 Individual differences in cone photoreceptor signals 61

2.1 Introduction . . . 62
2.1.1 Individual differences in cone spectral sensitivities . . . 62
2.1.2 Individual differences in L:M ratio . . . 64
2.1.3 Systematic differences between individuals . . . 66
2.1.4 Perceptual consequences of individual differences . . . 73
2.1.5 The scope of studying individual differences . . . 77
2.2 Experiment 1: Ethnic differences in the *OPN1LW* L(S180) vs L(A180) allele frequency . . 77
2.2.1 Introduction . . . 77
2.2.2 Methods . . . 80
2.2.3 Results . . . 81
2.2.4 Discussion . . . 83
2.3 Experiment 2: Systematic differences in L:M ratio . . . 85
2.3.1 Introduction . . . 85
2.3.2 Methods . . . 88
2.3.3 Results . . . 92

2.3.4 Discussion . . . 96
2.4 General Discussion . . . 97
2.5 Conclusion . . . 101
2.6 Acknowledgements . . . 102

3 Exploring the statistical regularities in photoreceptor signals from surfaces for illuminant estimation with machine learning 103

3.1 Introduction . . . 104
3.1.1 The computational challenge of illuminant estimation . . . 105
3.1.2 Illuminant perception thresholds . . . 106
3.1.3 Illuminant perception and colour constancy . . . 107
3.1.4 Illuminant perception as an environmental cue . . . 109
3.1.5 Illuminant estimation for computer vision . . . 109
3.1.6 Cues to the illuminant . . . 110
3.2 Experiment 1: Base Case . . . 111
3.2.1 Methods . . . 112
3.2.2 Results . . . 118
3.2.3 Discussion . . . 118
3.3 Experiment 2: Varying Illuminant Dataset . . . 119
3.3.1 Methods . . . 120
3.3.2 Results . . . 121
3.3.3 Discussion . . . 123
3.4 Experiment 3: Comparing surface statistics . . . 123
3.4.1 Methods . . . 123
3.4.2 Results . . . 124
3.4.3 Discussion . . . 125
3.5 Experiment 4: Strategies for combining multiple surfaces . . . 127
3.5.1 Methods . . . 127
3.5.2 Results . . . 128
3.5.3 Discussion . . . 129
3.6 Experiment 5: Availability of photoreceptor signals . . . 130
3.6.1 Methods . . . 130

3.6.2 Results . . . 130
3.6.3 Discussion . . . 131
3.7 General Discussion . . . 132
3.7.1 Connection to human illuminant perception . . . 132
3.7.2 Limitations . . . 133
3.7.3 Further directions with the machine learning approach . . . 134
3.8 Conclusion . . . 135
3.9 Acknowledgements . . . 135

4 Online visual psychophysics: is it feasible to study photoreceptor signals with online experiments? 137

4.1 Introduction . . . 138
4.2 The stages of display calibration and their impact on stimulus control . . . 140
4.2.1 Linearity of display output with input drive-signal . . . 140
4.2.2 Chromaticity of stimuli . . . 144
4.2.3 Illumination levels . . . 152
4.2.4 Spatial characteristics . . . 154
4.2.5 Timing . . . 158
4.2.6 Further Considerations . . . 160
4.2.7 Display calibration checklist . . . 161
4.3 Review of existing methods of perceptual calibration . . . 162
4.3.1 Linearity of display output with input drive . . . 163
4.3.2 Chromaticity of stimuli . . . 169
4.3.3 Illumination levels . . . 174
4.3.4 Spatial characteristics . . . 176
4.3.5 Timing . . . 178
4.3.6 Further Considerations . . . 179
4.3.7 Perceptual Calibration Checklist . . . 180
4.4 Discussion . . . 181
4.5 Conclusion . . . 184
4.6 Acknowledgements . . . 184

5 Demonstrating a multi-primary high-dynamic range display for photoreceptor isolating experiments 187

5.1 Introduction . . . 188

5.1.1 Melanopsin isolating experiments . . . 188

5.1.2 Cone and rod isolating experiments . . . 190

5.1.3 The MPHDR display . . . 191

5.2 Methods . . . 191

5.2.1 Hardware overview . . . 191

5.2.2 Calibration . . . 197

5.2.3 Display algorithm . . . 198

5.2.4 Designing photoreceptor isolating stimuli . . . 201

5.3 Results . . . 203

5.3.1 High-dynamic range and high luminance output . . . 204

5.3.2 Expanded colour gamut . . . 205

5.3.3 High-contrast photoreceptor-isolating stimuli . . . 207

5.4 Discussion . . . 209

5.5 Future development . . . 211

5.6 Conclusions . . . 214

5.7 Acknowledgements . . . 214

6 Novel metrics for characterizing the reproduction of photoreceptor signals 215

6.1 Introduction . . . 216

6.1.1 The human physiological response to light . . . 216

6.1.2 The reproduction of light by displays . . . 218

6.1.3 Towards more accurate physiologically relevant reproduction of light on displays 221

6.1.4 Quantifying a display's physiologically relevant reproduction of light . . . 222

6.2 Methods: Introducing three novel metrics and an equal-luminance photoreceptor excitation diagram . . . 223

6.2.1 Metric I: Photoreceptor Signal Reproduction Metric (PSRM) . . . 225

6.2.2 Metric II: Photoreceptor Signal Distortion Metric (PSDM) . . . 226

6.2.3 Metric III: Photoreceptor Correlation Distortion Metric (PCDM) . . . 227

6.2.4 Equal-luminance photoreceptor excitation diagram . . . 229

6.3 Results: Example applications of novel metrics . . . 231
6.3.1 Photoreceptor Signal Reproduction Metric (PSRM) . . . 232
6.3.2 Photoreceptor Signal Distortion Metric (PSDM) . . . 234
6.3.3 Photoreceptor Correlation Distortion Metric (PCDM) . . . 237
6.3.4 Equal-luminance photoreceptor excitation visualizations . . . 237
6.4 Discussion . . . 240
6.4.1 The applicability of the proposed metrics . . . 240
6.4.2 Limitations of proposed metrics . . . 241
6.4.3 Applying the metrics and visualization to multi-primary displays . . . 244
6.5 Conclusion . . . 245
6.6 Acknowledgements . . . 245

7 General discussion **247**

7.1 Chapter summaries . . . 248
7.1.1 Chapter 2: Individual differences in cone photoreceptor signals . . . 248
7.1.2 Chapter 3: Exploring the statistical regularities in photoreceptor signals from surfaces for illuminant estimation with machine-learning . . . 248
7.1.3 Chapter 4: Online visual psychophysics: is it feasible to study photoreceptor signals with online experiments? . . . 249
7.1.4 Chapter 5: Demonstrating a multi-primary high-dynamic range display for photoreceptor isolating experiments . . . 250
7.1.5 Chapter 6: Novel metrics for characterizing the reproduction of photoreceptor signals . . . 250
7.2 Future directions . . . 251
7.2.1 The genetic underpinnings of individual differences in L:M ratio . . . 251
7.2.2 Rod and ipRGC contributions to illuminant estimation . . . 251
7.2.3 Utilising the MPHDR display to study the photoreceptor signals . . . 252
7.2.4 Functional imaging of the human photoreceptor signals *in vivo* . . . 253
7.2.5 Dynamics of the human photoreceptor signals . . . 253
7.2.6 A complete model of the human photoreceptor signals . . . 254
7.3 Thesis summary . . . 254

List of Figures

1.1 The human photoreceptor signals 26
1.2 The human eye 27
1.3 Low-optical density absorbance spectra of cone photopigments 35
1.4 Optical density of cone photopigments as a function of field size 36
1.5 Macular pigment density as a function of field size 37
1.6 Lens density as a function of age 38
1.7 Contextual effect on colour perception 46
1.8 Colour constancy example 47
1.9 The principle of silent substitution 50

2.1 Example differences in the cone spectral sensitivities 63
2.2 Variability in the cone spectral sensitivities over a population of observers 64
2.3 Individual differences in L:M ratio 65
2.4 Heterochromatic flicker photometry (HFP) set-up 90
2.5 HFP settings across sessions and the participant cohort 93
2.6 L:M ratio estimates from HFP settings 94
2.7 Biological sex differences in HFP setting 94
2.8 Ethnicity differences in HFP setting 95
2.9 Season-of-birth differences in HFP setting 96

3.1 Schematic of the computational problem of illuminant estimation 106
3.2 Illuminants used in the base case 112
3.3 Reference dataset used in the base case 113
3.4 Schematic of training and testing the random forest classifier 115
3.5 Classifier evaluation visualizations 118
3.6 Classifier performance on Experiment 1 119

3.7 Classifier performance on Experiment 2 . . . 121
3.8 Classifier performance on Experiment 3a . . . 125
3.9 Classifier performance on Experiment 3b . . . 126
3.10 Classifier performance on Experiment 4 . . . 129
3.11 Classifier performance on Experiment 5 . . . 131

4.1 Simulation of impact of misestimating a display's gamma . . . 142
4.2 Simulation of impact of ignoring colorimetric calibration on hypothetical displays . . . 149
4.3 Simulation of impact of ignoring colorimetric calibration on example real displays . . . 150
4.4 Absolute luminance of displays and its effect on pupil size . . . 152
4.5 Spatial resolution considerations of stimuli delivered on a display . . . 157
4.6 Temporal responses of CRT and LCD . . . 159
4.7 Process for the systematic literature review of perceptual calibration methods . . . 163
4.8 Example of stimulus used in perceptual linearization task . . . 165
4.9 Examples of stimuli for perceptual colorimetric calibration task . . . 171

5.1 Schematic of MPHDR . . . 192
5.2 MPHDR hardware design . . . 194
5.3 Schematic of the driving algorithm of the MPHDR display . . . 199
5.4 Six effective primaries of the MPHDR display . . . 201
5.5 Display response curve for the MPHDR . . . 204
5.6 Colour gamut of the MPDHR display . . . 208
5.7 3D colour gamut of the MPHDR display . . . 209
5.8 Melanopsin isolating stimuli on the MPHDR . . . 209

6.1 An example of how a display can reproduce photoreceptor signals . . . 220
6.2 Real-world photoreceptor excitation correlations . . . 228
6.3 An extended equal-luminance photoreceptor excitation space . . . 230
6.4 Primaries of displays used in example applications of metrics . . . 232
6.5 Chromaticity reproduction and *PSRM* . . . 233
6.6 *PSDM* of an example three-primary display . . . 235
6.7 *PSDM* of two more example three-primary displays . . . 236
6.8 *PCDM* of the example three-primary displays . . . 237
6.9 Photoreceptor excitation diagram for the example three-primary displays . . . 238

6.10 Photoreceptor excitation diagram for the example five-primary displays 239

List of Tables

1.1 Table summarizing the structural and functional features of the different photoreceptor classes 28

2.1 Table summarizing the L(S180) vs L(A180) allele frequency estimates from various past studies 81

2.2 Table summarizing the *OPN1LW* L(S180) vs L(A180) allele frequency across ethnicity in the gnomAD database 82

2.3 Table summarizing the L:M ratio estimates from various past studies 87

3.1 Table of results of base case illuminant classifier 118

3.2 Table of results of classifier performance when number of illuminant classes is varied . 122

4.1 Screen specifications for example displays 155

5.1 Hardware components of the MPHDR 193

5.2 Table of global and local contrast range of the MPHDR 206

5.3 Table comparing the features of different display systems 211

Chapter 1

Introduction

Contents

1.1 Introduction 24

1.2 Human photoreceptor signals 25

1.3 Cones, rods, and ipRGCs 26

1.4 Individual differences in human photoreceptor signals 30

1.4.1 The standard observer 30

1.4.2 Variability between individuals 32

1.4.3 Quantifying variation: the CIE06 model 34

1.4.4 Variability in rods and ipRGCs 38

1.4.5 Individual differences in luminous efficiency 39

1.4.6 Modelling individual differences 39

1.5 Photoreceptor inputs to vision 39

1.5.1 Photopic vision 40

1.5.2 Scotopic vision 42

1.5.3 Mesopic vision 43

1.5.4 ipRGC inputs to vision 44

1.5.5 Vision beyond photoreceptor signals 46

1.6 Photoreceptor inputs to non-visual light mediated neural mechanisms 47

1.6.1 Circadian rhythm 47

1.6.2 Melatonin suppression 48

1.6.3 Pupil size 48

1.7 Methods of studying photoreceptor signals 48

1.8 Colorimetry 52
1.8.1 CIE 1931 XYZ 52
1.8.2 CIE 1931 xy chromaticity diagram 54
1.8.3 CIELUV 54
1.8.4 CIELAB and ΔE 55
1.8.5 MacLeod-Boynton Chromaticity Diagram 56
1.9 Thesis overview 57
1.9.1 Chapter 2: Individual differences in cone photoreceptor signals 57
1.9.2 Chapter 3: Exploring the statistical regularities in photoreceptor signals from surfaces for illuminant perception with machine-learning 57
1.9.3 Chapter 4: Online visual psychophysics: is it feasible to study photoreceptor signals with online experiments? 58
1.9.4 Chapter 5: Demonstrating a multi-primary high-dynamic range display for photoreceptor isolating experiments 58
1.9.5 Chapter 6: Novel metrics for characterizing the reproduction of photoreceptor signals 59
1.10 Acknowledgements 59

1.1 Introduction

All sensory processing of light in the external world is filtered through the photoreceptors. The first stage of neural processing of light, phototransduction, occurs when a photon of light is absorbed by the photopigment contained within the photoreceptor, beginning a cascade that ultimately results in an electrical signal in the cell, which can go on to trigger subsequent neural signalling. Without the photoreceptors we would be incapable of sensing light, and of activating the biological processes triggered by light.

The human photoreceptor signals are the topic of this thesis. Individual differences in the photoreceptor signals are discussed, as are some of the computational challenges the visual system faces. This thesis also focuses on methods through which the photoreceptor signals can be studied, and presents new technical capabilities in stimulus generation and control that will enable new tests of the photoreceptor inputs to light-mediated responses, as well as novel quantifications of the ability of stimulus presentation devices, such as displays, to reproduce the photoreceptor signals.

In this chapter, I will introduce the background material necessary for understanding the experimental chapters in this thesis. I will summarise the relevant aspects of our current understanding of the human photoreceptor signals and highlight the gaps in the current state of the field: some of which the experimental work presented in this thesis will begin to fill.

1.2 Human photoreceptor signals

Physically, light is electromagnetic radiation, which can be parameterised in terms of wavelength. Most of the light in the external world is comprised of a broad radiance spectrum, $P(\lambda)$, of different wavelengths of light at varying intensities. An example of such a radiance spectrum is shown in Fig. 1.1(a). Humans have five known photoreceptor types with which they can sense light: the three classes of cones, the rods, and the intrinsically photosensitive retinal ganglion cells (ipRGCs), which contain the photopigment melanopsin.

Each photoreceptor type has a known spectral sensitivity, $T_k(\lambda)$, which specifies the probability of that photoreceptor type absorbing light of a given wavelength. The spectral sensitivities of the five photoreceptors are depicted in Fig. 1.1(b) (Stockman et al., 1999; Stockman & Sharpe, 2000; Wyszecki & Stiles, 1982; Lucas et al., 2014). The three cone classes differ from each other in terms of the photopigment they express, and, as a result, each of the three cone classes are maximally sensitive to either short wavelengths (S), middle wavelengths (M), or long wavelengths (L). Conversely, there is only one type of rod (R) and melanopsin (I) photopigment. By integrating the spectral sensitivities of each photoreceptor type multiplied by the radiance spectrum of the physical light, one can calculate the signals, A_k, of each photoreceptor type, as in Eq. 1.1, and depicted graphically in Fig. 1.1(c):

$$A_k = \int_\lambda P(\lambda) \cdot T_k(\lambda)\, d\lambda$$
$$\text{where} \tag{1.1}$$
$$k \in [S, M, L, R, I].$$

The photoreceptor signals put fundamental limits on the human ability to sense light. One such limit is the range of wavelengths of light to which we are sensitive. We can only see light within the visible wavelength range (roughly 400-700 nm) because beyond that the probability that any of the photoreceptors in the eye may absorb a photon of light is essentially zero. Additionally, one can see that the broad physical radiance spectrum is reduced to a five-dimensional signal space by the photoreceptors. In other words, the photoreceptors downsample the external radiance spectrum, and

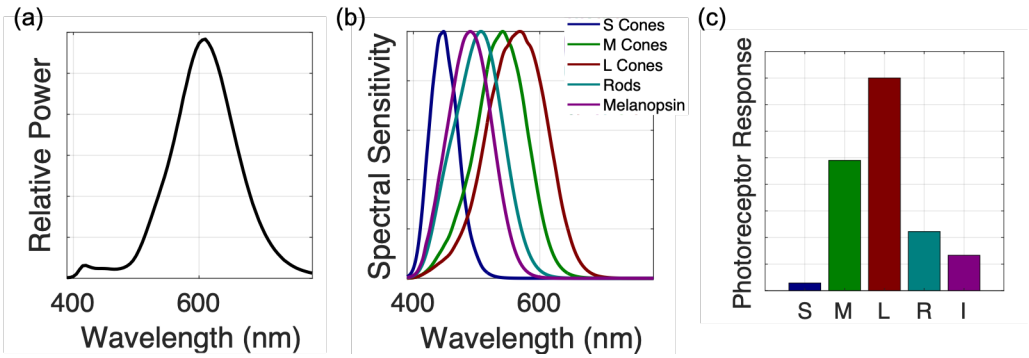


Figure 1.1: Figure showing how external light is sampled by the photoreceptors to generate the human photoreceptor signals. a) An example of a physical light spectrum. b) The spectral sensitivity functions of the five photoreceptors for the standard observer (Stockman et al., 1999; Stockman & Sharpe, 2000; Wyszecki & Stiles, 1982; Lucas et al., 2014). c) The relative responses of the five photoreceptors are calculated by integrating the product of the external light spectrum with the spectral sensitivities.

it is impossible for subsequent neural processing to reconstruct the complete radiance spectrum. All we can do is make comparisons between the photoreceptor signals and infer the possible external spectra that may have caused such patterns.

Due to the principle of univariance (Rushton, 1972), the signals of each photoreceptor class depend only on the number of photons absorbed. Considering the case where the external light is of just a single wavelength, it is impossible for the photoreceptor signals to differentiate between changes in intensity of the external light from changes in its wavelength. So too in the case of broad spectra of light can one find different physical radiance spectra that result in the same relative photoreceptor signals, and while thus physically distinct, the spectra are perceptually indistinguishable. In the case where the signals in the three cone types are equivalent between a pair of radiance spectra, the spectra are termed *metamers*.

1.3 Cones, rods, and ipRGCs

While all of the five photoreceptors share the same fundamental property of converting light into an electrical signal, they have notable differences from each other. This section provides a brief overview of the morphological, and functional differences between the photoreceptors. A summary of the key features is provided in Table 1.1 (Kolb, 2007, 2012b, 2012a; Do & Yau, 2010; Dartnall et al., 1983). The defining difference between the photoreceptors is the difference in spectral sensitivity, particularly the different wavelengths of peak sensitivity, as shown in Fig. 1.1(b).

Genetically, the S cone opsin is encoded for by *OPN1SW* which lies on chromosome 7. The L and M cone opsin genes, *OPN1LW* and *OPN1MW*, lie in tandem on the X chromosome (Nathans, Piantanida

et al., 1986; Nathans, Thomas et al., 1986; Vollrath et al., 1988). Rhodopsin, the photopigment in rods, is encoded for by the *RHO* gene which lies on chromosome 3 (Nathans, 1992), while melanopsin, the photopigment in ipRGCs, in encoded by *Opn4*, which is found on chromosome 10 (Provencio et al., 2000).

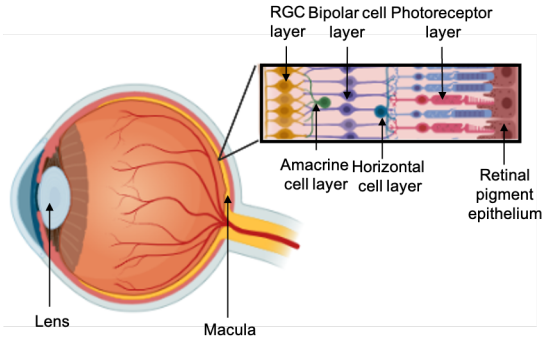


Figure 1.2: Figure showing the structure of the human eye. At the front of the eye is the lens, which focuses light onto the retina. The retina is at the back of the eye, and near the centre of the retina one finds the macula: the portion of the retina with a very high concentration of cone and rod photoreceptors which is responsible for central vision. The fovea, which has a diameter of ~5°, is the centre of the retina, the central ~2° of which is rod free, and where one finds the highest cone density. The foveola is the small pit found in the centre of the macula, which is about 1° in diameter and free from rods and capillaries. At the very centre of the foveola, one finds an S-cone free region (around 0.3° in diameter) (Polyak, 1941). At the retina, one finds many different cell layers, with the retinal ganglion cell (RGC) layer being the most anterior, then the amacrine cell layer, then the bipolar cell layer, then the horizontal cell layer, then the photoreceptor layer, and then finally, most posterior is the retinal pigment epithelium. The rods and cones are found in the photoreceptor layer, while the ipRGCs are found amongst the RGC layer. (Details on the anatomy of the eye and anatomy of the retina that informed the construction of this figure and figure caption were taken from Kolb, 2012b.)

Another key difference is in the location of the photoreceptors in the retina, as shown in Fig. 1.2. The cones and rods lie in the photoreceptor layer, at the back of the retina, while the ipRGCs lie amongst, as the name would suggest, the retinal ganglion cell (RGC) layer at the front of the retina (Kolb, 2012b; Graham & Wong, 2016). The cones and rods are much more plentiful than ipRGCs, with approximately 6-7 million cone cells in the human retina and 120 million rods, compared to around 1.2 million retinal ganglion cells, of which around 1 % contain melanopsin, resulting in just ~12,000 ipRGCs (Kolb, 2007; Kim et al., 2021; Münch & Kawasaki, 2013). While the cones are densely packed in the fovea, and sparser in the periphery, the rods are absent from the central ~2° of the fovea and more tightly packed in the periphery (Polyak, 1941). ipRGCs are absent from the fovea. ipRGCs are most abundant in the peri-fovea, paralleling the decrease in RGC density with increasing eccentricity (Mure, 2021).

The rods and cones also share many anatomical properties that differ from the anatomical properties of ipRGCs. For instance, the physical shape of the rods and cones is different than that of the ipRGCs. Rods and cones both have a long and thin anatomy, with the synapse terminal at the front

	Cones	**Rods**	**ipRGCs**
Photopigment	Cone opsin (long-, middle-, or short-wavelength-sensitive)	Rhodopsin	Melanopsin
Photopigment gene	*OPN1LW, OPN1MW, OPN1SW*	*RHO*	*Opn4*
Peak sensitivity of photopigment	~ 419 nm (S), ~ 530 nm (M), ~ 558 nm (L)	~ 498 nm	~ 480 nm
Retinal location	Photoreceptor layer	Photoreceptor layer	Retinal ganglion cell layer
Eccentricity	Densely packed in fovea, sparsely distributed in the periphery	Absent from the central ~ 2°, densely packed in the periphery	Absent from the fovea, most densely packed in the peri-fovea
Number	~6-7 million	~120 million	~12,000
Temporal resolution	High	Low	Very low
Sensitivity	Low	Very high	Low
Inputs	External light	External light	External light and rod and cone inputs
Dominant behaviour	Photopic and mesopic vision (including colour and spatial vision and motion perception)	Scotopic and mesopic vision	Non-visual light responses e.g. sleep-wake control, circadian entrainment, mood, alertness*

Table 1.1: Table summarizing the key morphological and functional attributes of the rods, cones, and ipRGCs (*prevailing view, there is growing evidence to suggest ipRGC contributions to the visual response, see Section 1.5.4). Most of the facts on this table concerning rods and cones are taken from Webvision (Kolb, 2007, 2012b, 2012a), while those concerning ipRGCs are from Do and Yau, 2010; Graham and Wong, 2016. The peak spectral sensitivity of the photopigment for the cones is from Dartnall et al., 1983. Any other references from which this information was drawn are indicated in the main text describing the relative feature.

tip of the cell, neighboured by the cell body containing the nucleus, followed by the inner segment containing the mitochondria and outer segment containing disks densely packed with the opsin (Kolb, 2012a). ipRGCs, on the other hand, have a central cell body and extensive dendritic networks (Graham & Wong, 2016).

The cones and rods do differ from each other in their anatomy with cones being more conical shaped and with shorter outer segments while rods are more cylindrical shaped with longer outer segments (Kolb, 2012a). In the periphery where both rods and cones co-exist, rods are smaller with a diameter of around 2 μm compared to cones with a diameter of around 6 μm. Foveal cones, however, are smaller than rods, with a diameter of around 1.5 μm, and are also longer than peripheral cones (Stricker & Power, 1872). This is an important point to note as it means the shape of the cones depends on retinal location, with cones being larger and sparser as one moves from the fovea to the periphery. It has been suggested that rod shape also depends on retinal eccentricity, but there is little data available

to quantify how, though it has been observed that rods are smaller closer to the fovea and larger in the periphery (Polyak, 1941).

Often the cones are grouped together when describing differences between the photoreceptor types, but it is important to note that S cones are actually quite distinct from L and M cones. S cones are completely absent from the very centre of the retina (approximately the central 0.3°), and make up only about 5-10 % of all cones in the retina (Polyak, 1941; Curcio et al., 1991; Castańo & Sperling, 1982). S cones also have shorter outer segments and longer inner segments than L or M cones (Kolb, 2012a).

ipRGCs are also typically thought of as just one subtype of photoreceptor, but in fact ipRGCs themselves have many distinct subtypes. Three subtypes of ipRGCs have been identified in humans: M1, M2, and M4 (Hannibal et al., 2017). Both the M2 and M4 subtypes are inner stratifying i.e. with dendritic branching towards the inner plexiform layer, while the M1 subtypes are outer stratifying (Hannibal et al., 2017). The different subtypes have slightly different anatomies, with the M1 ipRGCs having smaller somas and fewer dendritic branches than M2 ipRGCs (Kim et al., 2021). M4 cells also have smaller somas and fewer dendritic branches than M2 cells (Hannibal et al., 2017). M4 cells express very low levels of melanopsin compared to M1 and M2 cells (Hannibal et al., 2017; Estevez et al., 2012). Importantly the different ipRGC subtypes project to different downstream destinations, as described in Sections 1.5 and 1.6. M2/M4 cells much more abundant in the nasal retina, while M1 cells are more concentrated in the temporal retina (Hannibal et al., 2017).

A key difference in the light response between the ipRGCs, cones, and rods, is the operating range of each photoreceptor type. Rods are extremely sensitive and able to signal the absorption of a single photon of light (Hecht et al., 1942). Cones are much less sensitive than rods, and ipRGCs are the least sensitive of all the photoreceptors (Do & Yau, 2010; Lamb, 2016). However, the response of ipRGCs is long-lived, so sensitivity to continued light exposure is greater than sensitivity to a single flash of light (Do & Yau, 2010; Dacey et al., 2005). ipRGCs also exhibit much lower temporal resolution than rods or cones, with the cones having the highest temporal resolution of the three major photoreceptor classes (Kolb, 2012a; Do & Yau, 2010; Lamb, 2016).

Functionally, the rods, cones, and ipRGCs serve different purposes, innervating different downstream neural pathways. These are unpacked in detail in Section 1.5 and Section 1.6. Briefly, the rods mediate low-light vision, the cones enable colour, form, and motion perception under bright-light conditions, and the ipRGCs are responsible for driving the non-image forming responses to light. In addition to differences in the outputs, the photoreceptors also differ in terms of their inputs. By definition, they all receive input from external light, but ipRGCs are also innervated by rods and cones

(Dacey et al., 2005). Thus the electrical activity in ipRGCs will be in part due to the inputs from cones and rods, and in part due to the intrinsic phototransduction driven by the self-contained melanopsin. This is an important distinction as the melanopsin signal is not the same as the ipRGC signal, but is one of the input signals driving ipRGC activity. Another important distinction between ipRGCs vs rods and cones is that ipRGCs are *spiking* neurons, while rods and cones do not spike but rather respond via a graded potential (Do & Yau, 2010). Cones and rods are not devoid of communication between each other, and in fact it is well-established that rod signals can be carried through cones via rod-cone gap junctions (Hornstein et al., 2005).

To summarise, while similar in that the rods, cones, and ipRGCs all convert light to electrical signals, the different photoreceptors have notable structural and functional differences. The electrical signals from the photoreceptors are not as simple as the photoreceptor signals driven by the intrinsic light sensitivity of the photoreceptors as described in Section 1.2, due to the complex interactions between the photoreceptors at the photoreceptor level. While we are beginning to develop an understanding of the interactions between the photoreceptors and their inputs to downstream neural pathways (see Section 1.5 and 1.6) there is still a lot unknown about these mechanisms, and at present there is no quantitative model for these interactions.

1.4 Individual differences in human photoreceptor signals

1.4.1 The standard observer

Models of human photoreceptor signals are typically focused on developing a generalized description of retinal level processing across the human population. Such a "standard observer" model is widely used and does have practical advantages and applications such as in display design, where display manufacturers aim to reproduce the colours on a display to suit the average consumer. In terms of photoreceptor signals, a standard observer model aims to quantify two major features: 1) the spectral sensitivity of the photoreceptors; and 2) the luminous efficiency function, which describes the overall spectral sensitivity of the human eye. Luminous efficiency can be measured in both photopic and scotopic conditions.

The most widely used standard model of the spectral sensitivities of the photoreceptors is the CIE S026 standard, which includes, the Stockman-Sharpe 10° cone spectral sensitivity functions, $T_L(\lambda)$, $T_M(\lambda)$, $T_S(\lambda)$, the scotopic luminous efficiency function, $V'(\lambda)$, as the spectral sensitivity of the rods, $T_R(\lambda)$, and the melanopsin action spectrum as the spectral sensitivity of ipRGCs, $T_I(\lambda)$ (Fig. 1.1(b))

(Stockman et al., 1999; Stockman & Sharpe, 2000; Wyszecki & Stiles, 1982; Lucas et al., 2014; Stiles & Burch, 1959; CIE, 2006, 2018). The spectral sensitivity functions are measured at the level of the cornea.

The spectral sensitivity functions for the cones, rods, and ipRGCs have different experimental origins. The Stockman-Sharpe 10° cone spectral sensitivity functions are derived from a linear transformation of the Stiles and Burch, 1959, 10° colour matching functions (CMFs). The Stiles and Burch, 1959, 10° CMFs were measured from 49 participants, who completed a colour match across a 10-degree bipartite field, matching three primary wavelengths to a monochromatic target. To transform the Stiles and Burch, 1959, 10° CMFs into the 10° cone spectral sensitivities of Stockman-Sharpe, one must simply solve the linear equations in Eq. 1.2,

$$M_{10^\circ}\begin{bmatrix}\overline{r}(\lambda)\\ \overline{g}(\lambda)\\ \overline{b}(\lambda)\end{bmatrix}=\begin{bmatrix}l(\lambda)\\ m(\lambda)\\ s(\lambda)\end{bmatrix},$$

$$\text{where} \tag{1.2}$$

$$M_{10^\circ}=\begin{bmatrix}\overline{l_R} & \overline{l_G} & \overline{l_B}\\ \overline{m_R} & \overline{m_G} & \overline{m_B}\\ \overline{s_R} & \overline{s_G} & \overline{s_B}\end{bmatrix}.$$

To do so, Stockman and Sharpe, 2000, made spectral sensitivity measurements of dichromats (observers with only two functioning cone classes) of a known genotype. As the field size for the dichromat spectral sensitivity measurements was 2°, there is an intermediary step required to derive the 2° cone fundamentals. Thus, first, the matrix M_{2° is solved for the 2° cone fundamentals using the 2° Stiles and Burch, 1955, CMFs. In this derivation of the 2° cone fundamentals, the S cone sensitivity to the **R** primary is minimal, so $\overline{s_R}$ can be set to zero. For the L cone sensitivity, a weighting is applied to take the average of two prevalent polymorphic variants of the L cone gene (discussed in detail in Chapter 2). Once one has then obtained the 2° cone fundamentals by solving for M_{2° in the 2° case, the 10° fundamentals are constructed from fitting the 10° CMFs to the 2° cone fundamentals, allowing for adjustments in lens, macular, and photopigment optical densities (as known to vary with field size, see Section 1.4.3 and 1.4.4). One thus never directly solves for M_{10° in the derivation of the 10° cone fundamentals, but M_{10° can be solved for once the 10° cone fundamentals have been obtained through this curve fitting process (and importantly, this means the final set of 10° cone fundamentals are simply a linear transform of the Stiles and Burch, 1959, 10° CMFs).

The standard observer rod spectral sensitivity function is the same as the scotopic luminous

efficiency function, $V'(\lambda)$, reported by Wyszecki and Stiles, 1982, derived from measurements by Wald, 1945, and Crawford, 1949 (Wyszecki & Stiles, 1982). Wald, 1945, measured the absolute threshold of vision for a 1° at 8° offset from the fovea (to isolate rods) across the wavelength spectrum, for 52 participants, to estimate the rod spectral sensitivity curve, while Crawford, 1949, estimated the spectral sensitivity from brightness matches made across two 20° fields.

The standard observer photopic luminous efficiency function, $V(\lambda)$, is based on measurements by Sharpe et al., 2005, which derives the luminous efficiency as the weighted combination of the standard observer L and M cone spectral sensitivities based on the average relative contribution of L and M cones in the retina (as it is believed only the L and M cones feed into the luminance pathway, see Section 1.5 for more detail). Again, luminous efficiency is measured at the cornea.

Unlike the cone spectral sensitivities, rod spectral sensitivity (and scotopic luminous efficiency), and photopic luminous efficiency function, the melanopsin spectral sensitivity used in the standard observer does not come from human psychophysical data directly. Rather, the melanopsin spectral sensitivity is derived from the action spectrum of the melanopsin photopigment adjusted to account for pre-receptoral filtering mechanisms (Lucas et al., 2014; CIE, 2018). Unless otherwise specified, these standard observer functions are used throughout this thesis.

1.4.2 Variability between individuals

While a useful model, the standard observer notion fails to account for the fact that there are substantial individual differences that exist in the photoreceptor spectral sensitivities between different observers. At the extreme end are observers who are missing one or more of the photoreceptor types. Individuals missing one of the three cone types are known as dichromats. It is also possible to have an anomalous version of the cone photopigment. Individuals who have such a pigment are known as anomalous trichromats. Yet even within individuals who do not have anomalous pigments or missing photoreceptor classes, there is a large degree of individual variation. Thus a complete model of human photoreceptor signals would include a description of these differences.

Additionally, studying the disconnect between individual differences at the photoreceptor level and behavioural level can provide insight into cortical mechanisms and adaptation. Studying individual differences is also of interest to understand the causes of these differences which might provide clues as to why the retinal mechanisms develop differently in different individuals, including genetic and environmental drivers of photoreceptor fate. This section of the introduction provides a brief overview of the known sources of differences in the cone spectral sensitivities and luminous efficiency functions

between individuals. Systematic differences in the cone photoreceptors between individuals is the topic of Chapter 2. There is less known about, and a lesser degree of suspected, differences in rod and ipRGC spectral sensitivities between individuals, so here the focus is on the differences in the cone signals between individuals.

Variability within individuals

Just as there exist individual differences in the retinal mosaic between individuals, so too do there exist individual differences across the retinal mosaic within a given individual. As already discussed, in Section 1.3, there are eccentricity differences in the packing arrangement and diameter of the cones and rods, and spatial patterning of the ipRGCs. Importantly, parafoveal cones have a longer and less optically dense outer segment than their foveal counterparts (Elsner et al., 1993; Burns & Elsner, 1985). Such eccentricity differences in optical density are important as optical density impacts the width of the spectral sensitivity functions, due to Beer's law (Smith & Pokorny, 2003). The peak optical density of the macular pigment, which acts as a pre-receptoral filtering mechanism altering the light spectrum before it even reaches the photoreceptors, also varies with eccentricity (Moreland & Alexander, 1997). This thus, impacts the shape of the spectral sensitivities for all five photoreceptor classes across the retina. Therefore the appropriate spectral sensitivity functions depend on the retinal eccentricity being stimulated, which must be accounted for (see Section 1.4.4 below) when doing wide field stimulation of the retina.

Additionally, retinal vasculature varies across the retina, with the foveola being free of vessels, and vessel diameter increasing with eccentricity (Snodderly et al., 1992; Bird & Weale, 1974). Due to the retinas "back-to-front" layout (see Fig. 1.2), there will be regions of the retina where the photoreceptors lie in the shadow of the retinal vessels, regions where the photoreceptors are not blocked by the vessels, and regions inbetween where the photoreceptors are partially shadowed by the vessels. The latter is known as the penumbra (Adams & Horton, 2003). Photoreceptors blocked by, or partially blocked by, the retinal vasculature will have different spectral sensitivity functions than those in the open field. While it is possible to visualize one's own retinal vasculature (Spitschan et al., 2015), known as the Purkinje tree (Purkinje, 1819; Purkinje, 1823), it is difficult to quantify. Subsequently there is no quantification of the changes in spectral sensitivity across the retina as a result of the Purkinje tree shadow. Thus, even if one accounts for variation in the spectral sensitivities as a function of optical density across eccentricity, there will be irregular errors in the spectral sensitivities due to an individual's Purkinje tree.

1.4.3 Quantifying variation: the CIE06 model

Before light even reaches the photoreceptors, it is filtered by the lens and macular pigment, whose optical density, and thus the effective filtering of the incident light, can vary between and within observers (Webster & Macleod, 1988; Asano et al., 2016; Emery & Webster, 2019). Such differences in pre-receptoral filtering between observers will result in differences in the spectral sensitivities of all five photoreceptors. For the cones, differences in the optical density of the photopigment and in the wavelength of peak sensitivity of the photopigment also result in differences in the width and shift, respectively, of the cone spectral sensitivities (Stockman & Sharpe, 2000; CIE, 2006; Elsner et al., 1993; Winderickx et al., 1993; Asenjo et al., 1994; Merbs & Nathans, 1992; Neitz & Neitz, 1998; Sharpe et al., 1998).

Taken together, this gives eight underlying physiological factors that can differ between, and within, individuals, namely: the optical density of the lens, $D_{lens}(\lambda)$, the optical density of the macular pigment, $D_{mac}(\lambda)$, the optical density of the L cones, D_L, the M cones, D_M, and the S cones, D_S, and the peak wavelength sensitivity of the L cones, λ_{max_L}, the M cones, λ_{max_M}, and the S cones, λ_{max_S}. Firstly, I will explain how each of these factors feeds into the calculation of the cone spectral sensitivities, and thus how each factor may impact the shape of the cone spectral sensitivity functions. If one knows the value of each of these eight parameters for a given individual, one can derive their individualized cone spectral sensitivity functions. I will then outline how one may be able to better approximate an individual's cone spectral sensitivities using two easily measured parameters, namely age and field size (in degrees of visual angle).

The calculations of the cone spectral sensitivity functions, in terms of quanta, $l_q(\lambda)$, $m_q(\lambda)$, $s_q(\lambda)$, from the eight aforementioned parameters is taken from the CIE 2006 (CIE06) model (CIE, 2006). (Note, since $E = h\nu$, and $\lambda = \frac{1}{\nu}$, to convert from the quantal sensitivities to the energy sensitivities, one simply must multiply by λ and re-normalise). The cone spectral sensitivities, that is the probability that the light incident on the cornea is absorbed by the cones, as a function of wavelength, are calculated as in Eq. 1.3:

$$\begin{aligned} l_q(\lambda) &= a_l(\lambda) \cdot 10^{-D_{mac}(\lambda) - D_{lens}(\lambda)}, \\ m_q(\lambda) &= a_m(\lambda) \cdot 10^{-D_{mac}(\lambda) - D_{lens}(\lambda)}, \\ s_q(\lambda) &= a_s(\lambda) \cdot 10^{-D_{mac}(\lambda) - D_{lens}(\lambda)}. \end{aligned} \tag{1.3}$$

In the above equations, $a_l(\lambda)$, $a_m(\lambda)$, and $a_s(\lambda)$, are the retinal absorptance spectra of the cones, that is the probability of the cone type absorbing the light incident on the retinal photoreceptor

layer at a particular wavelength. We have already seen that the parameters $D_{mac}(\lambda)$ and $D_{lens}(\lambda)$ are the optical densities of the macular pigment and lens at a particular retinal location and for a particular individual. Thus one can see that the spectral sensitivities of the cones are the product of the transmittance spectra of the pre-retinal filtering mechanisms and the retinal absorptance spectra.

Let's now further unpack the derivation of the parameters that appear in Eq. 1.3. The retinal absorptance spectra of the cones is given as in Eq. 1.4:

$$\begin{aligned} a_l(\lambda) &= 1 - 10^{-D_L A_L(\lambda)}, \\ a_m(\lambda) &= 1 - 10^{-D_M A_M(\lambda)}, \\ a_s(\lambda) &= 1 - 10^{-D_S A_S(\lambda)}. \end{aligned} \tag{1.4}$$

One will again notice the familiar parameters, D_L, D_M, D_S, as the optical densities of the cones. $A_L(\lambda)$, $A_M(\lambda)$, $A_S(\lambda)$, are the low-optical density spectral absorbances of the L, M, and S photopigments, as plotted in Fig. 1.3. Thus the retinal absorptance spectra of the cones are derived from the photopigment optical densities multiplied by the absorbance spectra of the photopigment, scaled by optical density.

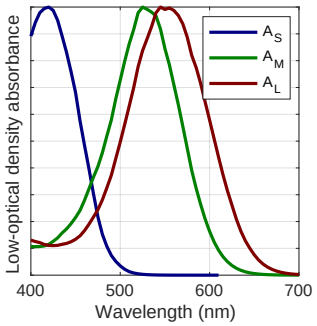


Figure 1.3: The spectra of the low-optical density absorbances of the L, M, and S photopigments, $A_L(\lambda)$, $A_M(\lambda)$, and $A_S(\lambda)$, in terms of energy.

One can see how the eight physiological parameters introduced as the drivers of variation between individuals can cause such differences when one considers the calculations of the cone spectral sensitivities as in Eq. 1.3 and Eq. 1.4. The optical densities of the lens, macular pigment, and the photopigments, simply input directly into these equations, while the peak wavelength sensitivity of each photopigment absorptance spectrum is simply a shift in the low-density absorbance spectrum as defined in the Asano et al., 2016 individual observer model (an extension of the CIE06 model), as in Eq. 1.5:

$$A_j(\lambda) = A_j(\lambda - s_j), \tag{1.5}$$

where $j \in [L, M, S]$ and s_j is the difference between $\lambda_{max\,j}$ for the individual and $\lambda_{max\,j}$ for the template (in this case, the standard observer model) (Asano et al., 2016). Thus, if one knows the standard observer low-density absorbance spectra, along with the eight physiological parameters that drive individual differences, $D_{lens}(\lambda)$, $D_{mac}(\lambda)$, D_L, D_M, D_S, λ_{max_L}, λ_{max_M}, and λ_{max_S}, then one can calculate the individualised cone spectral sensitivities for any given observer.

However, measuring these parameters directly is difficult. Fortunately, one can exploit known causes of individual differences in some of these parameters, namely the optical densities, such that one can calculate the individual optical densities from equations that take in easily measurable input parameters, namely age and field size, and the standard observer templates of these parameters. Let's unpack these equations one by one.

Firstly, the peak optical density of the cone photopigments, D_L, D_M, D_S, has a dependence on field size (Pokorny & Smith, 1976). As noted above, parafoveal cones are less optically dense than foveal cones, and thus a narrower spectral sensitivity function (Elsner et al., 1993; Burns & Elsner, 1985). In accordance with the CIE06 model (CIE, 2006), the optical density of the cone photopigments can be calculated as a function of field size, fs, as measured in degrees of visual angle, as in Eq. 1.6, and as plotted in Fig. 1.4:

$$\begin{aligned} D_L &= 0.38 + 0.54e^{-\frac{fs}{1.333}}, \\ D_M &= 0.38 + 0.54e^{-\frac{fs}{1.333}}, \\ D_S &= 0.30 + 0.45e^{-\frac{fs}{1.333}}. \end{aligned} \tag{1.6}$$

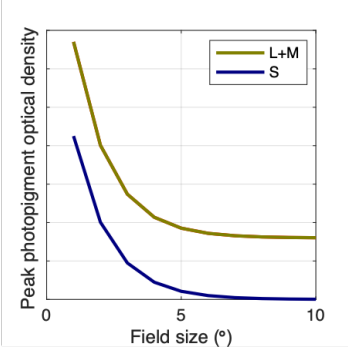


Figure 1.4: The peak optical density of the cone photopigment, D_L, D_M, and D_S, as a function of field size (in degrees of visual angle)

.

Secondly, the peak optical density of the macular pigment, D_{macmax} is also dependent on field

size, fs, in degrees of visual angle, as in Eq. 1.7 and in Fig. 1.5(a) (Moreland & Alexander, 1997):

$$D_{macmax} = 0.485e^{-\frac{fs}{6.132}}. \tag{1.7}$$

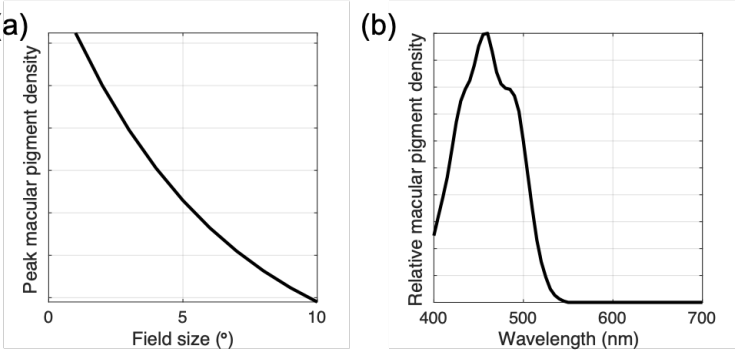


Figure 1.5: The spectrum of the optical density of the macular pigment. a) The peak density of the macular pigment, D_{macmax}, as a function of field size. b) The relative optical density of the macular pigment, $D_{macrel}(\lambda)$, which is assumed to be independent of field size, as a function of wavelength.

The optical density of the macular pigment is also wavelength dependent, $D_{macrel}(\lambda)$ as shown in Fig. 1.5(b). This relative spectral density of the macular pigment is assumed to be the same as a function of field size, such that the resultant macular pigment density, $D_{mac}(\lambda)$ can be defined as in Eq. 1.8:

$$D_{mac}(\lambda) = D_{macmax} \cdot D_{macrel}(\lambda). \tag{1.8}$$

Thirdly, and finally, the optical density of the lens is dependent on age. In fact, lens density is composed of two components, one of which is age dependent, $D_{lens1}(\lambda)$, as shown in Fig. 1.6(a), and one of which is age independent, $D_{lens2}(\lambda)$, as shown in Fig. 1.6(b). The optical density of the lens can be calculated for an individual of age, a, between the ages of 20 and 60 as in Eq. 1.9:

$$D_{lens}(\lambda) = D_{lens1}(\lambda)(1 + 0.02(a - 32)) + D_{lens2}(\lambda), \tag{1.9}$$

and for individuals over the age of 60 as in Eq. 1.10:

$$D_{lens}(\lambda) = D_{lens1}(\lambda)(1.56 + 0.0667(a - 60)) + D_{lens2}(\lambda). \tag{1.10}$$

The resulting $D_{lens}(\lambda)$ for a 20-, 40-, and 60-year-old is depicted in Fig. 1.6(c).

Thus, if one knows the individual's age and the field size at which they are viewing the stimuli, one can use the CIE06 model to improve the estimates from the means-based standard observer model

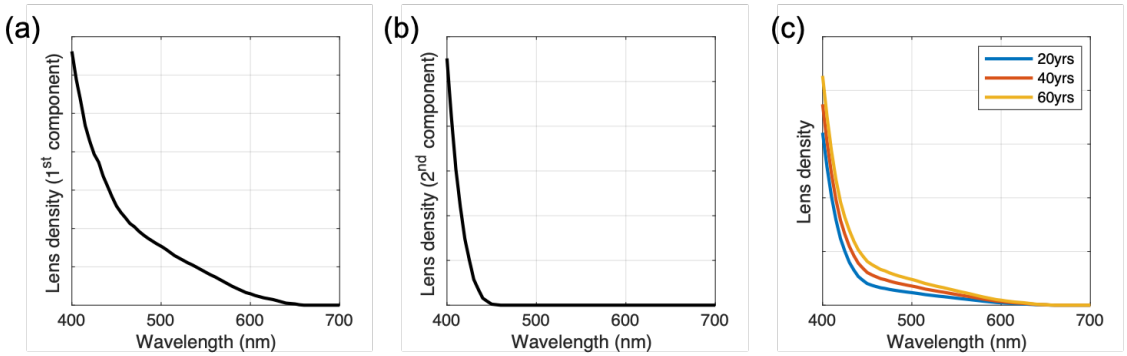


Figure 1.6: The spectrum of the optical density of the lens. a) The density of the first component of the lens, $D_{lens1}(\lambda)$, which varies as a function of age, as a function of wavelength. b) The density of the second component of the lens, $D_{lens2}(\lambda)$, which is age-independent, as a function of wavelength. c) The full lens density spectrum, $D_{lens}(\lambda)$, for a 20-year-old, 40-year-old, and 60-year-old individual.

towards personalised estimates of the cone spectral sensitivities for that individual (Stockman et al., 1999; Stockman & Sharpe, 2000; CIE, 2006).

1.4.4 Variability in rods and ipRGCs

There is less variation in the peak wavelength of sensitivity of the photopigment for rods and melanopsin, for example, there are no known polymorphisms of the rod photopigment genes (Sung et al., 1991). We also lack a quantitative descriptions of intra-observer variation in rods and and ipRGCs i.e. in photoreceptor packing density and optical pigment density across the retina. Thus, at present, there is no "bottom-up" model for deriving the rod and melanopsin spectral sensitivity functions as there is for the cones in the CIE06 model described above. This is not necessarily due to a lack of individual differences in rods and ipRGC signals, but a lack of a quantification of these differences.

In theory one could take a "top-down" approach and derive the rod photopigment absorptance spectra from the standard observer rod spectral sensitivity and then add in the quantified individual variation in the lens and macular pigment densities as parameterised by age and field size in the CIE06 model. For the melanopsin spectral sensitivity, as the spectral sensitivity is derived from the photopigment absorptance spectrum and passed through the pre-receptoral filtering mechanisms, the CIE S026 model does include an age correction that can be applied (CIE, 2018). Such a correction is not added for field size as there is not yet a standard quantification of how ipRGC density varies with field size. If age correction is applied, this results in the peak spectral sensitivity of rods and ipRGCs shifting to longer wavelengths as there is increased attenuation of short wavelengths due to increases in lens density (Spitschan, 2019b). However, for the rods, one must exercise caution as the $V'(\lambda)$ measurements of Wald, 1945, and Crawford, 1949, from which the rod spectral sensitivity function is

derived are only thought to represent observers under the age of 30 (Wyszecki & Stiles, 1982).

1.4.5 Individual differences in luminous efficiency

By virtue of the lack of account of individual differences in the rod spectral sensitivity, there is also no consideration of individual differences in scotopic luminous efficiency as the two are one and the same. However, there are large individual differences in photopic luminous efficiency. As the L and M cones sum into the luminance pathway (see Section 1.5.1), $V(\lambda)$ is determined by both the shape of the underlying L and M cone spectral sensitivities and by the relative weighting of the L and M cone spectral sensitivities in $V(\lambda)$. Therefore differences in $V(\lambda)$ are driven by both the variation in the underlying L and M cone spectral sensitivities and by differences in the weighting of L and M cone contributions to the luminance pathway. One major cause of the latter is believed to be due to differences in the relative number of L and M cones between individuals (Carroll et al., 2002; Hofer et al., 2005), though differences in postreceptoral weighting between individuals may also contribute.

1.4.6 Modelling individual differences

To summarise, the study of human photoreceptor signals must account for a consideration of individual differences. Standard observer models, while useful for some applications, do this only in part. Part of the reason these differences are not included in current models is due to gaps in our understanding of these differences, particularly in systematic drivers of these differences across individuals. These ideas are further expanded on in Chapter 2.

1.5 Photoreceptor inputs to vision

The main neural function driven by the photoreceptor signals is vision. Cone and rod signals are the primary drivers of vision, but ipRGC signals have been shown to potentially influence visual perception (as described in detail below) (Zaidi et al., 2007; Zele, Feigl et al., 2018; Spitschan et al., 2017; Cao et al., 2018; Brown et al., 2012; Zele, Adhikari et al., 2018). Vision can be subdivided into three major types of visual processing: scotopic, mesopic, and photopic vision (Hood & Finkelstein, 1986; Stockman & Sharpe, 2006). Scotopic light levels are well below the cone threshold, and thus scotopic vision is driven solely by rod signals. Photopic vision, on the other hand, occurs above rod saturation levels and thus is driven by just the cones. In the mesopic range, between the scotopic and photopic range, both rod and cone signals are available and feed into visual perception (Stockman & Sharpe, 2006; Zele

& Cao, 2015). There is not a clear distinction in the illumination levels that correspond to the three different types of visual processing, though broadly speaking the scotopic threshold i.e. the absolute threshold for seeing, is around -6 log cd/m^2, while cones begin to signal at around -3 log cd/m^2, and rod saturation begins around 1 log cd/m^2(Stockman & Sharpe, 2006). In this section I discuss the photoreceptor inputs to vision at scotopic, mesopic, and photopic light levels. I describe first the simple and well-established versions of these pathways, where only rods and/or cone signals drive these processes. I then include a discussion at the end of this section on how ipRGC signals may also contribute to vision.

1.5.1 Photopic vision

Photopic vision occurs in bright ambient light conditions and is associated with good colour vision and high spatial acuity. Photopic vision is driven by the cones. The cone signals are transmitted through the retina via various cone bipolar cells to specific RGCs which then project to the lateral geniculate nucleus (LGN) and from there to the cortex. Chromatic opponency, i.e. comparison of the cone signals, begins at the level of the RGCs, and there are three major types of RGC which have historically been implicated in colour vision. Traditionally, it has been understood that there are three distinct retinogeniculate pathways for the cone signals: the luminance pathway, the L-M-opponent pathway, and the S-opponent pathway (as reviewed by e.g. Casagrande and Xu, 2004; Solomon, 2021; Martinovic, 2014). However, there are many more retinal ganglion cell types which have since been identified (each with their own post-receptoral pathways) which suggest there is scope for our understanding of the role of each of these pathways to be updated (Kim et al., 2021).

Luminance pathway

The luminance pathway is achromatic and has a high temporal resolution (Dacey, 1996; Schiller, 2010). The magnocellular pathway, via the large bodied magnocellular cells in the bottom two layers of the LGN, is the main route for luminance information. The magnocellular layers are innervated by parasol RGCs, who receive a summed input from L and M cones transmitted via diffuse bipolar cells (with evidence suggesting that S cones do not contribute strongly, if at all, to the luminance pathway (Lee et al., 1988, 1989; Dacey & Lee, 1994; Sun et al., 2006a, 2006b)). Parasol cells compose around 10% of all RGCs (Dacey, 1994). Parasol cells can be ON-centre, such that their firing rate increases when the centre of their receptive field is illuminated, or OFF-centre, where their firing rate increases when the centre of their receptive field is not illuminated (Dacey & Lee, 1994). The magnocellular layers

primarily innervate the primary visual cortex (V1), specifically layer 4C(α) (Casagrande & Xu, 2004; Martinovic, 2014).

L-M-opponent pathway

The L-M-opponent information is transmitted via the parvocellular pathway (Derrington et al., 1984; Wiesel & Hubel, 1966). The parvocellular layers of the LGN are the top four layers, which have smaller cell bodies than magnocellular cells (Kim et al., 2021), and receive input from midget ganglion cells, which account for 60-80% of RGCs (Yamada et al., 1996; Dacey, 1993). Midget ganglion cells have an ON-centre, OFF-surround receptive field, receiving opponent signals from L and M cones via midget bipolar cells. In the fovea midget ganglion cell centres' make a one-to-one synapse with a midget bipolar cell, which make a one-to-one synapse with a single cone (Kling et al., 2019). It is unclear whether the surround of midget ganglion cells is of just the opposing cone type to the centre of the receptive field, or a mixture of L and M inputs, though the latter seems most probable (Lennie et al., 1991; Wool et al., 2018). The parvocellular layers projects to layers 4Cβ and 4A of V1 (Casagrande & Xu, 2004; Martinovic, 2014). The L-M-opponent pathway has a higher spatial resolution but lower temporal resolution than the luminance pathway (Schiller, 2010).

S-opponent pathway

Traditionally, it has been understood that S-opponent information is transmitted via S-ON bipolar cells, which synapse onto small bistratified RGCs (Dacey & Lee, 1994), which also receive opponent L and M inputs from diffuse bipolar cells (Szmajda et al., 2008; Percival et al., 2009). Small bistratified RGCs do not have a centre-surround receptive field organisation, and account for approximately only 4% of all RGCs (Kim et al., 2021; Dacey & Lee, 1994). Small bistratified RGCs project to the middle pair of koniocellular layers in the LGN (Kim et al., 2021; Dacey, 1994). The koniocellular cells are much smaller than the magnocellular and parvocellular cells and the koniocellular layers are much thinner and intermittent to the other layers in the LGN (Casagrande et al., 2007). The middle pair of koniocellular layers, which transmit the S-opponent information, project to V1 (Hendry & Reid, 2000). However, the story of S-opponent processing has recently become more complex due to evidence suggesting that S-OFF opponent information is also transmitted via S-OFF inputs to midget ganglion cells (Dacey et al., 2017; Wool et al., 2019).

DKL Colour Space

The three pathways described above are thought to be independent and led to the idea of three orthogonal mechanisms for photopic vision: L+M (luminance), S-(L+M) (S-opponent), and L-M (L-M-opponent). These axes form the orthogonal axes of the DKL colour space, which was developed in a famous study by Derrington et al., 1984, who derived these three *cardinal mechanisms* from cell recordings in the macaque retina. Though the idea of orthogonal axes of photopic vision had been shown psychophysically two years prior, in a study by Krauskopf et al., 1982, who first coined the term "cardinal axes". Krauskopf et al., 1982, showed that adaptation to one of the two colour directions or along the luminance axis selectively reduces discrimination thresholds along that direction, but not the other two, while adaptation to intermediate axes reduces thresholds in all directions. Reanalysis of this data, along with new evidence, by Krauskopf et al., 1986, however, suggested the presence of "higher-order" colour mechanisms, tuned to axes intermediate to the cardinal axes (Eskew Jr, 2009).

1.5.2 Scotopic vision

Scotopic vision is achromatic with poor spatial acuity, but allows us to see in low ambient light levels. Scotopic vision is the range below the cone threshold, so is where only rods signals are available. Scotopic vision is therefore perhaps the simplest case of vision: only one photoreceptor is signalling and the principle of univariance leads to the achromatic nature of scotopic vision, and to additivity (Stockman & Sharpe, 2006). At scotopic light levels, the human rod system is slow and sluggish, which corresponds to Stiles' π_0 mechanism (Stiles & Burch, 1959; Sharpe et al., 1989). The peak scotopic temporal contrast sensitivity is 5-9 Hz, which is lower than the photopic peak temporal contrast sensitivity of 8-10Hz, as is the maximum critical flicker fusion frequency (20-28Hz scotopic vs 50-60Hz photopic) (Zele & Cao, 2015; Delange, 1952; Conner & MacLeod, 1977; Sun et al., 2001a).

It is believed that at scotopic light levels the rods signals are transmitted through the rod bipolar pathway, sometimes called the primary pathway (Daw et al., 1990; Sharpe & Stockman, 1999). The hyperpolarizing rod signals lead to the depolarization of rod bipolar cells, which subsequently depolarise AII amacrine cells. The AII amacrine cells then innervate OFF-center RGCs while ON-center RGCs are innervated by ON/OFF cone bipolar cells which receive input from AII amacrine cells via gap junctions (Zele & Cao, 2015; Sharpe & Stockman, 1999; Verweij et al., 1999). It has been shown that these scotopic rod signals can then be transmitted to the visual cortex by the three retinogeniculate pathways described above, though recordings from macaque retina suggest the magnocellular pathway is the main route for rod signals (Lee et al., 1997; Crook et al., 2009; Field et al., 2009; Cao et al., 2010;

Sun et al., 2001b). A crucial feature of rod pathways is the large convergence that occurs as one moves through the pathway - with many rods converging onto one rod bipolar cell, many of which then converge on one AII amacrine cell, and many of which then converge onto RGCs (Sterling et al., 1987; Sterling et al., 1988). This pooling of the rod signal causes amplification of the rod signal enabling visual perception to occur in very low light levels (Hecht et al., 1942).

Despite the theoretical prediction that scotopic vision should be achromatic, there have been reports of colour perception persisting in what are traditionally thought of as scotopic light levels (Nagel, 1924; Pokorny et al., 2006, 2008; Elliott & Cao, 2013). It is difficult however to isolate rod function and attribute colour percepts to rod function alone, as the cones may in fact be active. It is possible that in truly scotopic vision, that colour percepts arise from the shared cortical mechanisms of the scotopic and photopic visual processes, and are relational colour appearances driven by attributing colour differences to the real intensity differences seen (Nagel, 1924).

1.5.3 Mesopic vision

Mesopic vision is where models begin to get complex, and despite many years of work, there is no consensus on how rods and cones contribute to mesopic vision. One of the major findings relating to mesopic processing is the duality in rod mediated vision, the first evidence for which came from temporal resolution measurements of a rod monochromat who lacked cone mediated vision entirely (Sharpe & Stockman, 1999; Hecht et al., 1948). The rod monochromat exhibited two distinct regions of maximum flicker fusion frequency depending on intensity. The first, more sensitive process, saturates at around 15 Hz, while the second, less sensitive, persists until around 28 Hz. These results suggest that mesopic rod-mediated vision operates through a distinct pathway to scotopic rod-mediated vision. Additional psychophysical evidence from normal trichromatic observers shows the existence of a rod-flicker null at 15 Hz, which provides further evidence for the existence of two separate pathways, and suggests that these pathways act in opposition to one another, and destructively interfere at 15 Hz resulting in the interim null (Sharpe et al., 1989; Sharpe & Stockman, 1999).

The two distinct temporal resolution limits of the rod system are typically explained by the existence of two distinct pathways through which rod signals are transmitted. The first, slow and sensitive pathway, is the primary pathway as described in the section above on scotopic vision. The faster, less sensitive signal is believed to be transmitted through the rod-cone gap junction pathway, sometimes called the secondary pathway, whereby rod signals "piggy-back" on the cone signals (Zele & Cao, 2015; Sharpe & Stockman, 1999; Verweij et al., 1999). It has also been suggested that a third pathway, the

tertiary pathway, where rod signals innervate OFF-cone bipolar cells directly, may also contribute to the transmission of the fast rod signal (Soucy et al., 1998; Hack et al., 1999).

While these three distinct routes for rod signals are clear, as are the pathways for cone signals, it is still not known how cone and rod signals are combined through these different pathways during mesopic vision. Human psychophysics in rod or cone isolating conditions has, however, uncovered some clues to this combination. For example, Sun et al., 2001b, showed that rod and cone signals linearly sum in the magnocellular pathway, but probability sum in the parvocellular pathway. The combination of rod and S cone signals is more complex, as psychophysical studies suggest that the interactions depend on contrast (Zele et al., 2012).

Electrophysiological recordings from macaque retina *in vitro* have also provided insights into how the rod and cone signals might combine in post-receptoral pathways. Intriguingly, recent evidence from macaque electrophysiological recordings of parasol RGCs, by Grimes et al., 2018, suggests that the secondary and tertiary pathways convey at most a small rod-signal, with the primary pathway being the dominant mediator of the rod signal, even under mesopic conditions. The primary pathway was even shown to speed up as light abundancy increases, suggesting it is not a change of pathway but a change in kinetics of the primary pathway due to the level of rod adaptation that accounts for the slow and fast rod signal (Grimes et al., 2018). This work also did not find an existence of a 15Hz flicker null for the rod mechanisms, which was also not observed in *in vivo* recordings of monkey midget RGCs (Cao et al., 2010).

To repeat what I wrote at the start of this section: mesopic vision is complex, and there is still no clear consensus on how rod and cone signals combine to contribute to mesopic vision. While there are clear candidate pathways for rod and cone signals in mesopic vision, it is still unclear how these signals combine through these various pathways, and thus what contribution the rod and cone signals make towards perception in the mesopic range.

1.5.4 ipRGC inputs to vision

It has been difficult to pin down ipRGC inputs to vision, primarily because of the difficulty in assessing ipRGC function in the human retina due to methodological limitations. Most of our understanding of ipRGC function has been gained through projection mapping in mice, via injection tracing and transgenic mouse models (Mure, 2021). Such techniques are not possible in humans, and thus morphological assessment of ipRGCs *in vitro* in the human retina has relied solely on immunostaining melanopsin: a technique known to miss ipRGCs that are detected via other methods in mice (Mure,

2021). It is therefore likely that ipRGC abundance is underestimated in humans, and that subtype characterisations are lacking. There is, however, converging (and recently, fast growing) evidence from mouse and human *in vitro* studies that suggest ipRGCs innervate visual pathways, and thus may contribute to vision. In the mouse retina it has been shown that the M4 subtype of ipRGCs project heavily to the LGN, providing a possible pathway for ipRGC inputs to vision (Estevez et al., 2012). In the macaque and human retina M1 and M2 ipRGC subtypes appear to project to the dorsal LGN (Liao et al., 2016).

Behavioural evidence from mouse and humans also suggests a role of ipRGCs in vision. In mice, ipRGCs have been shown to support pattern vision when mice are lacking functional cones and rods (Ecker et al., 2010). ipRGCs have also been shown to play a role in contrast sensitivity in mice (Schmidt et al., 2014). In humans, psychophysical evidence demonstrates that ipRGCs support brightness discrimination, a finding that converges with behavioural and electrophysiological results in mice (Brown et al., 2012). Additionally, a functional MRI response has been observed in participants viewing a melanopsin isolating stimulus, with the strength of the response scaling with the melanopsin contrast of the stimulus (Spitschan et al., 2017).

While the evidence does seem to be converging in support of a role of ipRGCs in vision, what is still poorly characterised is the interaction between rod and cone signals and the melanopsin-mediated intrinsic ipRGC signal. Further, it appears such interactions depend on ipRGC subtypes. In humans, M1 ipRGCs receive an S-OFF input (Patterson, Kuchenbecker et al., 2020), while M2 ipRGCs receive an S-ON input (Patterson, Mazzaferri et al., 2020). There is still much to quantify about these interactions.

One must also consider how the recent evidence, as summarised above, of ipRGC contributions to vision fits with the original colour matching functions from which the cone spectral sensitivity estimates were derived (Stiles & Burch, 1959). As pointed out by Danilova and Mollon, 2021, if melanopsin contributed to colour perception, then why is it possible for an observer to make a trichromatic (foveal) colour match? Why do they not need four primaries to equate the cone and melanopsin signals across the bipartite field? The fact that observers only need three primaries, not four, to make such foveal colour matches suggests that there are only three independent mechanisms underlying colour perception. If melanopsin contributed independently to the cones, four primaries would be needed to make a match. One way, perhaps, to solve this apparent conflict might be if melanopsin only feeds into colour perception through the same post-receptoral channels through which the cone signals mediate colour perception. Thus the four mechanisms at the receptoral level would be reduced to just three at the post-receptoral stage, and thus the downstream visual system would still only have

access to three independent signals. If this were the case, this would, thus, however, mean that the cone spectral sensitivity functions are contaminated by melanopsin contributions, which would thus have implications for the design of melanopsin isolating stimuli (see Section 1.7).

1.5.5 Vision beyond photoreceptor signals

While the photoreceptor signals impose a strict limit on vision, our *perception* of an external scene is influenced by other higher-level cortical processes. There are for example, many contextual dependencies on colour appearance. For example, the two central squares in Fig. 1.7 appear to be different colours, when in fact they have the same RGB values. This is an example of simultaneous contrast, which shows how influential surrounding context can be on our colour perception (Jameson & Hurvich, 1959).

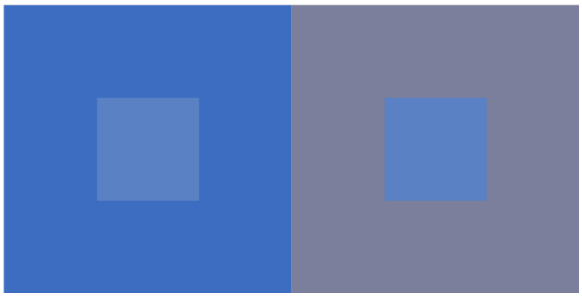


Figure 1.7: The two central squares appear to be two different colours, despite the fact that they in fact share the same RGB values. This is an example of the simultaneous contrast effect which demonstrates the impact of context on colour perception.

Contextual effects nicely demonstrate that colour appearance is not purely determined by photoreceptor signals. Another example of where our colour percept is not directly linked to the photoreceptor signals is the case of colour constancy (Smithson, 2005; Foster, 2011). Colour constancy describes the phenomenon where surface colours appear subjectively unchanging despite changes in the illuminant, and thus objective changes in the radiance spectrum reaching our eyes. An example of colour constancy is depicted in Fig. 1.8. Colour constancy represents a remarkable computational challenge that the visual system must face. From the photoreceptor signals alone there is no way for the visual system to reconstruct the full spectrum of the incident light, nor to disentangle which aspects of that spectrum are contributed by the illuminant and which are contributed by the surface. Yet, our visual system is able to achieve colour constancy (though not perfectly). It is not yet understood how the visual system is able to achieve a stable perception of object surface colour and correspondingly of illuminant colour from a simple pattern of photoreceptor signals. The computational challenges involved in colour constancy and illuminant perception, and how these problems might be bounded by the constraints

the external physical world imposes on our visual system is the topic of Chapter 3.



Figure 1.8: A photograph of Pembroke College Chapel Quad unedited (left) and under an orange filter (right). In both images the sky appears blue and the grass green, despite the fact that the sky and grass in each image have remarkably different RGB values, and when displayed out of context are clearly distinct colours. This demonstrates the property of colour constancy: the unchanging colour appearance of objects in the scene despite changes in illumination and thus objective chromaticity.

1.6 Photoreceptor inputs to non-visual light mediated neural mechanisms

ipRGCs are believed to primarily drive the non-visual light mediated mechanisms. By non-visual light mediated neural mechanisms I am referring to any process that is affected by external light stimulation, but which does not directly contribute to visual perception. In this section I will briefly review the major non-visual light mediated mechanisms and the ipRGC control of those functions, which have been reviewed in detail elsewhere (e.g. Spitschan, 2019b).

1.6.1 Circadian rhythm

The superchiasmatic nucleus (SCN) is often called the "master clock" (Moore, 2007). The SCN is situated in the hypothalamus and acts as an oscillator, to control the regulation of different body functions over a 24-hour period i.e. our circadian rhythm. The SCN innervates the pineal gland, which controls body temperature, and hormone production, including melatonin ("the sleep hormone") (Cajochen et al., 2003). The SCN also innervates the raphe nucleus, where serotonin is produced, and receives reciprocal inputs from the raphe-hypothalamic tract and pineal gland (Moore, 2007; Blume et al., 2019). However, the major input to the SCN is from ipRGCs via the retinohypothalamic tract. Thus the light in the external world acts as the timekeeper of the SCN and our circadian cycle.

Interestingly, in mice it has been shown that the degree of rod and cone coupling during mesopic vision depends on the time of day, suggesting a potential reciprocal output of the SCN back to the retina itself (Ribelayga et al., 2008).

In humans, M1 ipRGCs project to the (SCN) (Kim et al., 2021). Human chronobiology work has shown that the spectral sensitivity of circadian phase shifting broadly matches the spectral sensitivity of melanopsin, suggesting melanopsin signals are the main contributor of circadian shifts (Wright & Lack, 2001; Wright et al., 2004). There is, however some evidence that cone signals may also contribute to phase shifting in certain conditions (Gooley et al., 2010).

1.6.2 Melatonin suppression

As described above, melatonin is generated by the pineal gland, which is innervated by the SCN. Melatonin production is suppressed by light inputs through the retinohypothalamic pathway (Lewy et al., 1980). It is important to note that melatonin suppression and circadian rhythm are decoupled (Spitschan, 2019b). As with circadian phase shifts, the spectral sensitivity for melatonin suppression broadly matches the spectral sensitivity of melanopsin, suggesting intrinsic activity in the ipRGCs is the key driver of melatonin suppression (Nowozin et al., 2017; Prayag et al., 2019). There is however, recent evidence to implicate rods in melatonin suppression too (Vartanian et al., 2015).

1.6.3 Pupil size

M2 ipRGCs do not project to the SCN in humans, but rather to the olivary pretectal nucleus (OPN) in the thalamus (Kim et al., 2021). The OPN controls the pupillary light reflex (Klooster et al., 1995). It is clear that all photoreceptor signals can drive pupil size changes, but the nature of this control depends on ambient light level (Spitschan, 2019a). For example, at photopic light levels, L+M, melanopsin, and (weak) rod signals confer an excitatory input on pupil size, in conjunction with inhibitory signals from S cones (Spitschan et al., 2019).

1.7 Methods of studying photoreceptor signals

At the time of writing this thesis, it is not possible to directly study the five photoreceptor signals *in vivo* in a human observer. Therefore our understanding of the human photoreceptor signals results from the cumulation of many different indirect techniques to infer the photoreceptor signalling in the human retina.

Broadly speaking, these techniques can be divided into three categories: *in vitro* electrophysiological recordings at the single-cell, multi-electrode array, or gross scale of electroretinogram (ERG), from human, macaque, cat, and mouse retina; *in vivo* circuit mapping and manipulation in mice; and human psychophysics. The synopsis of visual and non-visual processes described up to this point has relied on evidence from all of these disciplines. This thesis, however, focuses solely on the use of human psychophysics to study photoreceptor signals. Human psychophysics offers distinct advantages over *in vitro* recordings and *in vivo* animal models, in that it offers an assessment of function in the fully intact, functioning system, in the organism that we are trying to understand. Mouse models, for example, are limited in that the retina of mice is known to be very different to that of humans, with mice expressing only two cone photoreceptor types, a relatively lower proportion of cones compared to rods, and no fovea, to name a few of these differences (Huberman & Niell, 2011).

The difficulty with human psychophysics, however, is isolating photoreceptor function. For example, traditionally, functionality was often attributed to the cones at light levels believed to be above the rod saturation level, but recent evidence suggests that the rods may saturate at higher illuminances than previously thought, and thus the contribution of rod signals to many visual functions may be underestimated (Lall et al., 2010; Frederiksen et al., 2021). As the spectral sensitivities of all five photoreceptor classes overlap, it is impossible to find a single wavelength on the visible spectrum that will activate only one of the photoreceptor classes. It would be desirable to be able to find such a wavelength, as this would be a simple solution to the problem of independent isolation of the photoreceptors.

If we instead consider a stimulus of not just one wavelength, but many wavelengths producing a spectrum of light, we can find pairs of spectra that produce the same response in four of the five photoreceptor classes, but a different response in one of the five photoreceptor classes, then we can exchange the two spectra to achieve isolation of an individual photoreceptor class. This method is known as silent substitution (Estévez & Spekreijse, 1982; Spitschan & Woelders, 2018). In the case where we keep the activations of the three cone classes constant, the two spectra are metamers. Without additional constraints, metamers will typically differ in their rod and melanopsin signals (as depicted in Fig. 1.9).

Independent control of the activations of the five photoreceptors requires control of five independent primary lights. Given control of the intensities of five independent primary lights, $p_j(\lambda)$, one can always find a solution for the relative intensities of those primaries, $\vec{C} = [r, g, b, c, m]$ to produce a specific relative set of desired photoreceptor activations, $\vec{A} = [A_L, A_M, A_S, A_R, A_I]$, given the spectral

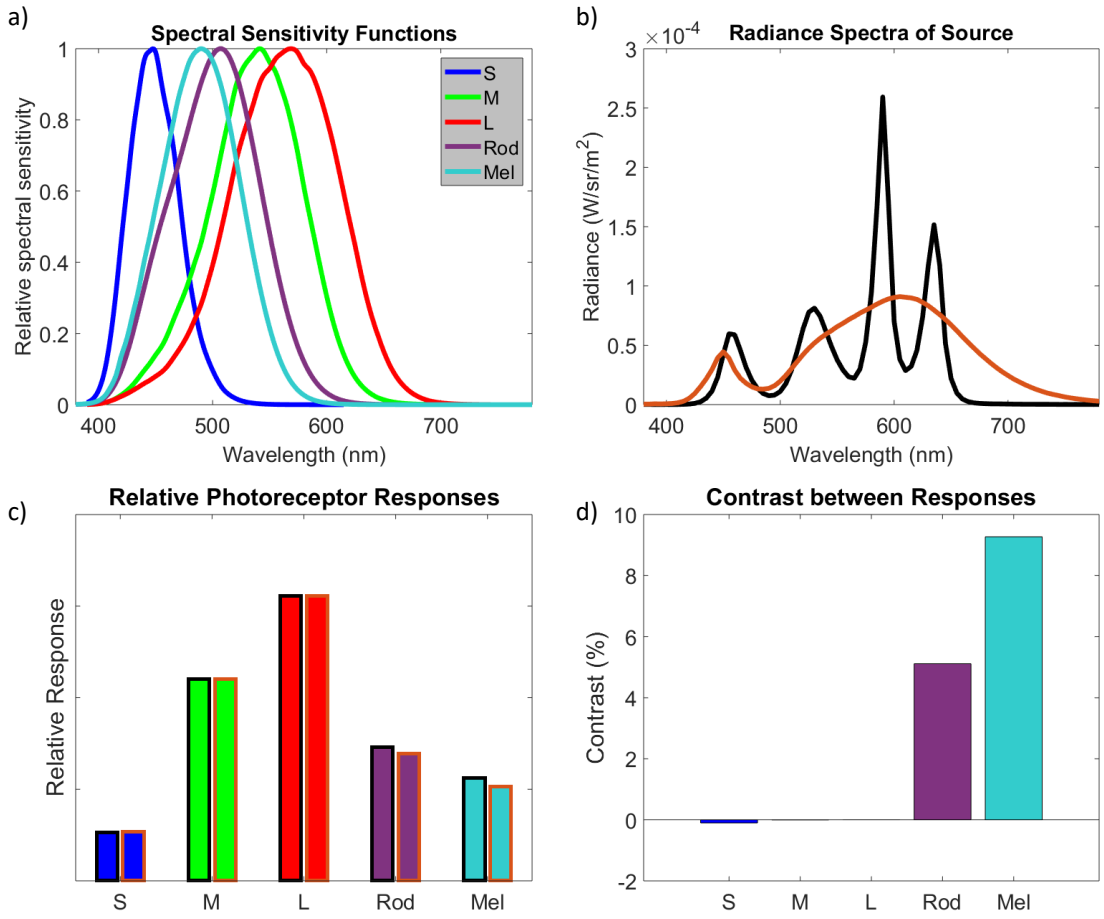


Figure 1.9: The principle of silent substitution, for the special case of metameric pairs. a) Energy-based spectral sensitivity functions of the photoreceptors. b) A pair of spectra, that are clearly different, yet yield metameric cone responses, with differing rod and melanopsin responses. c) Relative responses of the photoreceptors to the spectra shown in (b). d) Percentage difference in response for each photoreceptor type to the spectra shown in (b). The relative activity in the cones is equated for the two spectra while the relative activity in rods and in melanopsin is different, by 5.1% and 9.3%, respectively. The two spectra in (b) are metamers for cones.

sensitivities of the five photoreceptors, $T_k(\lambda)$, as in Eq. 1.11:

$$\vec{C} = B^{-1}\vec{A}$$

where

$$\vec{C} = \begin{bmatrix} r \\ g \\ b \\ c \\ m \end{bmatrix} \text{ and } \vec{A} = \begin{bmatrix} A_L \\ A_M \\ A_S \\ A_R \\ A_I \end{bmatrix} \tag{1.11}$$

and

$$B_{k,j} = \int_{\lambda} p_j(\lambda) \cdot \ T_k(\lambda)\, d\lambda$$

$$j \in [r, g, b, c, m]$$

$$k \in [L, M, S, R, I].$$

By definition therefore, one cannot independently isolate the photoreceptor signals using a traditional three-primary display. The best one can do with a three-primary display is to make assumptions about which photoreceptor mechanisms are likely to dominate (Zele & Cao, 2015). That is not to say that three-primary displays are useless for uncovering aspects of visual perception, and many important results about higher-order visual processing have come from experiments using three primary displays (of which there are far too many to even attempt to cite). What is important however, is that an experimenter is aware of the assumptions they make when using a particular device to probe vision, and the limits these place on the types of conclusions they are able to reach. This is particularly important in the case of online psychophysics, to which many researchers were forced to turn during the course of my DPhil research due to the impact of the COVID-19 pandemic (March 2020 - ongoing) on in-person psychophysics. Chapter 3 of this thesis focuses on the issue of display calibration for online visual psychophysics, specifically on the limits of calibration that can be achieved and thus the subsequent limits this puts on the type of experiments to probe vision that one can confidently run remotely.

To further how human psychophysics can quantify photoreceptor signals, one needs to expand the limits of the technologies that can deliver photoreceptor isolating stimuli. This demands display design and development. Chapter 4 of this thesis describes the development of such a novel display technology, which is unique in the control of photoreceptor function that it enables human visual

psychophysics to achieve. It is important to note that our photoreceptor isolating stimuli will only ever be as good as our estimates of the photoreceptor spectral sensitivities i.e. if there is some contamination of the melanopsin spectral sensitivity in the cone spectral sensitivity functions, then such designed melanopsin isolating stimuli may not, in practice, be melanopsin isolating. Perhaps one approach to improve confidence in the spectral sensitivity functions for such experiments may be to take a "bottom-up" approach and derive the spectral sensitivities from the photopigment absorptance spectra and the pre-receptoral filtering mechanisms. This idea is not further considered in this thesis, but is noted here for completeness.

1.8 Colorimetry

Colorimetry is the quantification of human colour perception. At the heart of colorimetry lies the colour-matching functions: the intensities of three independent primary lights required to make a match to a series of monochromatic test fields (Brainard & Stockman, 2010). Colour matching was the method by which the trichromatic nature of human colour vision was first demonstrated by Young, 1802. As normal trichromatic observers have three cone photoreceptor types, the colour-matching functions and the cone spectral sensitivities are simply a linear transform of each other (assuming no influence of rods and ipRGCs on colour matching). Thus colorimetry is a quantification of the cone photoreceptor signals. This section introduces some useful tools from colorimetry, including colour spaces, which are used throughout this thesis.

1.8.1 CIE 1931 XYZ

Perhaps one of the most ubiquitous tools in colorimetry, the CIE 1931 XYZ colour matching functions (CMFs) were the first attempt to quantify human colour vision. Colour matching data from Guild, 1931, and Wright, 1929, Wright, 1930, were used to derive the CIE 1931 XYZ CMFs (CIE, 1932). The CIE 1931 XYZ CMFs are the first notion of a standard observer: where a single set of functions, derived from the average of, in this case, 18 observers, can be used to describe the "average" observer. The cone spectral sensitivities, $\overline{l}(\lambda)$, $\overline{m}(\lambda)$, $\overline{s}(\lambda)$, can be directly derived from the set of CMFs, $\overline{r}(\lambda)$, $\overline{g}(\lambda)$, $\overline{b}(\lambda)$, via a linear, matrix transformation (Brainard & Stockman, 2010), as in Eq. 1.12:

$$M\begin{bmatrix}\overline{r}(\lambda)\\ \overline{g}(\lambda)\\ \overline{b}(\lambda)\end{bmatrix}=\begin{bmatrix}l(\lambda)\\ m(\lambda)\\ s(\lambda)\end{bmatrix},$$

where (1.12)

$$M=\begin{bmatrix}\overline{l_R} & \overline{l_G} & \overline{l_B}\\ \overline{m_R} & \overline{m_G} & \overline{m_B}\\ \overline{s_R} & \overline{s_G} & \overline{s_B}\end{bmatrix},$$

and $\overline{a_R}$, $\overline{a_G}$, $\overline{a_B}$, are the A-cone sensitivities to the **R**, **G**, **B** primary lights from which the colour matches were made, for $A \in [L, M, S]$. As the real **R**, **G**, **B** primary lights lead to $\overline{r}(\lambda)$ going negative, i.e. the **R** primary needed to be added to the target stimulus, in order for a colour match to be made, the CIE proposed a set of imaginary primaries, **X**, **Y**, **Z**, with associated CMFs, $\overline{x}(\lambda)$, $\overline{y}(\lambda)$, $\overline{z}(\lambda)$, which are always positive, and which can be derived from the real $\overline{r}(\lambda)$, $\overline{g}(\lambda)$, $\overline{b}(\lambda)$, CMFs by a simple linear transformation. It is these $\overline{x}(\lambda)$, $\overline{y}(\lambda)$, $\overline{z}(\lambda)$, CMFs for the imaginary **X**, **Y**, **Z**, primaries that are most commonly used in colorimetry (Brainard & Stockman, 2010).

From the CIE 1931 XYZ CMFs, $\overline{x}(\lambda)$, $\overline{y}(\lambda)$, $\overline{z}(\lambda)$, one can calculate the CIE 1931 XYZ tristimulus values, X, Y, Z, of a spectrum in the same way one computes the photoreceptor signals of a radiance spectrum, $P(\lambda)$, i.e. as the integral of the product of the spectral radiance and the CMFs, as in Eq. 1.13:

$$X=\int_\lambda P(\lambda)\overline{x}(\lambda)d(\lambda)$$

$$Y=\int_\lambda P(\lambda)\overline{y}(\lambda)d(\lambda) \quad (1.13)$$

and

$$Z=\int_\lambda P(\lambda)\overline{z}(\lambda)d(\lambda).$$

However, it is now well-established that the CIE 1931 XYZ CMFs are not the best estimates of CMFs, yet the CIE 1931 XYZ CMFs persist in colorimetry (Stockman, 2019). It is widely accepted that the most physiologically relevant CMFs are those back-derived from the linear transformations of the cone spectral sensitivities of Stockman and Sharpe, 2000 (Stiles & Burch, 1959). While the Stockman and Sharpe, 2000, cone spectral sensitivities have been widely adopted throughout the colour vision research community and industry as the cone fundamentals, the CMFs back-derived from these cone fundamentals have not yet been widely adopted by the community.

1.8.2 CIE 1931 xy chromaticity diagram

Another useful tool in colorimetry is the chromaticity diagram. Chromaticity coordinates are defined independently of intensity i.e. such that black and white at the intensity extremes of the achromatic axis would have the same chromaticity coordinates. By removing intensity information, chromaticity coordinates need only capture a two-dimensional representation of the human colour percept, which is thus easier to depict diagrammatically. The CIE 1931 xy chromaticity diagram is commonly used in industrial applications.

The CIE 1931 XYZ CMFs translate to the CIE 1931 xy chromaticity coordinates as in Eq. 1.14:

$$\begin{aligned} x &= \frac{X}{X+Y+Z} \\ y &= \frac{Y}{X+Y+Z} \end{aligned} \tag{1.14}$$

and

$$z = \frac{Z}{X+Y+Z} = 1 - x - y.$$

By design the Y tristimulus value of the CIE 1931 XYZ tristimulus values is the photopic luminance, and thus the most common representation of the CIE 1931 XYZ tristimulus values is the CIE 1931 xyY colour space. The luminance independent projection of the CIE 1931 xyY colour space is simply the CIE 1931 xy chromaticity diagram. Note, if the physiologically-relevant CIE CMFs become more widely adopted than the CIE 1931 XYZ CMFs then one can simply use the new XYZ CMFs to calculate the XYZ tristimulus values from which the xyY values are calculated. At the time of writing this thesis, luminance meters for colour science universally report CIE 1931 Y (i.e. the 1931 photopic luminance in cd/m^2), and colorimeters report the CIE 1931 xy coordinates. Since the CIE 1931 Y function is different than the physiologically-relevant CIE Y function, the measured luminance is dependent on the choice of CMFs.

1.8.3 CIELUV

A major drawback of the CIE 1931 xyY chromaticity diagram is that it is not perceptually uniform. That is to say there are larger colour appearance changes associated with a given separation of points in some regions of the chromaticity diagram than in others. The CIE 1976 $L^*u^*v^*$, often abbreviated to CIELUV, colour space was the first attempt to design a perceptually uniform colour space (CIE, 1976-b). The $L^*u^*v^*$ values are derived from a non-linear combination of the CIE 1931 XYZ tristimulus

values. In this thesis, we use only the CIELUV colour space for its associated chromaticity diagram, the CIE u'v' chromaticity diagram. The CIE u'v' chromaticity coordinates are calculated from non-linear combinations of the CIE 1931 xy values, as in Eq. 1.15:

$$u' = \frac{4x}{-2x+12y+3}$$
$$v' = \frac{9y}{-2x+12y+3}. \tag{1.15}$$

1.8.4 CIELAB and ΔE

1976 also saw the development of the CIELAB, or CIE 1976 $L^*a^*b^*$ colour space (CIE, 1976-a). Like the CIELUV space, CIELAB was designed to be perceptually uniform. The L^* value is common between the two colour spaces, and is an expression of perceptual lightness, where 0 is black and 100 is diffuse white. Crucially, this lightness value is relative to a predefined white point to which the observer is adapted (and in fact, so are the u^*v^* and a^*b^* values). The CIELAB space differs from the CIELUV space as the associated chromaticity coordinates for the CIELAB space, a^*b^*, are designed to represent the unique hues of colour vision. The a^* axis is intended to represent the red-green opponent colours, while the b^* axis represents blue-yellow.

Like CIELUV, CIELAB has been widely adopted by colour vision scientists, engineers, and the display industry. Of particular prevalence is the use of CIELAB to compute the colour difference, ΔE, between two colours as defined in CIELAB coordinates (CIE, 2000). The first proposed version of the colour difference formula is simply the Euclidean distance between the CIELAB coordinates of two colours, $L_1^*a_1^*b_1^*$ and $L_2^*a_2^*b_2^*$, as in Eq. 1.16:

$$\Delta E = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2}. \tag{1.16}$$

In this original formula for ΔE a value of approximately 2.3 corresponds to the just noticeable difference (JND) between two colours. However, this original formula for ΔE contained perceptual non-uniformities, and thus the formula for ΔE was updated in both 1994 and again in 2000 to reflect this. The CIE 2000 ΔE formula computes ΔE from the hue, lightness, and chroma of the two colours, including necessary corrections. The formula for the CIE 2000 ΔE is fairly involved, and it thus not shown here, but is the one used to calculate ΔE throughout this thesis. The CIE 2000 ΔE is designed such that a value of 1 corresponds to the JND (Sharma & Bala, 2017).

1.8.5 MacLeod-Boynton Chromaticity Diagram

The colour spaces described above show a range of attempts to quantify the perceptual dimensions of colour. An alternative method is to quantify the physiological dimensions of colour directly i.e. the relative cone photoreceptor signals. As described in Section 1.5.1 one way to do this is using the DKL colour space. An alternative approach, which also represents chromatically opponent neural pathways, is the MacLeod-Boynton chromaticity diagram (MacLeod & Boynton, 1979). In the MacLeod-Boynton chromaticity diagram the relative S cone signals compared to the summed L and M cone signal, $s_{\mathrm{MB}}(\lambda)$, is plotted against the L cone signals relative to the summed L and M signals, $l_{\mathrm{MB}}(\lambda)$, where $s_{\mathrm{MB}}(\lambda)$ and $l_{\mathrm{MB}}(\lambda)$ are defined as in Eq. 6.12:

$$s_{\mathrm{MB}}(\lambda) = \frac{a_S T_S(\lambda)}{V(\lambda)},\ l_{\mathrm{MB}}(\lambda) = \frac{a_L T_L(\lambda)}{V(\lambda)}$$

where

$$V(\lambda) = a_L T_L(\lambda) + a_M T_M(\lambda) \tag{1.17}$$

and

$V(\lambda)$ and $s_{\mathrm{MB}}(\lambda)$ peak at unity.

The MacLeod-Boynton chromaticity diagram is equiluminant by definition. The 10° cone spectral sensitivity functions (as in the CIE S026 standard) are used in the calculation of the MacLeod-Boynton chromaticity coordinates here and are normalised to unity peak (Stockman et al., 1999; Stockman & Sharpe, 2000; CIE, 2006). Thus in terms of energy, the scaling constants a_S, a_M, and a_L, required such that $V(\lambda)$ and $s_{\mathrm{MB}}(\lambda)$ peak at unity, are (Stockman & Sharpe, 2000; MacLeod & Boynton, 1979):

$$a_L = 0.692839$$

$$a_M = 0.349676$$

$$a_S = 0.0554786.$$

The traditional MacLeod-Boynton space offers a physiological representation of the cone signals, but does not include a quantification of the rod or melanopsin signals. An extension to include rod signals was proposed previously by Shapiro et al., 1996. Such a limited focus on the cone signals is pertinent throughout colorimetry, particularly in the case of describing the reproduction of light on displays. To date there are no widely used quantifications of the rod and melanopsin signal

reproduction on displays. Some proposed quantifications to fill this niche are the subject of Chapter 6 of this thesis.

1.9 Thesis overview

The human photoreceptor signals are the first step in the human sensation of light, and thus a characterization of the photoreceptor signals is a crucial component of a model of the early processing stages of vision and non-visual light mediated mechanisms. Not every aspect of the human photoreceptor signals could be covered in this introduction, which sought to introduce the relevant topics pertaining to the human photoreceptor signals that are discussed in this thesis. Some of the additional characteristics of the photoreceptor signals that one must consider for a complete model of the earliest stages of the human response to light are detailed in Chapter 7.

In the following sections I provide a brief overview of each of the five experimental chapters in this thesis.

1.9.1 Chapter 2: Individual differences in cone photoreceptor signals

While the degree of variation in the cone photoreceptor signals and cone photoreceptor mosaic between individuals is known to be large, what remains unknown is the extent to which these differences can be explained by group differences between individuals vs random variation. In Chapter 2, I discuss evidence of systematic differences between individuals related to ethnicity, biological sex, and environmental light exposure. I focus in particular on two pertinent differences between individuals: the frequency of a key polymorphism on the *OPN1LW* gene which drives relatively large differences in the L cone spectral sensitivity, and the relative number of L and M cones in the retina.

1.9.2 Chapter 3: Exploring the statistical regularities in photoreceptor signals from surfaces for illuminant perception with machine-learning

In natural scenes, illuminant colour can provide biologically relevant information about time-of-day, and local conditions of shade and heading. Illuminant estimation is also a vital step in many models of colour constancy, in which surfaces maintain stable colour appearance under changing illumination. Yet, in a world without physical constraints, a single sample of reflected light from a surface could in principle be produced from an infinite set of possible surface-illuminant pairs, and illuminant estimation is mathematically under constrained. This chapter presents a machine-learning approach

to quantify, for a dataset of real-world surface and illuminant spectra, the limits that can be placed on illuminant estimation from photoreceptor signals. Radiance spectra are reduced to five photoreceptor coordinates (the three cone classes, the rods, and melanopsin), which are in principle available to a human observer. Algorithms were trained on the combination of photoreceptor signals associated with 49,667 real-world surfaces labelled by the illuminant under which the surfaces were placed. In Chapter 3, I evaluate the properties of the surface statistics, illuminants, and photoreceptor signals that impact performance of the algorithm to predict illuminant chromaticity. This approach also allows me to quantify the advantage gained from combining signals from multiple surfaces, and the optimal strategy of such combination.

1.9.3 Chapter 4: Online visual psychophysics: is it feasible to study photoreceptor signals with online experiments?

Display calibration is crucial for accurate and precise reproduction of visual stimuli. In recent years, there has been a growing interest in running experiments using visual stimuli online to access larger and broader participant pools. This interest has gained additional momentum due to the COVID-19 global pandemic, which has restricted access to lab-based testing environments. This chapter considers perceptual methods that could be used for remote display calibration for online testing. Whilst online testing is attractive in terms of participant recruitment, the use of uncalibrated displays can result in large errors in reproduction of contrast, spatial frequency, chromaticity, and other stimulus characteristics, which can in turn lead to misinterpretation of the data collected. In Chapter 4, I present results of simulations showing the importance of these considerations. I review existing perceptual calibration methods that could be used to calibrate stimuli for online testing, discuss their limitations, and suggest additional approaches that could be used in the future to improve methods of perceptual calibration. Additionally, I discuss specific applications in visual science for which subsets of the requirements of display calibration might be relaxed.

1.9.4 Chapter 5: Demonstrating a multi-primary high-dynamic range display for photoreceptor isolating experiments

Chapter 5 describes the design, construction, calibration, and characterization of a multi-primary high-dynamic range (MPHDR) display system for use in vision research. The MPHDR display is the first system to my knowledge to allow for spatially controllable, high-dynamic range stimulus generation using multiple primaries. This chapter demonstrates the high luminance, high-dynamic

range, and wide colour gamut output of the MPHDR display. During characterization, the MPHDR display achieved a maximum luminance of 3200 cd/m^2, a maximum contrast range of $3,240,000:1$, and an expanded colour gamut tailored to dedicated vision research tasks that spans beyond traditional sRGB displays. I discuss how the MPHDR display could be optimized for psychophysical experiments with photoreceptor isolating stimuli achieved through the method of silent substitution. I present an example case of a range of metameric pairs of melanopsin isolating stimuli across different luminance levels, from an available melanopsin contrast of 117% at 75 cd/m^{2}to a melanopsin contrast of 23% at 2000 cd/m^2.

1.9.5 Chapter 6: Novel metrics for characterizing the reproduction of photoreceptor signals

In Chapter 6 three novel, physiologically relevant display metrics to characterize the full photoreceptor response to light are proposed, along with a visualization of the photoreceptor response in the form of a new equal-luminance photoreceptor excitation diagram (an extension of the MacLeod-Boynton chromaticity diagram). These methods allow for the quantification of photoreceptor signals beyond the cones to evaluate the reproduction capabilities of different displays, and could be used to inform the optimal design of multi-primary displays for the reproduction of real-world photoreceptor signals. The three metrics also allow a characterization of the fundamental problem of photoreceptor signal distortions introduced by displays with fewer than five primaries. The metrics and visualizations proposed here are not envisaged to be the only metrics used in the future to quantify the full human response to light from displays, but rather, a starting point for evaluating photoreceptor signals currently ignored in common display metrics.

1.10 Acknowledgements

The work presented in this chapter is entirely that of the author. Some of the text and figures in this chapter have been adapted from a publication (Hexley, Yöntem et al., 2020), of which the author of this thesis, ACH, is the first author, and for which ACH was awarded the 2020 JOSA A Emerging Researcher Best Paper Prize. Additionally some of the text was adapted from a preprint (Hexley et al., 2021b), of which ACH is the first author.

Chapter 2

Individual differences in cone photoreceptor signals

Contents

2.1 Introduction 62

2.1.1 Individual differences in cone spectral sensitivities 62

2.1.2 Individual differences in L:M ratio 64

2.1.3 Systematic differences between individuals 66

2.1.4 Perceptual consequences of individual differences 73

2.1.5 The scope of studying individual differences 77

2.2 Experiment 1: Ethnic differences in the *OPN1LW* L(S180) vs L(A180) allele frequency 77

2.2.1 Introduction 77

2.2.2 Methods 80

2.2.3 Results 81

2.2.4 Discussion 83

2.3 Experiment 2: Systematic differences in L:M ratio 85

2.3.1 Introduction 85

2.3.2 Methods 88

2.3.3 Results 92

2.3.4 Discussion 96

2.4 General Discussion 97

2.5 Conclusion 101

2.6 Acknowledgements 102

2.1 Introduction

As introduced in Chapter 1, the cone spectral sensitivities and their inputs into luminous efficiency are often specified by standard observer models. These models have been refined over the years (Stockman et al., 1999; Stockman & Sharpe, 2000; Smith & Pokorny, 1975), and as the main starting point in this chapter I take the standard model as adopted by the CIE (Stockman et al., 1999; Stockman & Sharpe, 2000; CIE, 2006). While such standard observer models are useful tools in colorimetry, they fail to account for the large variation across individuals (Webster & Macleod, 1988; Webster, 2015). Variability between different individuals can occur at many different levels of the early stages of colour vision, and to differing degrees. At the extremes, one has colour anomalous and colour deficient observers, but even within the colour normal population there are large individual differences that are not accounted for within the standard observer model. This chapter focuses on the two broad types of individual differences: individual differences in the cone spectral sensitivity functions and individual differences in the relative numbers of L and M cones in the retina.

2.1.1 Individual differences in cone spectral sensitivities

The cone spectral sensitivities are measured at the cornea, meaning there is room for individual differences at both pre-receptoral filtering stages in the eye and at the photoreceptor level. Before light even reaches the cones at the back of the retina, it is filtered by the lens and macular pigment, the density of each of which, and thus the effective filtering of the incident light, can vary between observers (Webster & Macleod, 1988; Asano et al., 2016; Emery & Webster, 2019). Then, when the light does reach the cones, differences in the optical density of the photopigment and in the wavelength of peak sensitivity result in differences in the width and shift respectively of the cone spectral sensitivities between individuals (Stockman & Sharpe, 2000; CIE, 2006; Elsner et al., 1993; Winderickx et al., 1993; Asenjo et al., 1994; Neitz & Neitz, 1998; Sharpe et al., 1998; Merbs, S. L., and Nathans, J., 1992).

As described in Chapter 1, within our current understanding there are eight physiological factors that underpin the derivation of the cone spectral sensitivities, namely: the optical density of the lens, $D_{lens}(\lambda)$, the optical density of the macular pigment, $D_{mac}(\lambda)$, the optical density of the L cones, D_L, the M cones, D_M, and the S cones, D_S, and the peak wavelength sensitivity of the L cones, λ_{max_L}, the M cones, λ_{max_M}, and the S cones, λ_{max_S}. Thus, differences between individuals can be explained

by differences in these parameters. However, it is tricky to measure these physiological properties for a given individual, and thus tricky to estimate any given observer's individualised cone spectral sensitivity functions.

Some of the variation between individuals is correlated with easily measurable parameters, namely age and field size. The degree of variability explained by age and field size is well quantified, and described by the CIE06 model (as explained in detail in Chapter 1) (CIE, 2006). Thus, if one knows an individual's age and the field size at which they are viewing the stimulus, one can use the CIE06 model to improve the estimates from the means-based standard observer model towards personalised estimates of the cone spectral sensitivities for that individual (Stockman et al., 1999; Stockman & Sharpe, 2000; CIE, 2006). For example, the personalised cone spectral sensitivities for a 20-year-old observer at a field size of 2° vs a 60-year-old observer at a field size of 10°, compared to the Stockman-Sharpe 10° standard observer are depicted in terms of energy in Fig. 2.1.

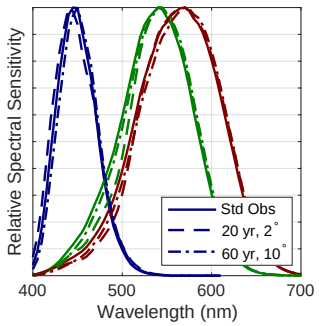


Figure 2.1: Differences in the cone spectral sensitivities, depicted in energy for a 2° 60-year-old observer, a 10° 20-year-old observer, and the 10° 32-year-old standard observer.

There is still additional variability between individuals, beyond that captured by age and field size in the CIE06 model. This variability was characterised by a model developed by Asano et al., 2016. Asano et al., 2016, estimate the expected variability in the eight physiological parameters, $D_{lens}(\lambda)$, $D_{mac}(\lambda)$, D_L, D_M, D_S, λ_{max_L}, λ_{max_M}, and λ_{max_S}, from a systematic literature review of reports of such variability. The Asano et al., 2016, model simulates the cone spectral sensitivities for populations of observers who exhibit individual variability in these parameters via a Monte Carlo simulation. An example of the variability in the 10° cone spectral sensitivities one can expect between a population of 50 colour normal observers, as generated by the Asano et al., 2016, Monte Carlo simulation, is shown in Fig. 2.2. Their model can be used to put boundaries on the differences seen in the colour normal population which is useful to inform ergonomic colorimetric design in displays and lighting.

What the Asano et al., 2016, model does not offer, however, is a full quantification of the systematic

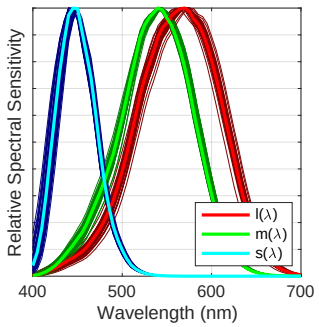


Figure 2.2: Expected extent of individual differences in the 10° cone spectral sensitivities in a population of 50 colour normal observers between the ages of 20 and 60. Observer age, along with the difference between each of the eight physiological parameters ($D_{lens}(\lambda)$, $D_{mac}(\lambda)$, D_L, D_M, D_S, λ_{max_L}, λ_{max_M}, and λ_{max_S}) compared to the means-based estimate of those parameters, is randomly sampled from the expected distribution of the variability seen in the population. For age, this is implemented by randomly selecting an age from a uniform distribution between the ages of 20 and 60. For the eight physiological parameters each parameter is randomly sampled from a normal distribution where the mean is the value of the parameter in the CIE06 model, and the standard deviation as estimated by Asano et al., 2016.

mechanisms behind this individual variability. Such quantification of the systematic drivers is interesting from a basic biological perspective to understand what factors contribute to individual variation in the cone spectral sensitivities, and thus the very earliest stages of colour processing. Without such a quantification, the Asano et al., 2016 model, does not offer an improved estimate of the spectral sensitivities for a *given* individual observer over the CIE06 model (for whom one does not know the values of the eight underlying physiological factors). Such improved estimates are desirable for given observers as these will ultimately feed into industrial applications of personalised colorimetry.

2.1.2 Individual differences in L:M ratio

Thus far we have considered differences in the cone spectral sensitivities between and within individuals i.e. differences that alter the amount of light absorbed by a single photoreceptor of each type. Yet, this is only the very first stage of vision. After phototransduction at the photoreceptor stage, the signals from the three cone classes are combined in the post-receptoral pathways: S-(L+M), L-M, and L+M, as discussed in Chapter 1. There is less known about the factors affecting retinal circuitry of the post-receptoral pathways, and thus, less known about the individual variation that occurs between them.

One known individual difference at the post-receptoral level is driven by the large individual differences observed in the relative numbers of L and M cones in the retinal mosaic (Carroll et al., 2002; Hofer et al., 2005; Carroll et al., 2000). The summed L and M signals feed into the post-receptoral luminance pathway, and thus differences in the relative numbers of L and M cones in the retina result in

differences in luminous efficiency between individuals, as shown in Fig. 2.3. The cone photoreceptors are arranged in a hexagonal packing arrangement in the retina, with smaller, more densely packed cones towards the fovea, and larger, sparser cones towards the periphery. S cones, which are not believed to feed into the luminance pathway (Lee et al., 1988, 1989; Dacey & Lee, 1994; Sun et al., 2006a, 2006b), are absent from the fovea, and rare elsewhere, accounting for only 5-10% of the cones (Curcio et al., 1991; Hofer et al., 2005; Roorda et al., 2001; Sabesan et al., 2015). The remaining cones are either L or M, which are arranged in an irregular, though not necessarily random, arrangement (Hofer et al., 2005; Roorda et al., 2001).

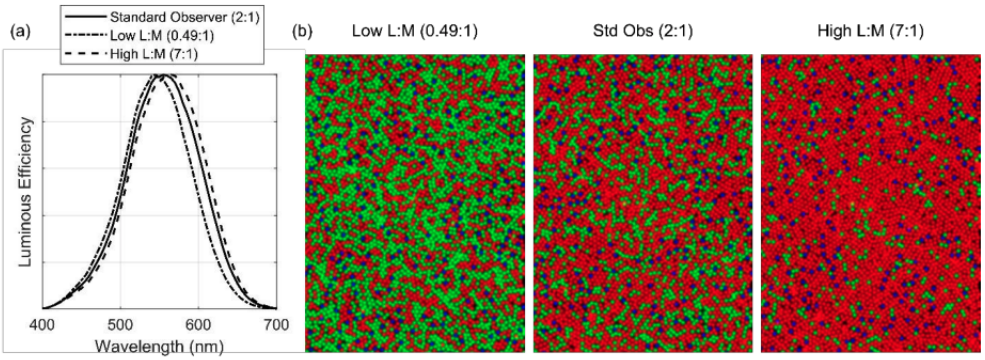


Figure 2.3: Individual differences in relative numbers of L and M cones in the retina. (a) The relative numbers of L and M cones underpins the shape of the luminous efficiency spectral sensitivity function, and thus differences in the relative numbers of L and M cones results in differences in luminous efficiency. (b) Simulations of the cone mosaic packing arrangement of three individuals' retinas, one with a low L:M ratio of 0.8:1, one with the average reported ratio of 2:1, and one with a high L:M ratio of 7:1 (Young & Smithson, 2021).

Numerous studies have investigated the degree of individual variability in the L:M cone ratio, and estimates range from 0.49:1 up to 16:1, with the average observer (as defined in the standard observer model) having a ratio of 2:1 (Sharpe et al., 2005; Carroll et al., 2002; Hofer et al., 2005; Carroll et al., 2000). It is worth being upfront about the uncertainty in estimates of L:M ratio. Most estimates of L:M ratio have relied on indirect methods such as heterochromatic flicker photometry (HFP) and electroretinography (ERG). Generally, the estimates derived from ERG and HFP have been shown to be similar (Kremers et al., 2000). However, the method for calculating L:M ratio from a HFP/ERG setting relies on using a template of the observer's cone spectral sensitivities (see Section 1.3.2, Eq. 2.2). Yet, we have just discussed that there is a large degree of variability in observers' cone spectral sensitivity functions, and thus part of the variability in HFP/ERG settings is likely to be due to variation in the cone spectral sensitivity functions between observers rather than L:M ratio. In fact, recent modelling work has suggested that all but the more extreme HFP/ERG settings could be accounted for by differences in lens density, macular pigment density, cone photopigment optical density, and the peak spectral

sensitivity of the L cone (He et al., 2021). In particular, it has been repeatedly shown that estimates of L:M cone ratio are highly sensitive to a particular single nucleotide (SNP) polymorphism on the L cone, which results in a serine to alanine substitution at codon 180 of exon 3, and a 2-7 nm shift in the λ_{maxL} of the L cone spectral sensitivity (Carroll et al., 2000; Bieber et al., 1998). Thus, throughout this chapter, I use L:M estimates from the literature where the observer's λ_{maxL} was known. It is also possible that different weightings might be given to the L and M signals in the luminance pathway for different individuals and account for some of the observed variability in HFP/ERG settings, and thus luminous efficiency, though at present there is no means by which to measure this.

Perhaps the most reliable estimates of L:M ratio we have are thanks to recent advances in retinal imaging technology which have allowed for a direct assessment of the cone ratio and packing arrangement in the living human retina (Hofer et al., 2005; Sabesan et al., 2015; Zhang et al., 2019). However, such cone typing retinal imaging techniques are difficult and long, and thus not many retinas have been characterised through such methods. These methods may also suffer from systematic errors of their own e.g. misclassifying neighbouring cones where there is the possibility of blur/contamination in the signals. Reassuringly for our indirect techniques, retinal imaging-based L:M measures broadly agree with the estimates from ERG, though ERG estimates seem to over-estimate the number of M cones compared to retinal imaging estimates (Hofer et al., 2005).

There is still a lot to be understood about the variability in the packing arrangement of L and M cones in the retina between individuals, and the impact of the irregular mosaic on vision. Further, the underlying driving mechanisms for determining the L:M ratio and arrangement in the cone mosaic is poorly understood. A more detailed discussion of the perceptual consequences of the differences in L:M ratio is found in Section 2.1.4, while a more detailed discussion of the underlying mechanisms responsible for the individual differences can be found in Section 2.1.3 below.

2.1.3 Systematic differences between individuals

Genetics

Before discussing genetic causes of systematic differences between individuals, I first briefly review the genetics of colour vision, which is relevant to understand such differences. The L, M, and S opsins are coded for by the *OPN1LW*, *OPN1MW*, and *OPN1SW* genes. *OPN1LW* and *OPN1MW* lie in tandem on the X chromosome, while *OPN1SW* is found on chromosome 7 (Nathans, Piantanida et al., 1986). *OPN1LW* and *OPN1MW* are >96% homologous, while both are only about 40% homologous to *OPN1SW* (Nathans, Thomas et al., 1986). This is believed to be due to the fact that the *OPN1MW*

and *OPN1LW* distinction is a relatively recent divergence, evolutionarily speaking (Nathans, Thomas et al., 1986; Neitz & Neitz, 2011). *OPN1LW* lies upstream of *OPN1MW* on the X chromosome, and it is typically believed that individuals have just one copy of *OPN1LW*, but multiple copies of *OPN1MW* (though it is also believed only the first copy of *OPN1MW* is expressed) (Nathans, Thomas et al., 1986). *OPN1LW* and *OPN1MW* expression is controlled by the locus control region (LCR), which lies upstream of the opsin tandem on the X chromosome, and which must bind to the promoter of one of the opsin genes to drive expression (Wang et al., 1992; Smallwood et al., 2002). It is believed that only one of the X-linked opsin genes is expressed in any given cell (Baylor et al., 1987; Mollon, 1999). Another important point about *OPN1LW* and *OPN1MW* expression to note, is that due to their location on the X chromosome, XX individuals have an extra copy of *OPN1LW* and *OPN1MW* than XY individuals. However, random X inactivation means that in any given cone cell only one of the two X chromosomes, and thus one of the *OPN1LW*/*OPN1MW* genes, is expressed (Lyon, 1961; Teplitz, 1965).

Differences in the opsin gene sequences between individuals are the major drivers of colour vision deficiencies (CVD) and are a considerable source of variation in the colour normal population (Neitz & Neitz, 2011). How do such sequence differences occur? Perhaps the most common cause is unequal recombination (Nathans, Piantanida et al., 1986; Mollon, 1997). During meiosis, where the paternal and maternal chromosomes line up and exchange pieces of DNA, a misalignment between the maternal and paternal chromosome can result in gene loss (if the unequal cross-over occurs between genes) or hybrid genes (if the unequal cross-over occurs within genes). Due to the high homology and close proximity of *OPN1LW* and *OPN1MW*, a misalignment of the two chromosomes during meiosis and thus such unequal cross-over is common (Nathans, Piantanida et al., 1986; Mollon, 1997). When such unequal cross-over occurs between genes, resulting in gene loss, dichromacy occurs (protanopia if the gene loss is the L gene, and deuteranopia if the loss is the M gene). Approximately 1 in 50 Caucasian XY individuals are either protanopes or deuteranopes (Neitz & Neitz, 2000). Protanopia or deuteranopia can also occur if unequal cross-over occurs within the genes, and results in a loss-of-function gene (Mollon, 1997; Neitz & Neitz, 2000).

Such loss-of-function in the L and M opsin genes is believed to be relatively uncommon, as unequal cross-over within a gene typically results in a *functional* hybrid gene, but one of a different peak spectral sensitivity than the original gene, with a λ_{max} between λ_{maxL} and λ_{maxM} (Mollon, 1997). Individuals with such hybrid genes will typically have a normal L/M opsin gene, and a hybrid L'/M' opsin gene. Thus such XY individuals will still express three different photopigments in their retina, but the L and M photopigments will be closer in spectral sensitivity than in a colour normal observer. These

individuals are known as anomalous trichromats (Neitz & Neitz, 2000). Deuteranomaly, where the individual has an intact L gene and an L' (i.e. an anomalous M) gene, is the most common case of anomalous trichromacy, affecting 1 in 20 Caucasian XY individuals (Jordan et al., 2010). Protanomaly, where the individual has an intact M gene and an M' (an anomalous L gene), affects about 1 in 100 Caucasian XY individuals (Jordan et al., 2010).

As with X-linked dichromacy, X-linked anomalous trichromacy is rare in XX individuals, as they need both copies of the L or M gene to be anomalous in order to exhibit the resulting anomalous phenotype (Jordan et al., 2010). However, around 12% of XX individuals are carriers of a hybrid pigment (Jordan et al., 2010). Due to random X inactivation, such XX individuals that are heterozygous for the normal and hybrid genes will in fact express four photopigments in their retina, the S opsin, along with the normal L and M opsins and the hybrid L or M opsin. Such individuals are known as *weak tetrachromats* (Jordan et al., 2010). The exact arrangement of the normal and hybrid L and M cones in their retina is unclear, as is how these different L and M cones may feed into post-receptoral pathways. There is evidence that one XX individual who is a carrier of anomalous trichromacy is able to utilise all four cone signals and enjoy behavioural tetrachromacy, though the majority of weak tetrachromats do not enjoy such luxury (Jordan et al., 2010).

Sequence differences in *OPN1SW* are less common, and so too are the resulting colour vision deficiencies: tritanopia (the loss of function in the S gene) and tritanomaly (where an individual has an anomalous S pigment). Estimates of the frequency of tritanopia range from 1 in 500 to 1 in 65,000 (Wright, 1952; Went & Pronk, 1985; Simunovic, 2010). S-linked colour deficiencies are rare because recombination errors are much less common in *OPN1SW* which sits on an entirely different chromosome to *OPN1LW* and *OPN1MW*. In fact, the main driver of S-linked colour deficiencies are point mutations that occur in *OPN1SW* (Mollon, 1997). Point mutations can also cause X-linked colour deficiencies, though many SNP point mutations that occur in *OPN1LW* and *OPN1MW* lead to only small shifts in the spectral tuning of the L and M photopigment (Neitz & Neitz, 2011). Such shifts are thus classified as part of the variation within the colour normal population, rather than colour deficiency. There are several common SNPs that occur in one of the six exons in the *OPN1LW* and *OPN1MW* sequences that shift the λ_{maxL} and λ_{maxM} in the colour normal population.

L(S180A): Of particular interest is the common SNP in exon 3 of the *OPN1LW* gene that causes a serine (S180) to alanine (A180) substitution at codon 180, that changes the λ_{maxL} from ~ 559 nm to ~ 555.5 nm, resulting in a 2-7 nm shift in the peak wavelength of the L cone spectral sensitivity function (Asenjo et al., 1994; Merbs & Nathans, 1992; Neitz & Neitz, 1998; Sharpe et al., 1998; Nathans et al.,

1992). Individuals with L(S180) will have an L cone spectral sensitivity that has a λ_{maxL} towards longer wavelengths i.e. greater separation between the L and M cone spectral sensitivities, while individuals who have the minor allele, L(A180), will have an L cone spectral sensitivity with a lower λ_{maxL}, i.e. less separation between the L and M cone spectral sensitivities. This shift is relatively large and particularly prevalent compared to other SNP induced shifts seen in the colour normal population.

Genetics of L:M Ratio: Little is known about the genetic underpinnings of the L:M ratio. However, L:M ratio has been suggested to have some degree of hereditability (De Vries, 1949), which has since been demonstrated empirically (Lutze et al., 1990), pointing to a genetic underpinning of the ratio (Lawrance-Owen et al., 2014). More recent evidence from retinal imaging to assess the structural arrangement of L:M cones in the retina demonstrates a non-random arrangement of the L and M cones in three XY individuals, which indicates non-random processes underline the determination of whether a cone is L or M (Hofer et al., 2005). A key candidate for such non-random behaviour would lie in genetic control.

It has been suggested that the relative proximity of *OPN1LW* and *OPN1MW* to the LCR is responsible for the favoured expression of L over M in the retina (Smallwood et al., 2002; Wang et al., 1999). It has also been suggested that differences in the structural layout of the X chromosome gene array may contribute to the differences in L:M ratio (McMahon et al., 2008). Various mechanisms for this have been suggested. Originally, it was suggested that this may simply be due to differences in the order of *OPN1LW* and *OPN1MW* on the X chromosome array (Smallwood et al., 2002; Wang et al., 1999). It has also been suggested that differences in the sequences of the *OPN1LW* and *OPN1MW* gene promoters may be driving the ratio (Smallwood et al., 2002). Another candidate mechanism attributes the differences in the ratio to differences in the length of intron 1 in *OPN1LW* (Mollon, 1999). However, it has been shown that significant differences in L:M ratio can be seen, despite nearly identical layouts in the X chromosome gene array (McMahon et al., 2008). Therefore, any role of the LCR in the control of L:M ratio must be a result of the more complex three-dimensional structure of the X chromosome array (McMahon et al., 2008). Additionally, RXRγ, the retinoid X receptor γ, has been implicated in control of the ratio (Lawrance-Owen et al., 2014). RXRγ, which is coded for by the *RXRG* gene on chromosome 1, is known to play a role in the differentiation of cones during development (Onishi et al., 2010; Swaroop et al., 2010), and it has been shown that a SNP upstream of *RXRG* has a tentative association with OSCAR test setting (Lawrance-Owen et al., 2014) (where the OSCAR test is a method to test relative sensitivity to red and green light, and thus taken as a proxy for L:M ratio (Estévez et al., 1983)).

Biological sex

It is worth re-emphasising the well-established consequence of the fact that the L and M cone genes lie on the X chromosome for the relationship between biological sex and colour vision. Due to random X inactivation, XX individuals that are heterozygous for any of the L and M hybrid genes will in fact express L and/or M cones with both sensitivities in their retina. This means that XX carriers of protanopia or deuteranopia, i.e. with one chromosome either missing *OPN1LW* or *OPN1MW*, will express disproportionately more M or L cones, than the cone type in which they are deficient. Similarly, XX carriers of protanomaly or deuteranomaly, i.e. with an anomalous copy of *OPN1LW* or *OPN1MW*, will express fewer normal M or L cones and more L/L' or M/M'. Thus, on average, XX carriers of protan or deutan deficiency will have a lower or higher L:M ratio than XY individuals, or XX individuals who are not carriers of CVD. This has been experimentally verified by Jordan and Mollon, 1997, who investigated L:M ratios of the mothers of sons with CVD. Biological sex also contributes to differences in macular pigment density, which is higher in XY individuals than XX individuals (Hammond Jr et al., 1996).

Environmental factors

As with any aspect of human biology, genetics cannot work alone, and there exists a delicate balance between the interactions of genes and environment in determining phenotype. Environmental and lifestyle factors are well-known to contribute to individual differences in macular pigment and lens density, which in turn contribute to the cone spectral sensitivities. For example, smoking increases the lens density (Hammond et al., 1999), and diabetes accelerates lens density increases (Lutze & Bresnick, 1991), while increase in dietary carotenoid intake increases macular pigment density (Ciulla et al., 2001; Bone et al., 2003), as does increased body fat (Nolan et al., 2004).

It is also possible that the expression of cone opsins may be optimised for the given environmental light. Humans are not born with the same retinal mosaic they exhibit as an adult. Foveal cone density increases during the first 4-6 years of life (Hendrickson & Zhang, 2019), from approximately only ~25,000/mm^2 at birth, compared to ~200,000/mm^2 during adulthood (Hendrickson & Zhang, 2019; Yuodelis & Hendrickson, 1986), most likely due to migration of cones after birth towards the fovea, as there is no evidence for generation of new cone cells postnatally (Hendrickson & Zhang, 2019; Hendrickson et al., 2016). The fovea is immature at birth (Abramov et al., 1982), narrowing and deepening after birth (Yuodelis & Hendrickson, 1986), and is believed to reach maturity sometime between 15 and 45 months (Hendrickson & Yuodelis, 1984). Additionally, at birth, cone outer segments

are short and immature, and elongate by a factor of seven by 15 months, when they are still only half of the length of the outer segment observed in the adult retina (Yuodelis & Hendrickson, 1986). This postnatal development of the fovea leaves scope for the influence of the lighting environment during development to influence foveal packing, and therefore variation in local L:M ratio.

Perhaps even more importantly, there is evidence which shows transient S opsin expression in L/M cells in the foetal and neonatal human retina (Xiao & Hendrickson, 2000; Cornish et al., 2004). In particular there is a marked decrease in the number of cells which co-express S + L/M in the 8-month retina (0.01-0.03% of the cone population), compared to the number in the 4-day postnatal human retina (3-10% of the cone population), which suggests that L/M opsin expression is not complete at birth (Cornish et al., 2004). The presence of S + L/M cells in the neonatal retina implies that mutually exclusive expression of an opsin is not necessary for cone survival (Xiao & Hendrickson, 2000; Cornish et al., 2004). This is in line with evidence from artificial constructs in the mouse retina, where it has been shown that when one of two genetic reporters replaces the L and M opsin gene in an artificial DNA construct expressed in transgenic mice, some cones in the mouse retinal tissue express both genetic reporters (though the vast majority of cones express one or the other) (Wang et al., 1999). Mature cones, however, do seem to rarely express multiple opsins, with only a very small number of S + L/M cones appearing in the immunostained adult retina (0-0.02% of the cone population) (Xiao & Hendrickson, 2000; Cornish et al., 2004), which is corroborated by results from electrophysiological recordings of single cones from the adult primate retina (Baylor et al., 1987; Mollon, 1999). It is possible, therefore, that at birth an L/M cone could co-express L and M opsin (and that some subset of those will also express S opsin), and that opsin expression is regulated by a postnatal process that could be influenced by the postnatal environment. To the best of my knowledge, there has not been any study to date that separates between L and M opsin expression in human tissue immunostaining, due to the high homology of the L and M cones which make them difficult to differentiate through such techniques (Nathans, Thomas et al., 1986; Xiao & Hendrickson, 2000; Cornish et al., 2004). Therefore, it is not yet known at what stage in development the L or M fate of the cone cell is determined.

Evidence from animal models does indeed suggest that environmental light conditions do influence cone fate during postnatal development. For example, in guinea pigs, who only have two cone types, M and S, M cone density was higher in a group of animals reared in an environment lit by green light compared to a group reared in violet light (Hu et al., 2011). Chickens raised in an environment with a red-light cut-off filter had reduced L opsin expression compared to chickens raised in a white light setting (Gisbert et al., 2020). Such studies suggest that early environmental light exposure contributes

to selectively determining which opsins are expressed during postnatal retinal development.

As environmental lighting exhibits a large geographical and seasonal dependence (Webster et al., 2007), if environmental lighting does contribute to cone cell fate in humans as it does in animal models, one may expect to see systematic geographical and seasonal differences in L:M ratio and cone spectral sensitivities between individuals (Webster, 2014; Hagen et al., 2019). Whether or not such an effect exists remains to be established.

Ethnicity

In addition to having an active impact on the retinal development of a given individual, environmental influences may also present themselves passively through ethnicity differences, reflecting long-term evolutionary tuning of the eye to a given geographical location (Race, Ethnicity, and Genetics Working Group et al., 2005; Liu et al., 2006). It is well-established that CVD frequency varies with ethnicity, with CVD being less common in Asian, African, and Native American populations, than Caucasian populations (Delpero et al., 2005; Birch, 2012). Ethnic differences have also been observed in both L:M ratio and L(S180) vs L(A180) allele frequency. Though to date comparisons across ethnicity have typically relied on a relatively a small number of ethnic groups, and small number of observers within each group (Sharpe et al., 2005; Carroll et al., 2002; Hofer et al., 2005; McMahon et al., 2008; Hagen et al., 2019; Deeb et al., 1995; Hayashi et al., 2001; Kuchenbecker et al., 2014; Yamauchi et al., 2013).

Generally, it is reported that the L(A180) minor allele is most frequent amongst Norwegians, where it actually is more common to have L(A180) (55%) than L(S180) (45%) (Hagen et al., 2019), followed by Caucasians (56% L(S180) and 44% L(A180) (Stockman & Sharpe, 2000)), while the minor allele is quite uncommon in Japanese observers (80% L(S180) and 20% L(A180) (Deeb et al., 1995; Hayashi et al., 2001)) and rare among African/African-Americans (96% L(S180) and 4% L(A180) (McMahon et al., 2008)). (Note, here "Caucasian" is used to refer to the sample of observers from which the 56/44 frequency estimate is derived i.e. British and central European Caucasian individuals. Norwegians is used to refer to the Norwegian Caucasian population, which is noted separately from Caucasians here simply because the Norwegian-specific estimate was derived later than, and is different from, the original "Caucasian" estimate).

Further, the highest reported average L:M ratio is from a sample of Norwegian observers (~3.95:1 (Hagen et al., 2019)), while East Asian (~1:1 (Kuchenbecker et al., 2014)) and African/African-American observers (~1.86:1 (McMahon et al., 2008)) have been reported to have lower ratios on average than the average observed amongst Caucasian observers (~2.6:1 (Sharpe et al., 2005; Carroll et al., 2002; Hofer

et al., 2005)). One must however bear in mind that some of the perceived difference in L:M ratio with ethnicity may be due to confounding systematic differences in the cone spectral sensitivities across ethnicity. For example, ethnicity has also been shown to correlate with macular pigment density, with Black observers having a lower macular pigment density, and South Asian observers a higher macular pigment density than Caucasian observers (Iannaccone et al., 2007; Howells et al., 2013).

Ethnicity and biological sex differences are particularly intriguing, as not only do they provide strong evidence for a genetic underpinning of observed individual differences, (and in the case of ethnicity, a hint of a potential environmental driver of such genetic drift), if well characterised they may also provide an easily measurable parameter, like age and field size, to be fed into models of individual differences to tune the cone spectral sensitivities for a given observer. Thus ethnicity and biological sex differences not only provide insight onto the biological mechanisms underlying individual differences, but also an opportunity to account for such differences in ergonomic colorimetry design and vision science experiments.

2.1.4 Perceptual consequences of individual differences

Colour appearance

Despite the individual differences at the retinal level, perceptually, colour experience appears to remain remarkably stable both across and within observers (Emery & Webster, 2019). At the extreme end, anomalous trichromats do not report colour experiences much different than normal trichromats, despite marked differences in their cone spectral sensitivities that would predict much greater colour appearance differences (Boehm et al., 2014; Bosten, 2019; Tregillus et al., 2021; Isherwood et al., 2020). It has been suggested that adaptive control in the postreceptoral gain which determines the relative strength of the cone signals into the postreceptoral channels may act as the compensatory mechanism to diminish such expected differences (Boehm et al., 2014). Recent work with fMRI showed a reduction in cortical response in V1 for anomalous trichromats compared to normal trichromats during both a simple fixation and high-attention fixation task with L-M- and S-opponent stimuli, but no difference in areas V2v and V3v, suggesting a possible cortical origin of this compensation (Tregillus et al., 2021).

Yet, when individual differences in colour perception do appear, it is striking. A common place where individual differences are observed is in the perception of unique hues, i.e. the red and green that contain no blue or yellow, and the blue and yellow that contain no red and green (Hering, 1878). Unique hue settings, while mostly stable within an individual, are highly variable across individuals, yet these differences do not seem correlated with individual differences in retinal mechanisms (Webster

et al., 2000). Notably, it has been shown that unique yellow settings do not correlate with L:M ratio as one might have predicted (Brainard et al., 2000). Perception of unique hues, namely unique yellow, has also been suggested to vary within an individual, and exhibit a seasonality dependence (Welbourne et al., 2015).

An interesting case is the perception of white (the "mother of all unique hues" (Mollon, 2006; Bosten et al., 2015)). Perception of white appears to be relatively stable between observers (Boehm et al., 2014), as well as within an observer over a lifespan, despite age-related changes in lens density (Werner & Schefrin, 1993). However, achromatic settings show a large shift in the "yellow" direction directly after cataract surgery (where the lens is replaced with an artificial, clear lens), gradually drifting back to the achromatic setting before surgery over time (Delahunt et al., 2004). A common feature of the observed appearance differences is the role of adaptation, which presumably operates at the post-receptoral level to scale the cone signals (von Kries, 1905), with differences resulting from different scaling of the cone signals in different conditions or in different individuals.

Chromatic and spatial discrimination

While colour appearance differences may not correlate with individual differences in the available cone photoreceptor signals, it has been shown that at the very limits of perception, i.e. discrimination *thresholds*, individuals do exhibit differences as one would anticipate from photoreceptor inputs. In particular, Hood et al., 2006 showed that chromatic discrimination thresholds of a red-green square wave grating were higher for deutan carriers than either controls or protan carriers. Deutan carriers are more likely to have a more extreme L:M cone ratio, than controls or protan carriers. Protan carries who lacks a copy of *OPN1LW* (or have an anomalous L cone gene) will comparatively over-express M cones compared to an individual who has an intact copy of both *OPN1LW* and *OPN1MW* on both chromosomes, but this over-expression of M cones will actually bring their ratio closer to a balanced retina. Therefore, if anything, one may expect protan carriers to have *reduced* thresholds compared to controls, though individual variability within the colour normal group means that such a difference would be hard to measure at the group level (Carroll et al., 2002; Hofer et al., 2005; Carroll et al., 2000; Hood et al., 2006).

It is not just chromatic discrimination thresholds that will be impacted by L:M ratio differences, but *spatial acuity* thresholds. Chromatic and spatial discrimination are inherently tied. Consider trying to discriminate between high intensity and low intensity regions of a two-dimensional achromatic square-wave grating as eye movements translate the grating across the retina. If one were to sample

the stimulus with a sampling array of just one cone type, the only relative change in signal would be due to the change in intensity of the stimulus. If, however, one were to sample the stimulus with a sampling array consisting of two cone types and extract information about the changes in intensity by the summation of the signals of each type, then the change in the signal would partially be due to changes in intensity of the stimulus, and partially due to which cone type was sampling which region of the stimulus. Therefore, the achromatic signal is predicted to be less noisy for uniform regions of a single cone type. Conversely, as we know, comparison between two cone types is crucial for colour vision, and thus chromatic signalling comes at the expense of spatial signalling. The challenge for the retina is to find the balance of the two which results in some optimal trade-off (Lewis & Zhaoping, 2006).

In terms of the individual differences in the photoreceptor signals we've discussed, there are two likely ways in which the cone signals may be balanced to maximise spatial over chromatic acuity. Firstly, as described in the example above, simply having more cones in the retinal mosaic of the same type will maximise spatial acuity, by reducing chromatic noise, at the expense of chromatic discrimination. And indeed, as explained above, deutan carriers (the extreme case, with a retina very heavily L cone dominated) do have elevated chromatic discrimination thresholds (Hood et al., 2006), though it is unclear whether they have an advantage in achromatic spatial acuity. It has also been shown that some dichromats, who by definition will have diminished chromatic discrimination, have significantly higher visual acuity than normal trichromats (Jägle et al., 2005). Perhaps most strikingly, Danilova et al., 2013, demonstrated a direct correlation between spatio-chromatic resolution and OSCAR test settings, which are a known proxy of L:M ratio. Their experiment used Landolt Cs to determine spatial resolution thresholds for L and M isolating stimuli for colour-normal observers and found that the majority of their observers had higher contrast sensitivity for L isolating stimuli than M, as one would expect in an L dominated retina. Importantly, they also show that the ratio of L to M contrast sensitivity is positively correlated with the L:M cone ratio, i.e. that the advantage of spatial resolution for L isolating stimuli increases as the relative numbers of L cones increases. Their study thus provides the first direct evidence that density of a particular cone class can impact spatial resolution for targets that isolate that particular cone class. Whether or not there is also an impact on achromatic spatial resolution, as one may predict, is yet to be tested.

Secondly, a retina could favour spatial discrimination by having an L and M cone that were closer to each other in terms of spectral sensitivity. In fact, the common *OPN1LW* S180A polymorphism would shift the L cone spectral sensitivity closer to that of the M cone, and thus make the two cone

signals more correlated, reducing chromatic noise to maximise spatial discrimination (Hagen et al., 2019). It has in fact been shown experimentally that chromatic discrimination, as measured in a spatio-chromatic contrast sensitivity task, is higher on average in individuals with L(S180) than L(A180) (Dees et al., 2015). Spatial acuity benefits have not been directly demonstrated among individuals with L(A180). There are, however, hints of evolutionary evidence, from both humans and primates, that the L(A180) minor allele may occur more frequently in individuals living in low light environments: where there is perhaps more need for the eye to favour spatial over chromatic discrimination. For example, the owl monkey, the only known nocturnal monkey, is believed to only ever express L(A180) (Jacobs, 2008). Further, ethnicity differences in L(S180) vs L(A180) allele frequency seem to correlate with seasonal minima in environmental light. The L(A180) minor allele is significantly more frequent in Norwegians, who have descended from individuals living close to the Arctic circle, and thus in extremely low-light conditions for half of the year, compared to East Asians, descended from more equatorial regions that are less light-starved (with the L(A180) minor allele reported to occur in 55% of Norwegians compared to only 20% in East Asians) (Hagen et al., 2019).

Another consideration in spatial and colour vision, is the chromatic aberrations of the eye which mean the eye cannot be simultaneously optimally focused for both L and M cones (Danilova et al., 2013). It therefore may be the case that part of the spatial acuity benefit for some dichromats arises from the eye being in focus for the single cone class (Jägle et al., 2005). For instance, if one had a predominantly L heavy retina, then one may do better having a longer focal length of the eye to preferentially focus the long wavelength information on the retina, at the expense of some defocus in the middle and shorter wavelengths. Indeed, in an experiment with chickens it has been shown that L:M ratio correlates with the focal properties of the eye, with higher L:M ratios correlating with a longer vitreous chamber depth and greater refraction (Gisbert et al., 2020; Gisbert & Schaeffel, 2018). There is also evidence that high L:M ratios, and the L(S180A) polymorphism, are associated with lower prevalence of myopia (which occurs when the eye is too long, and the focus is in front of the retina) in humans (Hagen et al., 2019; Zhou et al., 2015). While these two results suggest correlations in opposing directions, one must remember that chicken L and M cones are not directly related to human L and M cones, and thus the main takeaway here should be the fact that any correlation between eye length/myopia and L:M ratio is observed. Interestingly, in the chicken experiments, environmental lighting during development was shown to influence the relationship between L:M ratio and eye length, with chickens reared wearing red spectacles expressing fewer L cones and more hyperopia (where the eye is too short) (Gisbert et al., 2020).

In sum, it has been shown that both spatial resolution and chromatic discrimination, and indeed even the physical size of the eye, depends on individual differences in the cone photoreceptor signals, namely in L:M ratio and the L(S180A) polymorphism. There are hints that these individual differences are influenced by environmental light exposure, something which may be observed both in a given individual and in longer-term evolutionary signatures, such as ethnicity. Yet despite these observed differences of threshold level spatial resolution and chromatic discrimination with individual differences in retinal mosaic, these differences do not appear to propagate to differences in colour experience between these observers.

2.1.5 The scope of studying individual differences

Studying individual differences provides an opportunity to uncover mechanisms underlying colour vision. Firstly, systematic differences between groups of individuals can reveal the genetic and environmental driving factors underlying the individual differences in the first place. And secondly, studying the perceptual consequences of these differences, at both threshold and suprathreshold levels, can inform our understanding of the neural processing of colour. Thus, while we are continually improving our knowledge of individual differences in the cone photoreceptor signals and their consequences for vision, there is still much that we can learn from studying this variability between individuals. In other words, the scope for research into individual differences is vast. In this chapter, I present two experiments that contribute to the field of study on individual differences. In the first experiment, I present results from a population-level assessment of the frequency of the L(S180) vs L(A180) alleles across ethnicity. In the second experiment, I present an analysis of systematic differences between observers on heterochromatic flicker photometry (HFP) settings, where HFP settings are taken as a proxy of L:M ratio.

2.2 Experiment 1: Ethnic differences in the *OPN1LW* L(S180) vs L(A180) allele frequency

2.2.1 Introduction

As described above, the *OPN1LW* S180A polymorphism is of particular interest as a source of variability among the colour normal population. The reason for the special status of this polymorphism is twofold. Firstly, it leads to a relatively large spectral shift in the λ_{maxL}, and thus subsequently causes the M and L signals to be more or less correlated with each other (Asenjo et al., 1994; Merbs & Nathans,

1992; Neitz & Neitz, 1998; Sharpe et al., 1998; Nathans et al., 1992). The L(S180A) polymorphism is a sufficiently large spectral shift to cause measurable differences between individuals in performance on a psychophysical task such as the previously reported differences in spatio-chromatic sensitivity (Dees et al., 2015). Additionally L(S180) and L(A180) individuals have two distinct matching ranges on the Rayleigh anomaloscope (Dees et al., 2015; Winderickx et al., 1992; Thomas & Mollon, 2004). Thus, while a 2-7 nm shift in the L-cone spectral sensitivity may not seem large, it represents a considerable increase in the separation between the L and M cones and better signal-to-noise ratio in discrimination along this colour axis.

Secondly, this polymorphism is believed to be particularly common among the colour normal population (Stockman & Sharpe, 2000; Winderickx et al., 1993; Neitz & Neitz, 1998; Sharpe et al., 1998). Other polymorphisms, which cause similarly large shifts, such as the corresponding M(A180) to M(S180) substitution, are reported to be much less prevalent than the L(S180A) polymorphism (with 93 – 94% of Caucasian XY individuals thought to have the dominant allele (Winderickx et al., 1993; Neitz & Neitz, 1998)). The fact that the L(A180) minor allele is believed to be so frequent is important because the standard observer L cone spectral sensitivity is defined to be the weighted average of the L(S180) cone spectral sensitivity and L(A180) cone spectral sensitivity, weighted by the estimated proportions of L(S180) and L(A180) in the population.

Despite the long-held belief that the L(S180) vs L(A180) substitution is common, evidence is accumulating to suggest this is only the case among Caucasian observers. Historically, as with almost any field in science, ethnic differences have been overlooked in colour vision research, with the majority of work into the cone spectral sensitivities in the 1980s-2000s happening in American and UK based labs, recruiting (unintentionally) biased samples of Caucasian observers. In fact the four samples, with a grand total of 309 observers, from which the 56% L(S180), 44% L(A180) estimate is derived used exclusively Caucasian XY individuals (Stockman & Sharpe, 2000; Winderickx et al., 1993; Neitz & Neitz, 1998; Sharpe et al., 1998). Yet, even as early as the mid-1990s, work from Deeb et al., 1995 suggested that the L(A180) minor allele was significantly less frequent among Japanese XY individuals (16% L(A180), $n = 49$, $z = 3.71$, $p < 0.001$, *one-tailed*). Why the observed frequency among Japanese individuals was not included in the frequency estimates from which the standard observer was derived in the late 1990s is unclear. Though, in their paper Stockman and Sharpe, 2000, are careful to state that their estimate of L(S180) vs L(A180) allele frequency is for the Caucasian XY population.

Further evidence from Hayashi et al., 2001, in the early 2000s provides additional support to the observations of Deeb et al., 1995, that the L(A180) minor allele is significantly less frequent among

Japanese XY individuals than Caucasian XY individuals (21% L(A180), $n = 119$, $z = 4.40$, $p < 0.001$, *one-tailed*). Interestingly, it seems as though interest in deriving estimates of L(S180) vs L(A180) allele frequency ceased between the early 2000s and today, with very few papers reporting L(S180) and L(A180) frequencies, and those that do, doing so indirectly, when *OPN1LW* was sequenced as a means to improve estimates of L:M ratio from ERG/HFP settings (Carroll et al., 2000; Bieber et al., 1998). There is no doubt this must be at least partially driven by the infamous difficulty in sequencing *OPN1LW*. Two studies from the mid 2000s that contain reports of L(S180) vs L(A180) allele frequency in Caucasian observers report a frequency that is not statistically significantly different from the 44% L(A180) estimate used for the Caucasian population (Carroll et al., 2002: 38% L(A180), $n = 50$, $z = 0.79$, $p = 0.43$, *two-tailed*; Sharpe et al., 2005: 42% L(A180), $n = 38$, $z = 0.23$, $p = 0.82$, *two-tailed*).

Estimates of L(S180) vs L(A180) allele frequency in African/African-American observers can be derived from the data presented by McMahon et al., 2008, in their paper estimating L:M ratios in African/African-American XY observers. Again, estimates of the peak spectral sensitivity of the L cone, and thus indirectly of L(S180) vs L(A180) genotype, were measured by McMahon et al., 2008, to improve estimates of L:M ratio, and the authors do not explicitly report the frequency of L(S180) vs L(A180) in their paper. Yet, from the data reported in their paper, they find only one observer with a peak sensitivity that would indicate they had L(A180), and 26 observers with a peak sensitivity indicative of L(S180). Their results thus show that the L(S180) vs L(A180) allele frequency in African/African-American XY observers is significantly lower than the 44% L(A180) estimate from Caucasian XY individuals (4% L(A180), $n = 27$, $z = 4.06$, $p < 0.001$, *one-tailed*). The L(S180) vs L(A180) allele frequency derived from McMahon et al.,'s cohort of African/African-American observers and its difference from the frequency in Caucasian observers was first noted explicitly by Hagen et al., 2019, who also report that in their sample of Caucasian Norwegian XY individuals L(A180) is more common than L(S180) (55% L(A180)), though this increase is not statistically significant from the reported 44% L(A180) in the Caucasian XY population (55% L(A180), $n = 60$, $z = 1.57$, $p = 0.06$, *one-tailed*).

The frequencies reported above are summarised in Table 2.1. In addition to the comparisons listed above, I note that the L(S180) vs L(A180) allele frequency observed in the Japanese XY individuals, i.e. 20% L(A180) ($n = 168$, pooling the observations from Deeb et al., 1995, and Hayashi et al., 2001), is significantly higher than the L(S180) vs L(A180) allele frequency in African/African-American individuals (4% L(A180), $n = 27$, $z = 2.02$, $p = 0.02$, *one-tailed*). I also note the considerable variability of the L(S180) vs L(A180) allele frequency across different samples of Caucasian observers, with estimates ranging from as low as 26% L(A180) (Sharpe et al., 1998) to as high as 55% L(A180) (Hagen et al., 2019).

Despite the variability, none of the L(A180) proportions within the Caucasian samples are significantly different from the 44% L(A180) estimate derived from the 309 Caucasian XY observers by Stockman and Sharpe, 2000.

A major takeaway from the results in Table 2.1 is that there are still many ethnicities underrepresented. For the ethnicities that there is datum available, sample sizes are small, particularly for African/African-American observers. And many prominent ethnicities are not represented at all, such as Latino, and South Asian; while East Asian observers are only represented by Japanese individuals. There is thus a need to estimate the L(S180) vs L(A180) allele frequency across a much larger cohort of observers and a more expansive set of ethnicities. In this experiment, I make use of the large-scale exome and genome sequencing database gnomAD (Genome Aggregation Database v2.1.1) to obtain such estimates (Karczewski et al., 2020).

2.2.2 Methods

gnomAD is an aggregated dataset from a wide-range of large-scale exome and genome sequencing projects (Karczewski et al., 2020). gnomAD v2.1.1 contains a total of 125,748 exome and 15,708 whole-genome sequences. gnomAD contains data from seven ethnic groups: Latino/Admixed American, Ashkenazi Jewish, European (non-Finnish), European (Finnish), South Asian, East Asian, and African/African-American. Note, that gnomAD separates European observers into Finnish and non-Finnish Europeans, due to repeated population bottlenecks in the Finnish population. Here, I use the term Caucasian to mean non-Finnish European as this most closely matches the populations that informed the estimate of L(S180) vs L(A180) allele frequency by Stockman and Sharpe, 2000. Any individuals' sequences that do not classify clearly into one of the seven major groups are classified as Other. The numbers of exome and genome sequences available for each ethnicity in the gnomAD v2.1.1 database are shown in Table 2.2.

The exome and genome sequencing projects incorporated in the gnomAD dataset use next-generation sequencing (NGS) technology. Briefly, NGS works by breaking the genome into millions of small fragments of DNA, sequencing each small fragment, and then using computational techniques to reconstruct the full genome/exome sequence from these fragments. gnomAD makes summary data, including population statistics of allele frequency, available to browse freely on their website, but full sequence data is not publicly available, bar a few examples. For this experiment, I accessed the freely available data on *OPN1LW* and analysed the estimated frequencies of the L(S180) vs L(A180) allele frequency in the *OPN1LW* aggregate population data on the basis of ethnicity.

Study	**Observers**	**n**	**Fraction L(A180)**
Winderickx et al., 1993	Caucasian American XY individuals (74 colour normals, 13 single-gene deuteranopes, 22 deutan defect) colour normal XY individuals	109	0.44
Deeb et al., 1995	Japanese XY individuals	49	0.16
Neitz & Neitz, 1998	Caucasian American colour normal XY individuals	130	0.48
Sharpe et al., 1998	Caucasian Central European colour normal XY single-gene deuteranopes	27	0.26
Schmidt et al., as reported by Sharpe et al., 2000	Caucasian Central European colour normal XY individuals (36 colour normals, 2 single-gene deuteranopes)	38	0.39
Hyashi et al., 2001	Japanese colour normal XY individuals	119	0.21
Carroll et al., 2002	Caucasian American colour normal XY individuals*	50	0.38
Sharpe et al., 2005	Caucasian Central European colour normal XY (35) and XX (3) individuals who are homozygous for L(A180)†	38	0.42
McMahon et al., 2008	African and African American colour normal XY individuals*	27	0.04
Hagen et al., 2019	Norwegian colour normal XY individuals	60	0.55

Table 2.1: Table summarising the L(S180) vs L(A180) allele frequency from various past studies. The first column indicates the study, the second, the observers from whom the fraction L(A180) estimate was derived, the third, the number of observers from whom the L(S180) vs L(A180) allele frequency is estimated (i.e. the number of observers who met the inclusion criteria, see below), and the, fourth the fraction of observers who had L(A180). The studies are listed chronologically in the table. In the two studies labelled with * the observers were not classified as L(S180) vs L(A180), but the peak wavelength sensitivity of their L pigment was reported. Observers who have a peak sensitivity of 559 nm were classified here as L(S180) and observers with a peak sensitivity of 555.5 nm as L(A180). Observers whose peak sensitivity was a wavelength other than 559 nm or 555.5 nm were excluded here (which led to the exclusion of $n = 12$ observers in Carroll et al., 2002, and $n = 0$ observers in McMahon et al., 2008). Sharpe et al., were the only study to contain XX individuals, indicated by †. Two additional XX individuals were included in their study but were heterozygous for L(S180) and L(A180). They are excluded here and only XY individuals and XX who were homozygous are included for the estimation of L(A180) proportion.

2.2.3 Results

The frequency of the L(S180) vs L(A180) allele across the eight ethnic categories from the gnomAD database are shown in Table 2.2. The frequency is reported across the whole population within

Pop.	**Exo.**	**Gen.**	**Total**	**A180**	**XY A180**	**XX A180**	**XX HOM**
Latino/ Admixed American	17,296	424	17,720	0.34	0.33	0.35	0.27
Ashkenazi Jewish	5,040	145	5,185	0.30	0.30	0.29	0.26
European (non-Finnish)	56,885	7,718	64,603	0.26	0.25	0.27	0.24
Other	3,070	544	3,614	0.26	0.25	0.27	0.21
European (Finnish)	10,824	1,738	12,562	0.20	0.18	0.21	0.19
South Asian	15,308	*	15,308	0.11	0.11	0.11	0.16
East Asian	9,197	780	9,977	0.09	0.06	0.09	0.11
African/ African-American	8,128	4,359	12,487	0.05	0.04	0.05	0.05

Table 2.2: Table summarizing the gnomAD *OPN1LW* L(S180) vs L(A180) allele population frequency results. The columns (from left to right) show ethnicity, the number of exome sequences, the number of whole-genome sequences, the aggregate number of exome and genome sequences, the frequency of L(A180) among all individuals in a given ethnicity, the frequency of L(A180) among XY individuals, the frequency of L(A180) among XX individuals, and finally, the percentage of the XX individuals with the L(A180) allele who are homozygous for L(A180). The frequency estimates have been rounded to 2 decimal places. Note, due to the low number ($n = 31$) of South Asian genomes available, the South Asian genome sequences are reported in the Other category in the gnomAD database.

each ethnicity, and for XY and XX individuals separately, along with a report of the proportion of XX individuals who are homozygous for L(A180). It can be seen that within an ethnic group the L(S180) vs L(A180) allele frequency does not vary much between XY and XX individuals. However, there are stark differences in L(S180) vs L(A180) allele frequency between ethnic groups, with 1 in 4 Caucasian observers having the minor allele (L(A180)), compared 1 in 10 East Asian observers, and only 1 in 20 African/African-American observers. All pairwise comparisons between ethnicities are highly significant ($p < 0.001$), except for the difference between the Caucasian and Other groups. The lack of significance here is not too surprising as the Other category contains ethnicities that do not neatly group into one of the seven other categories and are thus likely to comprise largely of individuals from mixed ethnic backgrounds, which, given the majority of the gnomAD contributing datasets come from US laboratories, are likely to be part Caucasian.

Note, there are small but significant differences in the frequency estimates from the genome and exome sequencing results, when the two are considered individually compared to pooled. The overall odds ratio for the frequency of L(A180) in the exome vs genome data is 1.25 (and no odds ratio for a

single ethnicity is greater than 1.26) indicating a universally small effect size between these data types, so here I report the frequency seen in the aggregated genome and exome database, to maximise the amount of data analysed.

2.2.4 Discussion

One can see the estimate for the L(S180) vs L(A180) allele frequency in Caucasian XY individuals is significantly lower than traditional reports from small samples (26% L(A180) from the 64,603 Caucasian individuals in the gnomAD data vs 44% L(A180) from the 309 observer pooled sample estimate (Stockman & Sharpe, 2000), $z = 6.91$, $p < 0.001$, *two-tailed*). The opsin array on the X chromosome is notoriously difficult to sequence with conventional PCR given the high homogeneity between the different X opsin genes. Classical estimates of L(S180) vs L(A180) allele frequency have relied on long-range polymerase chain reaction (PCR) to amplify *OPN1LW* and *OPN1MW*, usually from exon 2 to exon 5, as (Dees et al., 2015). This is quite a different approach to the NGS method, and to date it is unclear whether sequences derived from NGS and long-range PCR corroborate with each other. Despite this uncertainty, there is merit in utilising sequences derived by NGS methods. One major advantage of NGS and the gnomAD dataset is that it provides a unique opportunity to access data from thousands of individuals i.e. much bigger datasets than could ever be sequenced in a sample using long-range PCR by a single laboratory, as has been the alternative approach to date.

Without verification of how well NGS results corroborate with traditional long-range PCR sequencing results for *OPN1LW* it is difficult to know whether this difference is due to inaccuracies in the NGS method, inaccuracies in the long-range PCR method, or the fact that previous estimates relied on samples that were too small and biased to produce accurate estimates. More work is needed to verify the effectiveness of NGS against traditional long-range PCR measures.

There are reasons for both faith and doubt in the NGS method. The coverage, i.e. the number of unique reads that include a given nucleotide, of *OPN1LW* is relatively high (Karczewski et al., 2020). Coverage is a metric commonly used to determine the confidence that the determined nucleotide at a particular location is indeed that nucleotide, and thus the high coverage of *OPN1LW* indicates that one can put faith in the sequence data. However, despite the high coverage of *OPN1LW*, the *OPN1MW* coverage in the gnomAD dataset is poor. This suggests that there are some issues in correctly identifying *OPN1MW*, which might be due to cross-contamination with *OPN1LW*. It is, however, possible that the limited coverage of *OPN1MW* is due to an issue other than the separation of *OPN1LW*. Additionally, given the low coverage of *OPN1MW* and high coverage of *OPN1LW*, one may propose that *OPN1MW*

sequences are being misclassified as *OPN1LW* and pooled into the *OPN1LW* data (it is unlikely that *OPN1LW* are being misclassified as *OPN1MW* as if this were the case then one would expect better coverage for *OPN1MW*). However, if *OPN1MW* sequences were corrupting the *OPN1LW* data then one would expect a *higher* prevalence of alanine at codon 180, and thus a higher frequency of the L(A180) minor allele than in previous sample estimates, yet the results presented here show a *lower* prevalence of the L(A180) allele than expected.

It is also worth noting that classical long-range PCR estimates are not perfect and are also susceptible to L/M cross-contamination. While these methods do seem to work as expected, it is possible that some *OPN1MW* may be amplified by the primer designed to isolate *OPN1LW* and, while this is likely to be a sufficiently small degree of contamination that would be unlikely to be detected by Sanger sequencing methods, there is no guarantee that the classification between L and M genes will be perfect. It is worth pointing out that if there were contamination in the classical long-range PCR estimates, then one might expect to see an overestimate of L(A180) frequency, if M sequences which have alanine at codon 180 are misclassified as L. This is thus one possible explanation for the differences between the NGS results and long-range PCR results, though traditionally the colour vision genetics community has more faith in the accuracy of long-range PCR.

Ideally, one would want a sample of single-gene deuteranopes to derive estimates of L(S180) vs L(A180) allele frequency, to avoid any cross-contamination of M/L. It is interesting to note that in the Caucasian sample containing solely single gene deuteranopes (Sharpe et al., 1998), the L(S180) vs L(A180) allele frequency (26% L(A180)) is not statistically significantly different than the gnomAD estimate (26% L(A180), $z = 0.07$, $p = 0.94$, *two-tailed*); and while not statistically significantly different from the 44% estimate from the 309 Caucasian XY sample in a two-tailed z-test for proportions ($z = 1.81$, $p = 0.07$, *two-tailed*), in a one-tailed test to see if the L(A180) frequency is lower in single-gene deuteranopes, as one would anticipate *a priori* if one believed the L(A180) frequency was being overestimated in colour normals/multi-gene deuteranopes due to cross-contamination from M(A180), the difference is significant ($z = 1.81$, $p = 0.04$, *one-tailed*).

As the purpose of this experiment is primarily to analyse differences in L(S180) vs L(A180) allele frequency across ethnicities, as long as there are not systematic differences in NGS and long-range PCR estimates *across ethnicity* then it doesn't matter if the two estimates agree or not; any differences between the two sequencing methods can be treated as "noise" in the L(S180) vs L(A180) allele frequency estimate. There is no reason that sequencing error should result in systematic biases across ethnicities, and thus the *relative* frequencies of L(A180) across the different groups can be analysed with

confidence. For example, instead of saying that 1 in 4 Caucasian observers have L(A180), compared to 1 in 10 East Asian observers, and 1 in 20 African/African-American observers, we can perhaps have more confidence in saying L(A180) is almost 3 times as prevalent in Caucasian observers compared to East Asian observers, and over 5 times as prevalent than in African/African-American observers (Table 2.2). One potential caveat to this argument is if the number of copies of *OPN1MW* varies across ethnicity, then it is possible that more of the *OPN1LW* sequences get confused with the *OPN1MW* sequences when there are more copies of *OPN1MW*, and thus that L(A180) allele frequency is underestimated in groups where there are more copies of *OPN1MW*. However, there is no evidence to date to suggest the number of copies of *OPN1MW* varies with ethnicity. Such a potential confound may also be a topic of future work in assessing the differences in NGS and long-range PCR.

Interestingly, for the most part, the frequency of L(A180) seems to be higher in ethnicities that historically would have lived closer to the Arctic circle, with L(A180) being a rare allele in groups descendent from equatorial populations (Race, Ethnicity, and Genetics Working Group et al., 2005; Liu et al., 2006). This fits with the hypothesis that L(A180) may offer improved spatial acuity at the expense of some chromatic discrimination and is thus the favoured allele when environmental light levels are low (Hagen et al., 2019). Though, of course this observation is not sufficient evidence for an evolutionary driver in play here.

2.3 Experiment 2: Systematic differences in L:M ratio

2.3.1 Introduction

While it is well known that there are large individual differences in L:M cone ratio, it is less well understood is what systematic differences exist in L:M ratio between different sub-populations. Uncovering such systematic differences may give clues as to what determines the L:M ratio in a given individual. There are plenty of plausible systematic differences one might expect to observe in L:M ratio, based on the underlying genetics and ontogenesis of the retina, as discussed in the Introduction to this Chapter.

Biological Sex

Firstly, given that some 12% of XX individuals are carriers of CVD (Jordan et al., 2010) one may predict more skewed ratios in XX individuals, and a greater variance in L:M ratio among XX individuals than XY individuals. To date, there has been little reported on differences in L:M ratio between XX and XY individuals. Jägle et al., 2006, report no significant difference between L:M ratio between XX and

XY individuals, though their small sample size may not have given enough power to detect such a difference. On the other hand, using much larger samples, Hagen et al., 2019, do find a significantly higher proportion of L cones in XX individuals than XY individuals.

Ethnicity

Perhaps the one relatively well characterized systematic difference in L:M ratio is observed between ethnic groups. Studies that have reported ethnic differences in L:M ratio are summarised in Table 2.3. Hagen et al., 2019 showed a significantly higher L:M ratio in Norwegian XY individuals, than the 2.7:1 estimate of Caucasian XY individuals derived from ERG estimates of L:M ratio (Carroll et al., 2002; Hofer et al., 2005). Additionally, McMahon et al., 2008 showed the 2.7:1 ratio in Caucasian XY individuals was significantly higher than the L:M ratio estimated from their sample of African and African American XY individuals. Further, it has been reported in a few conference abstracts that East Asians have a lower L:M ratio than Caucasians (Kuchenbecker et al., 2014; Yamauchi et al., 2013), though none of these reports have appeared as published articles to the best of my knowledge, nor has the data been made available, and thus there is not sufficient information in the conference abstracts alone to demonstrate this difference is reliable. In the same conference abstract Kuchenbecker et al., 2014, also report a significantly lower L:M ratio in Native Americans than in Caucasians, though again there is not enough evidence in the published literature to support this. While it seems clear there is an association between ethnicity and L:M ratio, there is little or no datum available on L:M ratio estimates in most minority ethnic groups, and no comparisons of L:M ratio between minority ethnic groups. One must also remember to be conscious of the possible confounds in systematic differences across ethnicity in the cone fundamentals used to derive L:M estimates which may contribute to some of the observed variability in L:M ratio across ethnic groups (He et al., 2021; Iannaccone et al., 2007; Howells et al., 2013).

Seasonality

Beyond genetic determinants of L:M ratio, due to the postnatal stage of retinal development (as detailed in Section 1.1.3), it is plausible that early light exposure during development influences the L:M ratio, yet no studies to date, to the best of my knowledge, have investigated such a relationship. As described in the Introduction to this Chapter, the eye must find a balance between spatial acuity and chromatic discrimination, and one way to favour spatial resolution would be to have a highly skewed L:M ratio. Such a trade-off would be preferred for individuals living in low-light environments. Thus I

Study	**Observers**	**n**	**Mean L:M**	**Mean** $\%L$	**L:M/%L range**
Carroll et al., 2002 (unadjusted)	Caucasian American colour normal XY individuals	62	1.86:1	65%L	28-93%L
Hofer et al., 2005	Caucasian American colour normal XY individuals (7) + XX individual (1) who was protan carrier	8	3.70:1	79%L	0.37-16.5:1
Carroll et al., 2002 and Hofer et al., 2005 (ERG correction applied)	Caucasian American colour normal XY individuals and one XX protan carrier	70	2.72:1	73%L	37-95%L
Sharpe et al., 2005	Caucasian British colour normal XY (35) and XX (5) individuals	40	2.26:1	N/A	0.47-15.82:1
McMahon et al., 2008	African and African American colour normal XY individuals	27	1.86:1	65%L	42-85%L
Hagen et al., 2019	Norwegian colour normal XY individuals	60	3.95:1	79.8%L	49.9-100.3%L
Hagen et al., 2019	Norwegian colour normal XX individuals	33	6.14:1	86%L	64.3-99.5%L

Table 2.3: Table summarizing the L:M ratio estimates from a range of past studies. Note, all of the above reports used ERG to estimate L:M ratio, except for Sharpe et al., 2005 who used HFP, and Hofer et al., 2005 who combined ERG estimates with direct measures from retinal densitometry. In ERG estimates it is conventional to report %L cones, where $\%L = L/L+M$. Here, for ease of comparison, %L cones has been converted to L:M ratio as $\%L/(100-\%L)$. Range estimates are given in terms of the method of reporting used in the study in question. Following evidence from Hofer et al., 2005, which suggested ERG estimates of L:M ratio over-estimate the number of M cones, a correction was applied to the ERG estimates derived by Carroll et al., 2002, hence the appearance of the data from Carroll et al., 2002, twice in this table. In the second appearance of Carroll et al., 2002, the data is combined with the 8 observers from Hofer et al., 2005 to give a single mean L:M estimate from ERG recordings of Caucasian American observers as reported by McMahon et al., 2008. It has been reported in a few conference abstracts (Kuchenbecker et al., 2014; Yamauchi et al., 2013) that East Asians have a statistically lower L:M ratio than Caucasian observers, though the corresponding figures are not reported and thus could not be included in this table.

propose that in populations where light exposure changes dramatically across the seasons (Webster et al., 2007), such as the UK, one may find seasonal differences in the development of L:M ratio. That is individuals born in the winter, when the daytime is short, may tend to have higher L:M ratios than individuals born in the summer, who do not need to compromise chromatic discrimination in favour of spatial resolution as daylight is so abundant during the first few months of their development.

The Current Study: *In this experiment I investigated the above three systematic differences in L:M ratio. I began by recruiting a large cohort of observers, without specifically recruiting on the basis of ethnicity, biological sex, and season-of-birth, in order to create a large normative dataset which was going to form the basis of my participant cohort for all of the planned experiments for this thesis. I had then planned to complement this dataset by recruiting specifically from minority ethnic groups who were underrepresented in the original normative dataset. However, due to the global impact of the COVID-19 pandemic (March 2020 - ongoing) I was unable to recruit and test the ethnicity targeted dataset. I was also unable to collect a large enough normative dataset for my subgroup comparisons of biological sex and season-of-birth to be sufficiently, statistically powered. In order to detect a deviation in L:M ratio between any of my sub-groups of the magnitude observed by McMahon et al., 2008, i.e. from 2.7:1 (SD=0.12), to 1.86:1, at a power of 80% and a* 5% *significance level, I would need* $n = 30$ *for each subgroup. None of the subgroup comparisons here have an* $n = 30$ *and thus sufficient statistical power. Here I report the preliminary results of the analysis on the normative dataset that I was able to collect. I plan to use postdoctoral time to collect the remainder of this dataset and complete this experiment. I decided to include this underpowered, preliminary psychophysical experiment in my thesis as it tests novel hypotheses, and demonstrates my ability to design, set-up, and calibrate a Maxwellian-view set-up, and collect human psychophysical data.*

2.3.2 Methods

Heterochromatic flicker photometry (HFP) was used to estimate the L:M ratio in a large cohort of observers (Wagner & Boynton, 1972). HFP is a widely used technique to estimate L:M ratios (Sharpe et al., 2005; Gunter et al., 2006). In HFP, an observer views a stimulus composed of two lights each of fixed wavelength, oscillating in counterphase, in Maxwellian view. One of those two lights is designated the reference light and its irradiance is fixed, while the observer is asked to adjust the irradiance of the second, test, light. For a range of test irradiances, the observer will perceive the stimulus to be flickering, while for a specific range of test irradiances, the observer will perceive the stimulus to be a steady light. The standard model of luminance processing predicts that the point at which the stimulus appears as a steady light will occur when the luminance of the test light is equal to the luminance of the reference light for the individual observer. The aim is to measure this so-called flicker null point for each observer.

Traditionally in HFP, the observer is asked to find a flicker null point for a range of wavelengths of the test light. The individual observer's luminous efficiency function can then be extracted by fitting

this datum, and from this one can estimate the individual observer's L:M cone ratio (Sharpe et al., 2005; Lennie et al., 1993). Here, a variant on this traditional HFP method is used to estimate an individual's L:M ratio in a much shorter time frame to make it possible to collect data on a large sample more efficiently. The observers are asked only to make a setting for one fixed wavelength for the test light. To derive an estimate of L:M ratio, I then model the individual's luminous efficiency function as a linear combination of the standard observer L and M cone spectral sensitivities and find the appropriate weight of L and M to achieve the given null setting (Stockman & Sharpe, 2000; CIE, 2006; Sharpe et al., 2005; Stockman et al., 2008). This method requires only about fifteen minutes per observer to estimate their L:M cone ratio.

Stimuli

The test and reference lights were Winger Electronics WEERD14-CS 5mm red DIP LED (dominant wavelength 625 nm) and Roithner Lasertechnik 5mm LED 545-01 (dominant wavelength 545 nm), respectively. An Intor 630/10/50 bandpass filter and an Intor 545/10/50 bandpass filter are placed in front of the red and green LEDs respectively. To confirm the filters and LEDs behave as expected, the spectral radiance of the test and reference light were measured using a Cambridge Research Systems SpectroCAL MKII Spectroradiometer. The reference light spectrum was confirmed to have a peak wavelength of 545 nm (10 nm FWHM) and the test light, a peak wavelength of 630 nm (10 nm FWHM). The 545 nm filter was chosen for the reference light after modelling showed that this was the wavelength that would give the highest possible power output from the reference light (given the limited buying options of LEDs in the mid-wavelength range) while eliminating the influence of macular pigment density and S cone contributions that occur at lower wavelengths. The LEDs were driven by 10-bit pulse-width modulation (PWM) via an Arduino UNO with an Adafruit 16-Channel 12-bit PWM/ServoDriver shield. The LED output intensity was confirmed to be linear with respect to the driving input value. The waveform of the modulation is sinusoidal with the test and reference lights modulated in counterphase at a frequency of 25 Hz. The reference and test lights were modulated from 0 Td to the maximum measured output retinal illuminance of each light. The maximum of the test light was the variable manipulated by the observer. The observer viewed the combined stimulus through a Maxwellian view system with a 2mm artificial pupil, resulting in a field size of 2° (Fig. 2.4).

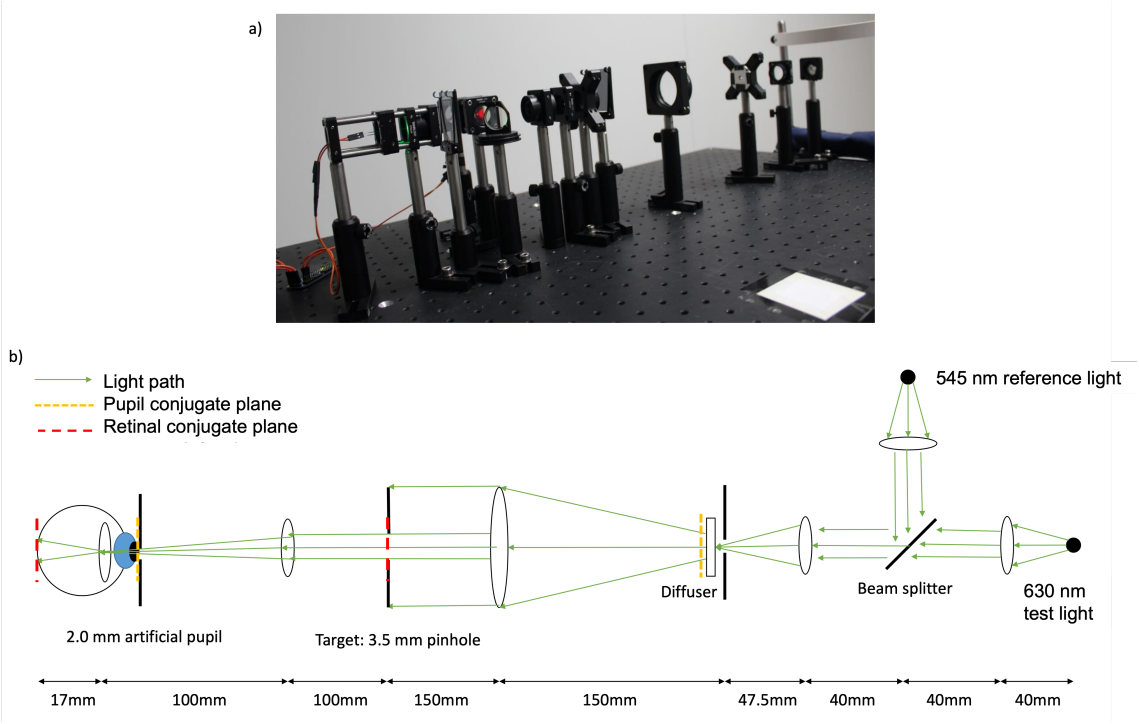


Figure 2.4: HFP set-up. a) Photograph and b) optical diagram of the Maxwellian-view set-up used for the HFP experiment. The optical diagram is not drawn to scale, but the correct distances between components are noted on the diagram

Calibration

The Maxwellian view system was aligned using a Thorlabs TLS001-635 T-Cube Laser Source. To calibrate the Maxwellian-view system used for the HFP set-up the radiant power, ($P(\lambda)$), of the LEDs was measured with a ThorLabs S130C Slim Photodiode Power Sensor in microwatts (μW) and converted to retinal illuminance in trolands (Td) using the method described by Nygaard and Schuchard, 2001. The maximum retinal illuminance of the reference light was measured to be 2.77 $\log_{10}$ Td. The maximum retinal illuminance of the test light, which was varied by the observer, was measured to be 2.39 $\log_{10}$ Td at the minimum possible setting available to the observer up to 2.99 $\log_{10}$ Td at the maximum available setting.

Participants

Participants were recruited from the city of Oxford, through the use of poster adverts, email notices, Facebook, and word of mouth. Ethical approval for this study was granted from the Medical Sciences Inter-Divisional Research Ethics Committee (MS IDREC) under reference number R53478/RE006. Participants who were found during screening to have normal/corrected-to-normal visual acuity of worse than 6/12 or a colour vision deficiency were excluded from further testing (refer to details below).

Participants were also excluded for making implausible settings on the HFP task (i.e. a setting that suggested a lack of L or M cones, suggesting colour vision deficiency). 94 participants, aged 18-35, completed the task. Of those, one subject was excluded on the basis of visual acuity, five were excluded on the basis of colour vision deficiency, four participants were excluded due to implausible HFP settings, and three participants were excluded due to incomplete datasets. The results of the remaining 81 participants are described here.

Trials

At the start of the experimental session, relevant demographic information, including ethnicity, month and year of birth, biological sex, and familial history of colour vision deficiency were first collected with a self-reporting questionnaire that the participants completed. The ethnic groups listed on the questionnaire were the same as those used on the UK census, including an option "Do not wish to say", as well as "Other", with a further option to write-in an answer. Participants then completed basic tests for visual acuity and colour vision. Visual acuity was tested monocularly using a Snellen Eye Chart. The eye with the best acuity was used for the remainder of the experiments. If acuity was the same in both eyes, then the dominant eye, determined using the Miles test, was used. Colour vision was tested using HRR pseudoisochromatic plates (illuminated by a D65 lamp) and the Rayleigh match on an Oculus anomaloscope. For a participant to be classified as being colour deficient, they had to fail both the HRR pseudoisochromatic plates and the Rayleigh match. A failure on the HRR pseudoisochromatic plates was recorded if an individual incorrectly identified one or more symbols on the screening plates and subsequently failed to identify one or more symbol on the diagnostic plates, or on the re-test of the screening plates. A failure on the Rayleigh match was recorded when the observer had a Rayleigh match setting within the range of settings determined to be indicative of CVD.

The participants then completed the HFP task. The HFP task was split into two sessions, with a short break in between. Each session contained two practice trials followed by five actual trials. A short break was taken between sessions to prevent chromatic adaptation to the reference light. The initial maximum retinal illuminance of the test light on each trial was between 2.69 $\log_{10}$ Td and 2.87 $\log_{10}$ Td, selected via a uniformly distributed random number generator. Observers were instructed to use the left and right arrow keys on a keyboard to increase and decrease the intensity of the test light, and change direction only when they noticed the amount of flicker begins to increase. Each observer made nine reversals, with the flicker null point being calculated as the mean of the final four reversals.

Estimating L:M cone ratio

One can analytically extract the L:M cone ratio of a particular observer from their HFP setting. One can calculate the relative L cone sensitivity to the 630 nm test light, L_{630}, and to the 545 nm reference light, L_{545}, and the relative M cone sensitivity to the test light, M_{630}, and the reference light, M_{545}. Here the 2°, rather than the 10°, standard observer L and M spectral sensitivity functions are used to calculate the L and M sensitivity to the test and reference lights as the stimulus had a 2° field size. At the flicker null point for a given observer, the effective luminance of the test light will equal the effective luminance of the reference light. Assuming that each cone contributes with a weighting of one to effective luminance, and n is the ratio of L to M cones for a particular observer, the effective luminance for a given observer, for a light of wavelength λ, is given by Eq. 2.1,

$$E_\lambda = n \cdot L_\lambda + M_\lambda. \tag{2.1}$$

We can therefore analytically solve for n as in Eq. 2.2:

$$n \cdot L_{630} + M_{630} = n \cdot L_{545} + M_{545}$$

$$n = \frac{M_{545} - M_{630}}{L_{630} - L_{545}}. \tag{2.2}$$

2.3.3 Results

HFP Settings and L:M Ratio

HFP settings are reported as $\log_{10}$ of the ratio of the retinal illuminance of the test light to the retinal illuminance of the reference light at the flicker null point. Fig. 2.5(a) shows the mean HFP setting made in the first session compared to the mean HFP setting made in the second session. Participants were self-consistent with their settings on the first and second session, with the settings between the two sessions being strongly positively correlated ($r = 0.90, p < 0.001$). A two-sample paired t-test on the mean setting made during the first session and the mean setting made during the second session supported the null hypothesis, i.e. there was no significant difference between the mean setting made on each session ($t = 0.32, p = 0.75$).

The self-consistency of individual participants on the HFP task demonstrates not only the reliability of the psychophysical method, but also the appropriate choice of the retinal illuminance setting of the test and reference lights. If the test and reference lights are set at too low a retinal illuminance

then there is likelihood of rod interference corrupting the results (Collins et al., 1972; Alexander & Fishman, 1984). However, if the retinal illuminance is set too high then there is a chance that the observer will begin to adapt to the reference light, and the results will be corrupted by chromatic adaptation (Stockman et al., 2008; De Vries, 1948). Given the consistency of observers across sessions, the results presented here are unlikely to be corrupted by chromatic adaptation. At the chosen retinal illuminance level, the scotopic illuminance is 3.07 $\log_{10}$ Td, which is well into the range where rods are considered to be saturated, and thus the results are also unlikely to be corrupted by rod interference (Wyszecki & Stiles, 1982; Wald, 1945; Crawford, 1949; Shapiro et al., 1996; Aguilar & Stiles, 1954).

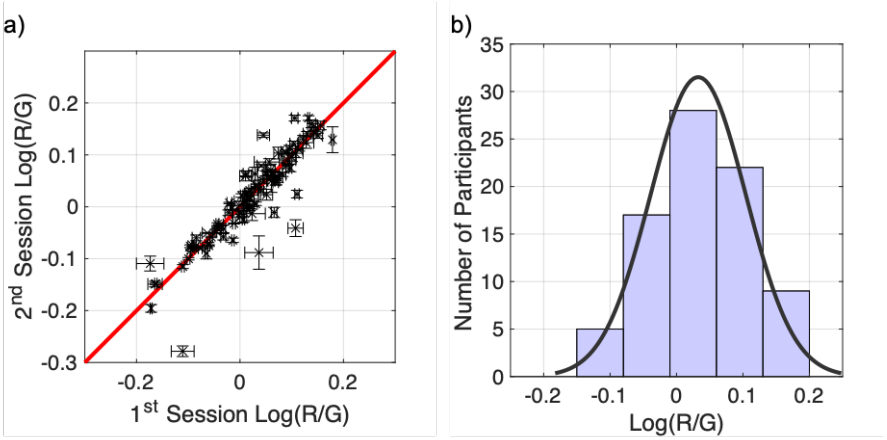


Figure 2.5: HFP settings across sessions and the participant cohort ($n = 81$). a) HFP setting made, reported as log10 of the ratio of the retinal illuminance of the test light to the retinal illuminance of the reference light, on the first HFP session compared to the second HFP session are strongly positively correlated ($r = 0.90$, $p < 0.001$). b) Histogram showing the distribution of HFP settings (pooling the data across both sessions and taking the overall average) made across the sample of observers.

The sample distribution of the mean HFP settings (pooling data from both sessions) across the 81 observers who met the inclusion criteria is shown in Fig. 2.5(b). Fig. 2.6 shows the mapping between HFP settings made and L:M cone ratio estimates. The distribution of L:M cone ratios across the sample of observers, derived from the HFP settings made, is depicted as a box plot in Fig. 2.6. Note that lower HFP settings correspond to a higher L:M ratio. The mean estimated L:M cone ratio of the sample is 2.26:1 ($2.26 \pm 0.196 : 1$). The maximum L:M cone ratio estimate in our sample is 10.8:1 while the minimum estimate is 0.76:1, which is in accordance with the range reported in the literature (Sharpe et al., 2005; Carroll et al., 2002; Hofer et al., 2005). As L:M ratio is a skewed metric by nature, I use HFP settings, which are normally distributed across the sample, for the subsequent sub-group comparisons.

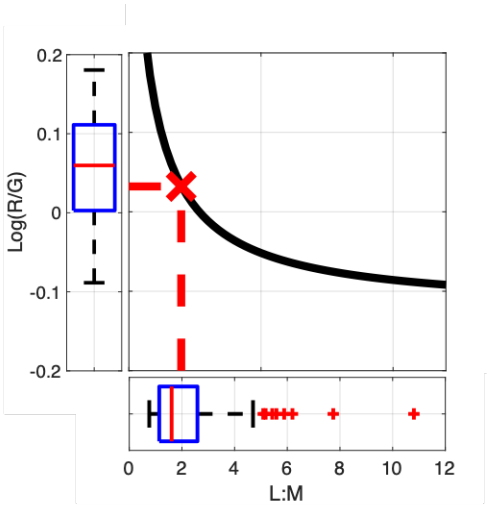


Figure 2.6: L:M ratio estimates from HFP settings. Left: Box plot showing the distribution of HFP settings in our sample. Centre: Mapping of HFP setting made to the estimated L:M cone ratio. The red dashed lines show the setting the standard observer, with an L:M ratio of 1.98:1, would make on the HFP task (Sharpe et al., 2005). Bottom: Box plot showing the distribution of estimated L:M cone ratio in our sample.

Biological Sex

Fig. 2.7 shows boxplots of the HFP settings for XX and XY individuals. A two sample F-test of equal variances showed no significant difference between the variance of HFP settings for XX individuals to XY individuals ($F = 1.11$, $p > 0.99$, $two-tailed$). A two-sample t-test showed no significant difference between the mean HFP setting for XX observers ($n = 51$) to that of XY observers ($n = 24$), ($t = -1.29$, $p = 0.20$, $two-tailed$).

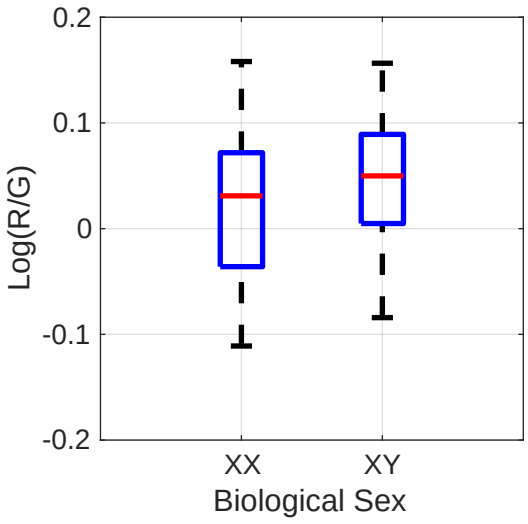


Figure 2.7: Box plots showing the distribution of HFP settings across biological sex (for XY $n = 24$, for XX $n = 51$).

Ethnicity

Of the 81 participants who met the inclusion criteria, 63 identified as Caucasian, 11 identified as Asian (which includes both South and East Asian), none identified as African or Black/Black British, 4 identified as Mixed, one identified as Other, and two declined to answer. Due to the large difference in the sub-group size of Caucasian participants compared to non-Caucasian participants, I plot the HFP settings made by Asian individuals (Fig. 2.8(a)) and Mixed individuals (Fig. 2.8(b)) on a case-by-case basis compared to the distribution of HFP settings made by the Caucasian majority subgroup. On the individual case-by-case basis there does not appear to be strong noticeable differences in L:M ratio between Mixed individuals and Caucasian individuals, with the four Mixed individuals evenly spread throughout the distribution of Caucasian observers. On the other hand 3 of the 11 Asian individuals have HFP settings that fall in the 95^{th} percentile of all settings, and of the remaining 8, 6 have HFP settings above the mean of the whole sample. This hints that Asian individuals may have higher HFP settings on average than Caucasian individuals, and consequently lower L:M ratios.

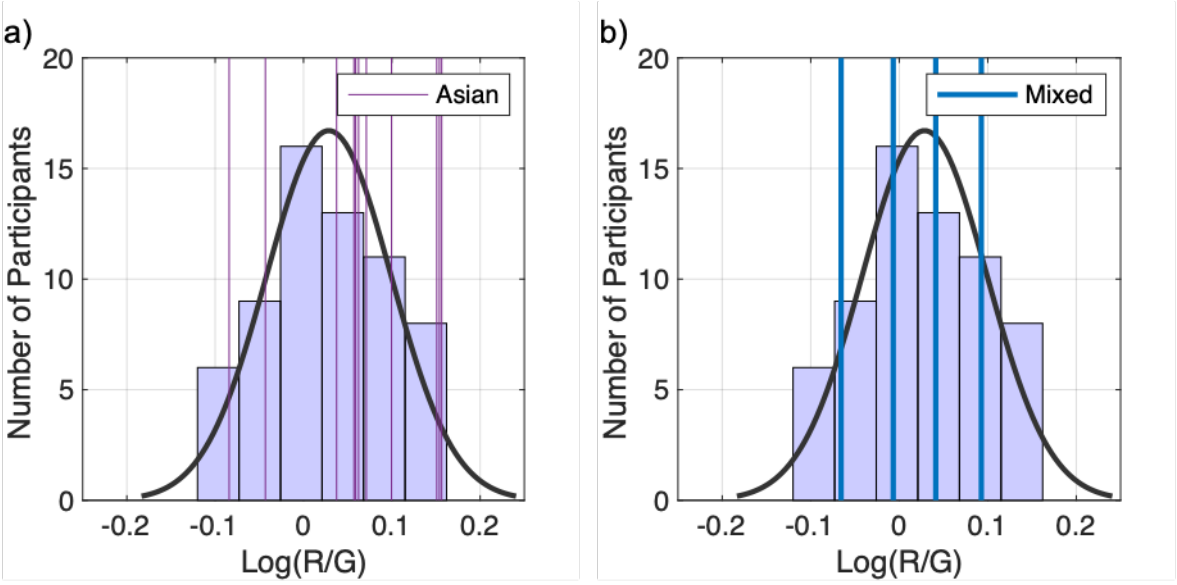


Figure 2.8: HFP settings in Asian ($n = 11$) (a) and Mixed ($n = 4$) (b) individuals, plotted on a case-by-case basis superimposed on a histogram depicting the distribution of HFP settings among Caucasian observers.

Seasonality

Participants reported their month and year of birth, and this was then classified into one of the four seasons using an approximation to the astronomical categorisation of seasons (such that months 1-3 are Winter, months 4-6 Spring, months 7-9 Summer, and months 10-12 Autumn). The astronomical categorisation of the seasons splits the year into groups between the two solstices and equinoxes, and

is thus a better split based on environmental light conditions than the metrological categorisations, which is based on temperature (and which would see each season category shifted forward by a month e.g. months 12-2 as Winter etc). Boxplots depicting the HFP settings made by observers from each season-of-birth are shown in Fig. 2.9. A one-way ANOVA showed no significant differences between the seasons of birth ($F(3,74) = 1.74$, $p = 0.17$). Likewise a two-sample t-test showed no differences when the sample is divided into two groups from spring to autumn equinox i.e. months 4-9 (when days are longest) vs autumn to spring equinox, i.e. months 10-3 (when days are shortest) ($t = -0.12$, $p = 0.91$, $two-tailed$).

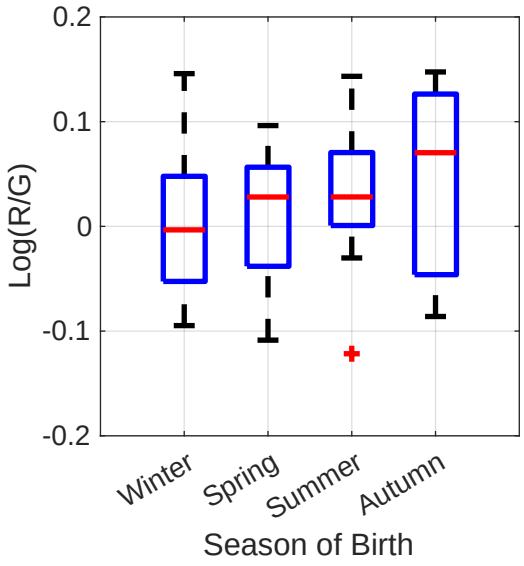


Figure 2.9: Boxplots showing HFP settings between observers with different seasons of birth. Observers are grouped by season-of-birth based on their month of birth, where Winter is January-March, Spring April-June, Summer July-September, and Autumn October-December. This grouping was chosen to best match the periods between the winter solstice and spring equinox, the spring equinox and summer solstice, the summer solstice and autumn equinox, and the autumn equinox and winter solstice respectively. For winter birthdays $n = 21$, for spring $n = 20$, for summer $n = 19$, and for autumn $n = 18$.

2.3.4 Discussion

In summary all systematic differences investigated showed null results. There was no significant difference in HFP settings between XX and XY individuals, nor were there significant differences between observers with different seasons of birth, nor could I observe significant differences between ethnicities. However, these null results do not exclude any of these systematic differences being present, as this experiment as presented here was hugely underpowered to detect such differences. Additionally, within each subgroup comparison presented here I do not control for the other factors i.e. controlling for season-of-birth while testing across biological sex, due to the low sample sizes. One

would ultimately need to control for the other variables within each sub-group analysis.

As noted at the start of this chapter, HFP settings are only a proxy of L:M ratio. While it is possible to estimate L:M ratio from HFP settings (see Eq. 2.2), using the standard observer L and M spectral sensitivity functions to do so, it is well known, that estimates of L:M ratio are very sensitive to which spectral sensitivities are used to estimate the relative sensitivities of the L and M cones to the test and reference lights. L:M ratio estimates are particularly sensitive to variations in the wavelength of peak sensitivity of the L cone spectral sensitivity (Carroll et al., 2000; Bieber et al., 1998). Since a lot of this variability arises from the *OPN1LW* S180A polymorphism, to improve estimates of L:M ratio one would ideally use either the L(S180) or L(A180) spectral sensitivity that corresponds with the individual's genotype to derive an estimate of L:M ratio from HFP score.

This is particularly important given that the L(S180) vs L(A180) allele frequency also varies systematically with ethnicity, meaning that L:M ratios derived from HFP settings from observers who have not been genotyped may be systematically under or overestimated for particular sub-groups. Consequently this would mean that while an observer may appear to have say a low HFP setting, implying a high L:M ratio, they may in fact have an average L:M ratio and an L cone spectral sensitivity that is shifted towards the longer wavelengths which accounts for the increased sensitivity to the red light and the subsequent low HFP setting. After the psychophysical data for the normative cohort reported here was collected, I had planned to re-recruit these observers for genotyping to improve L:M estimates. However, due to delays in getting ethical approval for such data collection and the impact of the COVID-19 pandemic it was impossible to collect the saliva samples from my observers necessary for genotyping, despite developing and testing an appropriate long-range PCR protocol with our collaborator Dr Stephanie Halford. It is worth noting that there are unlikely to be large confounds due to systematic lens density differences (i.e. as driven by age) in this cohort of observers, as the observers' ages were over a limited range, and the effects of lens density are relatively small at the wavelengths of the test and reference light used here. Similarly, there is unlikely to be large confounds due to macular pigment density differences given the choice of wavelength for the test and reference light should be beyond the wavelength region where macular pigment density is a concern.

2.4 General Discussion

This chapter reports empirical tests of systematic differences in the cone photoreceptor signals available to different individuals. Specifically this chapter discusses the ethnic differences in the frequency

of L(S180) vs L(A180) allele and ethnic, biological sex, and season-of-birth differences in L:M ratio. This chapter achieves this by pulling together various reports of systematic differences in the literature and complementing these existing datasets with an analysis of the large-scale gnomAD database and preliminary psychophysical experiments conducted by the author.

Firstly, analysis of existing reports in the literature of L(S180) vs L(A180) allele frequency and of the gnomAD database reveals large ethnic differences in the frequency of the L(S180) vs L(A180) alleles. These results highlight the need for a rethink of the L cone spectral sensitivity estimate. The current standard observer estimate is derived as the average of the L(S180) and L(A180) spectral sensitivities, weighted by the frequency of L(S180) and L(A180) in the population. Yet, the weighting used in the standard observer model is derived solely from estimates of L(S180) and L(A180) frequency in a sample of Caucasian XY individuals, and the results presented here show that the frequency of the L(S180) vs L(A180) alleles significantly differs across ethnicity. Thus, it is necessary to include a consideration of ethnicity in the standard model.

One way to do so, might be to reweigh the L(S180) and L(A180) contributions to the L cone fundamental based on the L(S180) and L(A180) frequency observed in a given ethnic group. As explained in Chapter 1, and outlined in Eq. 2.3, the CMFs and cone fundamentals are a simple linear transform of each other and the standard observer 10° Stockman and Sharpe, 2000 fundamentals are a linear transform of the Stiles and Burch, 1959 10° colour matching functions (CMFs):

$$\mathbf{M}\begin{bmatrix}\overline{r}(\lambda)\\ \overline{g}(\lambda)\\ \overline{b}(\lambda)\end{bmatrix}=\begin{bmatrix}l(\lambda)\\ m(\lambda)\\ s(\lambda)\end{bmatrix},$$

where (2.3)

$$\mathbf{M}=\begin{bmatrix}\overline{l_R} & \overline{l_G} & \overline{l_B}\\ \overline{m_R} & \overline{m_G} & \overline{m_B}\\ \overline{s_R} & \overline{s_G} & \overline{s_B}\end{bmatrix}.$$

Thus to derive the transformation between the CMFs and the cone spectral sensitivities one must simply solve for the matrix **M**. To do so, Stockman and Sharpe, 2000 used spectral sensitivity data of dichromats of a known genotype to determine these unknowns. As the spectral sensitivity measurements of the dichromats are made at a field size of 2°, an intermediary step in the derivation of the 10° standard observer cone fundamentals is the derivation of the matrix required to linearly transform the 2° **StilesBurch1955** CMFs to match the spectral sensitivity measurements of the dichromats (as

explained in detail in Chapter 1). As the S cone sensitivity to the **R** primary is minimal, $\overline{s_R}$ can be set to zero, one is left to derive the unknowns of matrix **M**, as in Eq. 2.4,

$$\begin{bmatrix} \overline{l_R}/\overline{l_B} & \overline{l_G}/\overline{l_B} & 1 \\ \overline{m_R}/\overline{m_B} & \overline{m_G}/\overline{m_B} & 1 \\ 0 & \overline{s_G}/\overline{s_B} & 1 \end{bmatrix} \begin{bmatrix} \overline{r}(\lambda) \\ \overline{g}(\lambda) \\ \overline{b}(\lambda) \end{bmatrix} = \begin{bmatrix} \overline{l}(\lambda) \\ \overline{m}(\lambda) \\ \overline{s}(\lambda) \end{bmatrix}. \tag{2.4}$$

It is at this intermediary step that the weighted average of the L(S180A) polymorphism influences the cone fundamentals. The $\overline{l_R}/\overline{l_B}$ and $\overline{l_G}/\overline{l_B}$ parameters are derived as the weighted average of the L(S180) and L(A180) specific $\overline{l_R}/\overline{l_B}$ and $\overline{l_G}/\overline{l_B}$; where for L(S180), $\overline{l_R}/\overline{l_B} = 5.28554$ and $\overline{l_G}/\overline{l_B} = 16.80098$, while for L(A180), $\overline{l_R}/\overline{l_B} = 4.15278$ and $\overline{l_G}/\overline{l_B} = 16.75822$ (Stockman and Sharpe, 2000). Thus one would simply need to reweigh the polymorphism frequencies for a given ethnic group to re-derive the cone fundamentals with an account of ethnicity. If one wanted to adopt a single means-based L cone spectral sensitivity function, it would perhaps be more appropriate to weight the ethnic specific fundamentals based on the demographics of the overall population.

However, at the heart of the issue, is that the Stiles and Burch, 1959, colour-matching functions (CMFs) from which the cone fundamentals are derived, are likely from only Caucasian XX and XY observers (though ethnicity of observers is not explicitly reported by Stiles and Burch, 1959). It is likely that the CMFs themselves therefore vary across ethnicity, as we already know of systematic variation in the underlying parameters e.g. macular pigment density (Iannaccone et al., 2007; Howells et al., 2013). The ethnic differences seen in CVD (Delpero et al., 2005; Birch, 2012), also suggest further systematic (and potentially correlated) differences in the molecular genetics of colour vision, as yet discussed. There is therefore no simple fix to account for this ethnicity bias. To date there are no colour-matching data available for non-Caucasian populations.

The cone fundamentals of (Stockman & Sharpe, 2000) have been adopted as the standard spectral sensitivity estimates by the Commission International de l'Éclairage (CIE) ((CIE, 2006)), meaning these functions are subsequently used by display manufacturers, architects, and lightning designers to inform ergonomic colour design. It is therefore important that an accurate estimate of the cone spectral sensitivity functions is obtained not just to inform vision science, but also because such estimates feed into a multitude of industrial and commercial applications. The results presented here call for the need to re-evaluate which cone fundamentals are adopted by industry, with an open-eyed approach to the systematic biases that are introduced when ethnicity is ignored. Ultimately, this will likely entail a move away from means-based accounts towards personalised fundamentals.

It is worth highlighting the fact that by taking a weighting of the L(S180) and L(A180) spectral sensitivities to derive a single-means based function the result is an L cone spectral sensitivity that is not perfectly appropriate for any individual (except, perhaps, XX heterozygotes who express both L(A180) and L(S180)). Another option, therefore, would be to simply use the L(S180) spectral sensitivity as this is the predominant allele observed in the population. By using just the L(S180) spectral sensitivity, one would therefore be using the correct spectral sensitivity estimate for an individual the majority of the time. Ultimately, to be sure which spectral sensitivity function one should use for a given individual, one really must genotype the individual.

Such genotyping is particularly prudent to improve estimates of L:M ratio derived from HFP, as in the second experiment presented in this chapter. In this experiment, I observed no significant difference in HFP settings on the basis of biological sex, ethnicity, or season-of-birth. The lack of a significant difference in HFP setting with biological sex is somewhat at odds with recent reports in the literature (Hagen et al., 2019), though there is no clear consensus in the literature on whether or how L:M ratio is associated with biological sex (Jägle et al., 2006). As my observers were not recruited on the basis of ethnicity, my sample was heavily biased towards Caucasian observers, and thus I was unable to make statistical comparisons of HFP settings across ethnicity. On an individual case-by-case basis, the HFP settings of Asian observers did tend to fall above the average HFP setting in the Caucasian sample, hinting at a difference in L:M ratio between Asians and Caucasians. To the best of my knowledge, the results presented here are the first to consider a relationship between environmental light exposure during development (as measured by season-of-birth) and L:M ratio. While I did not find a significant association between season-of-birth and HFP setting in my sample, it is possible that such an effect was drowned out by uncertainty in L:M ratio contribution to HFP settings (vs individual variability in cone spectral sensitivities) and the lack of control for other variability i.e. biological sex and ethnicity between individuals. It is also possible that due to large artificial light exposure in the modern world, any impacts of seasonal variations in natural light exposure are masked and thus do not drive differences in L:M ratio, even if there exists a mechanism by which light exposure could do so. With more observers, a more appropriate analysis here would have been fitting a Fourier series as a function of month of birth, rather than pooling across months into seasons. At the heart of the issue is that the results presented here are not from a large enough cohort of observers and thus the statistical analyses are underpowered, due to the disruption of the COVID-19 pandemic on participant recruitment and testing.

I plan to continue the work presented in this chapter during postdoctoral work. In my future work,

I will genotype observers through both NGS and long-range PCR methods as a means of 1) assessing the validity of NGS methods for sequencing *OPN1LW*, and thus the utility of the gnomAD database for assessing the frequency of the L(S180) vs L(A180) alleles in the Caucasian population, and 2) to use individualised estimates of cone spectral sensitivities to estimate L:M ratio from HFP. I also plan for future work to test the additional hypothesis regarding genetic control of L:M ratio by investigating a possible candidate factor for controlling the ratio: RXRγ (Lawrance-Owen et al., 2014). In particular, Lawrance-Owen et al., 2014, found the SNP rs16844995 which lies in the sequence upstream of *RXRG*, accounted for 2.6% of the variance in OSCAR test settings, where individuals homozygous for the major allele have higher average OSCAR settings than those homozygous for the minor allele, with heterozygotes somewhere in between, such that the minor allele corresponds to a shift towards lower L:M ratio. Interestingly, in the gnomAD database, the expected frequency of the minor allele in the Caucasian population is 0.19, which is very similar to the 0.21 estimate derived by Lawrance-Owen et al., 2014. Additionally, in the gnomAD dataset the frequency of the minor allele is increased in East Asians compared to Caucasians, with a frequency of 0.26, which mirrors the observed direction of L:M ratio differences between the two ethnic groups. There are, however, not large differences in the allele frequency among other ethnic groups. Taken together, *RXRG* and the upstream introns are clear candidates for a role in the L:M ratio that require further investigation.

2.5 Conclusion

This chapter presents results on the systematic differences in L(S180) vs L(A180) allele frequency and L:M cone ratio. I find evidence for ethnicity differences in the frequency of the L(S180) vs L(A180) alleles from the large-scale gnomAD database. I do not find significant differences in HFP settings as a proxy of L:M ratio across biological sex, ethnicity, or season-of-birth, though this is likely due to the low statistical power of the study. While I do not find evidence for such systematic differences, previous literature suggests that there are sex and ethnic differences in L:M ratio (Hagen et al., 2019). As suggested by Hagen et al., 2019, such ethnic differences in L:M ratio and L(S180) vs L(A180) allele frequency point to a possible evolutionary drift towards an eye that favours spatial resolution over chromatic discrimination in populations that ancestrally would have lived in relatively low-light conditions for long periods of the year (Race, Ethnicity, and Genetics Working Group et al., 2005; Liu et al., 2006). It remains unclear whether there is an impact of an individual's early light exposure during development on L:M ratio, or whether genetics alone (and if so, what genetic mechanisms), are

responsible for the observed variability.

It is worth highlighting that the differences presented in this chapter focus only on two known sources of variability between individuals, namely L(S180A) polymorphism and L:M ratio. Variability between individuals' cone fundamentals and luminous efficiency is by no means limited to variability in these two physiological factors, and systematic differences have been demonstrated previously in other physiological factors e.g. macular pigment density (Hammond Jr et al., 1996; Ciulla et al., 2001; Bone et al., 2003; Iannaccone et al., 2007; Howells et al., 2013). To fully utilise individual differences both as a tool for vision science and ergonomic design and as a means to understand driving factors underlying the variability between individuals, one would ideally want a full characterisation of all of the physiological factors that influence the cone photoreceptor signals. Collecting such a dataset, where one knows the CMFs, macular pigment density, lens density, genetics, L:M ratio, and demographics of a large cohort of individuals is an impossibly large task, and thus the approach taken here is strategic to try and pin down the variability in two of the many relevant parameters.

2.6 Acknowledgements

The work presented in this chapter on L(S180) vs L(A180) allele frequency across ethnicity is part of a collaborative project, led by ACH, with Dr Takuma Morimoto (University of Oxford and Justus Liebig Universität Giessen), Dr Michelle McClements (University of Oxford), and Prof. Hannah Smithson (University of Oxford). I would particularly like to thank Michelle for her expertise on cone opsin genetics. I would also like to thank Dr Stephanie Halford (University of Oxford) who was also part of this collaborative project before her retirement (September 2021) and was particularly instrumental in developing the protocol for genotyping L(S180A). The work on L:M ratios is part of a collaborative project, led by ACH, with Dr Sarah Regan (University of Oxford, DPhil, 2015-2020), Dr Laura Young (Newcastle University), and Prof. Hannah Smithson. Data collection for the HFP experiment was split between ACH, Dr Sarah Regan, and Miss Rebecca Williams. The analysis of the data and the write-up of the results presented in this chapter was entirely the work of ACH.

Chapter 3

Exploring the statistical regularities in photoreceptor signals from surfaces for illuminant estimation with machine learning

Contents

3.1 Introduction 104

3.1.1 The computational challenge of illuminant estimation 105

3.1.2 Illuminant perception thresholds 106

3.1.3 Illuminant perception and colour constancy 107

3.1.4 Illuminant perception as an environmental cue 109

3.1.5 Illuminant estimation for computer vision 109

3.1.6 Cues to the illuminant 110

3.2 Experiment 1: Base Case 111

3.2.1 Methods 112

3.2.2 Results 118

3.2.3 Discussion 118

3.3 Experiment 2: Varying Illuminant Dataset 119

3.3.1 Methods 120

3.3.2 Results 121

3.3.3 Discussion . . . 123
3.4 Experiment 3: Comparing surface statistics . . . 123
3.4.1 Methods . . . 123
3.4.2 Results . . . 124
3.4.3 Discussion . . . 125
3.5 Experiment 4: Strategies for combining multiple surfaces . . . 127
3.5.1 Methods . . . 127
3.5.2 Results . . . 128
3.5.3 Discussion . . . 129
3.6 Experiment 5: Availability of photoreceptor signals . . . 130
3.6.1 Methods . . . 130
3.6.2 Results . . . 130
3.6.3 Discussion . . . 131
3.7 General Discussion . . . 132
3.7.1 Connection to human illuminant perception . . . 132
3.7.2 Limitations . . . 133
3.7.3 Further directions with the machine learning approach . . . 134
3.8 Conclusion . . . 135
3.9 Acknowledgements . . . 135

3.1 Introduction

One of the many higher-level computational challenges the human visual system faces is illuminant perception. The light in our external world is constantly changing. In the modern world we are always moving between artificially lit indoor environments, to outdoor settings. Natural daylight shows a notable change in chromaticity and intensity over the course of a day, and with changes in the weather (Judd et al., 1964; Spitschan et al., 2016). In addition to these daily changes in lighting driven by our movements and the changing world around us, there are noticeable longer-scale seasonal changes in natural light (Webster et al., 2007), and consequently exposure, to artificial light, that we experience. Further, the illumination of a scene is inherently tied to our perception of the scene itself, causing shadows and reflections which can provide useful cues to shape and heading (Nascimento et al., 2016). Thus the perception of the illumination in our external world can provide information about our environment that the visual system may exploit.

3.1.1 The computational challenge of illuminant estimation

Illuminant estimation, however, is not computationally simple for our visual system. The vast majority of light we sense from real-world scenes and environments comes from a complex set of illuminants reflecting from the surfaces of objects that make up the scene. Thus, in the simplest case, the light reaching our eyes at any given instant is the resulting radiance spectrum at each spatial location, $E(\lambda)$, that arises from the combination of the spectral reflectance of the object surfaces, $R(\lambda)$, and the illuminant spectra of the light sources in the scene, $I(\lambda)$, as in Eq. 3.1,

$$E(\lambda) = R(\lambda) \cdot I(\lambda). \tag{3.1}$$

Under photopic conditions, this radiance spectrum (at a particular spatial location) is then down-sampled through the cone photoreceptors in the retina. Each cone type (for the normal trichromatic human observer, S, M, L) has a certain spectral sensitivity function, ($T_S(\lambda)$, $T_M(\lambda)$, $T_L(\lambda)$), with which it samples the external radiance spectrum, $E(\lambda)$, such that the resultant cone photoreceptor signals, A_S, A_M, A_L, are determined by the integral of the product of the spectral sensitivity functions with the radiance spectrum, as in Eq. 3.2,

$$\begin{aligned} A_S &= \int T_S(\lambda)E(\lambda)d\lambda, \\ A_M &= \int T_M(\lambda)E(\lambda)d\lambda, \\ A_L &= \int T_L(\lambda)E(\lambda)d\lambda. \end{aligned} \tag{3.2}$$

A schematic of the computational challenge described above is depicted in Fig. 3.1. Our colour percept arises from the comparison between these cone photoreceptor activations in post-receptoral pathways. Crucially, this means that our visual system does not have access to the full spectral information about the incoming radiance spectrum. Not only is there no way to undo the sampling of the radiance spectrum by the cone photoreceptors to recover the full spectral content of the light entering the eye, it is also impossible to undo the multiplication that resulted in that radiance spectrum in the first place: that is to say to separate the contributions of the illuminant and the surface reflectance to the radiance spectrum. Therefore, to estimate the illuminant, the observer must either have explicit access to the illuminant in the visual scene, which is rarely available, or the observer must somehow undo this multiplication to achieve an estimate of the illuminant (Maloney & Yang, 2003). This is a mathematically impossible and underdetermined task.

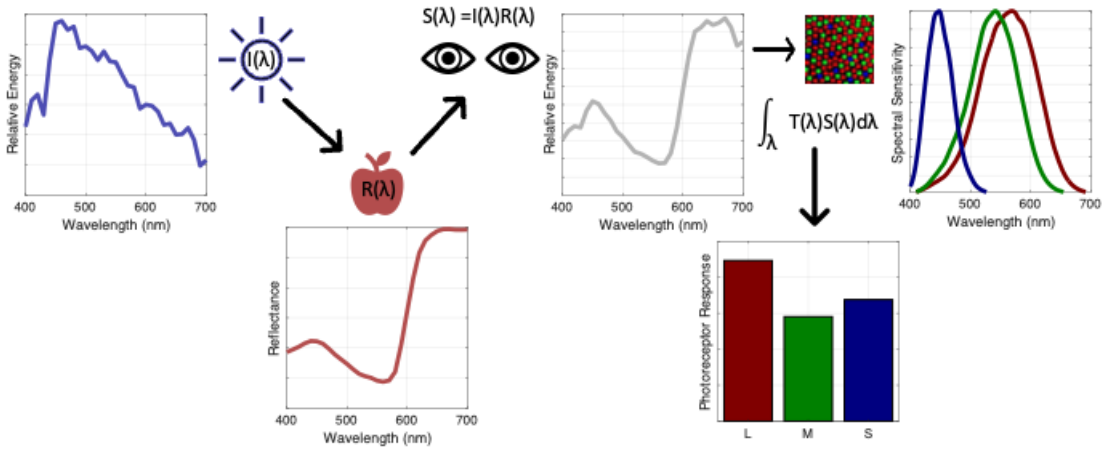


Figure 3.1: A schematic depicting the computational problem of illuminant estimation. This schematic demonstrates the simplest case: where a single, flat matte surface with a given surface reflectance, $R(\lambda)$, is illuminated by a single light source with a given illuminant spectrum, $I(\lambda)$. The radiance spectrum that enters the eye is the product of the illuminant spectrum and surface reflectance, $I(\lambda)R(\lambda)$. This radiance spectrum is then downsampled by the three cone photoreceptor types in the human trichromatic retina. Each cone type has a certain spectral sensitivity with which it samples the radiance spectrum, reducing the full spectral content down to the cone photoreceptor quantal catch. These photoreceptor signals represent the sole proximal sources of information to which downstream brain areas have access. Reliable illuminant estimation must overcome this projection into a lower-dimensional space and multiplication to separate the chromaticity of the illuminant from the chromaticity of the surface reflectance.

3.1.2 Illuminant perception thresholds

Despite the fact that the computational problem is under-constrained and ill-posed, our visual systems are, in practice, able to undo the effect of downsampling and multiplication in order to ascribe a separate perception of the chromaticity of surfaces in a scene and the chromaticity of the illumination (Craven & Foster, 1992; Zaidi, 1998). Illuminant perception, in and of itself, has not received as much attention in the literature as the closely related problem of colour constancy (see Section 5.1.3). However, early work demonstrated that observers are able to consistently judge the illuminant in a scene, demonstrating that our visual system is able to extract an explicit percept of the illuminant (Katz, 1936; Beck, 1959, 1961).

There is still a lot that is not understood about illuminant perception. The problem of illuminant perception is beginning to gain traction in the literature, with recent work characterizing thresholds for illuminant discrimination directly (Pearce et al., 2014; Radonjić et al., 2016; Álvaro et al., 2017; Radonjić et al., 2018; Aston et al., 2019). Not only does this work provide further evidence of the visual system's ability to detect changes in illuminant chromaticity, it is the first attempt to quantify the limits of such

perception directly.

Thus far, discrimination thresholds have been measured along four different chromatic directions from a neutral white and suggest a reduced sensitivity along the blue illumination direction, which is referred to as a "blue bias" (Pearce et al., 2014; Radonjić et al., 2016; Radonjić et al., 2018). Additionally, there appears to be a "neutral bias", such that sensitivity is higher when the illuminant moves towards a more chromatically biased illuminant (Aston et al., 2019). Interestingly, it has also been shown that there is no significant difference in illuminant discrimination thresholds between red-green dichromats and normal trichromats along the daylight locus, which one may have predicted due to the reduced cone signals available to dichromats compared to normal trichromats (Álvaro et al., 2017).

Taken together, these results suggest the visual system encodes priors for natural illumination conditions and likely changes in natural illumination. It is less clear how illumination discrimination depends on the make-up of surfaces in the scene, as there has been conflicting evidence suggesting both that surface content matters for illuminant discrimination (Radonjić et al., 2016; Radonjić et al., 2018), and that illuminant discrimination is independent of scene content (Álvaro et al., 2017). The idea that illuminant perception depends on the existence of priors is consistent with a Bayesian framework, which has been widely discussed as a model for colour constancy (Brainard & Freeman, 1997). Whether or not such a Bayesian model holds for illuminant perception still lacks direct evidence, and importantly the statistical regularities in the real-world environments that an observer may use to inform the construction of any priors are not well characterized.

3.1.3 Illuminant perception and colour constancy

Often, the study of illuminant perception is entwined with the study of colour constancy. Colour constancy is the ability of the human visual system to ascribe a constant percept of colour to an object, undoing the influence of the spectral bias introduced by the illuminant chromaticity (Smithson, 2005; Foster, 2011; Hurlbert, 2007). It is easy to see how illuminant perception and colour constancy get conflated, as if one is able to perfectly estimate the illuminant, then one should be able to perfectly disentangle the illuminant from the surface chromaticity, and thus achieve perfect colour constancy (Aston et al., 2019; Maloney & Wandell, 1986).

Colour constancy has received a lot of attention among colour vision scientists as it is a particularly useful attribute for the human visual system, primarily because colour is a powerful cue to object recognition. For example, colour can help us to identify the ripe fruit amongst the surrounding foliage. If our percept of surface colour changed with illumination conditions then we would not be able to

use colour to inform object identification, as the surface colours of objects in the world appear to change throughout the course of the day as the sky turns from a warm orange at sunrise, through bright blues (or greys) during the day, through to deep blue at twilight. Yet, we know from both the wealth of literature on colour constancy (Smithson, 2005; Foster, 2011) and from our own personal experience, that our colour percepts of objects remain fairly constant. The question of how our visual system achieves colour constancy has captivated vision scientists for decades. In fact, the phenomenon of colour constancy was first discussed over three centuries ago (Mollon, 2003). Yet, the mechanisms that drive colour constancy are still not well understood.

Baked in to almost every model of colour constancy is some form of illuminant estimation, either implicit or explicit. That is to say somehow the visual system must account for the changes in the illuminant chromaticity in order to achieve constancy (Helmholtz, 1866). It may be the case that the observer is explicitly judging the illuminant chromaticity in order to achieve this. Indeed such an approach was first suggested by Ives, 1912, in perhaps the first explicit model of colour constancy (Brill, 1995). Ives showed that if one knew the 3D colour coordinates of the illuminant, then one could scale the 3D colour coordinates from the surfaces in a scene via multiplicative scaling in order to achieve constancy (von Kries, 1905). The fact that such multiplicative scaling is able to undo the effect of the illuminant is not a theoretically provable phenomenon, rather it relies on statistical regularities in the set of surfaces and illuminants likely to be encountered. In other words, the use of a multiplicative scaling to undo the impact of the illuminant, relies on a spectral change in the illuminant causing multiplicative scaling of the cone signals from all of the surfaces in the scene in the first place. While this is not the case for any arbitrary set of surfaces, and arbitrary illuminant change, it has been shown the be approximately true for a large sample of surfaces and illuminants found in the real-world (Foster & Nascimento, 1994).

It is worth noting that one does not need an explicit estimate of the illuminant in order to achieve constancy, one just needs to know the gain factor to apply to the cone signals to adjust for illuminant changes (Smithson, 2005). In fact, it has been empirically shown that an explicit estimate of the illuminant does not correlate with colour constancy (Rutherford & Brainard, 2002; Granzier et al., 2009; Weiss et al., 2017). Despite such evidence, there is still keen interest in understanding illuminant perception as a possible mechanism that may contribute to colour constancy.

3.1.4 Illuminant perception as an environmental cue

It is often stated that illuminant perception is an interesting phenomenon as an estimate of the illuminant may help inform judgements of time of day, season, and weather (Radonjić et al., 2016; Radonjić et al., 2018); since the illuminant spectrum is known to vary over the course of the day and with weather (Judd et al., 1964; Spitschan et al., 2016). While there is a growing body of evidence that human observers can perceive the colour of the illuminant, there is little direct evidence as to whether that illuminant estimate is useful in providing such environmental cues. In fact, the limited evidence available actually suggests that observers are quite poor at using such illuminant estimates in distinguishing time of day, apart from distinguishing between sunset, sunrise, and midday (Granzier & Valsecchi, 2014). Some observers do seem to be able to use the illuminant estimate to distinguish seasonality and temperature, though it was not tested how well these observers were able to explicitly perceive the illuminant, and thus it may be the case that the unreliable estimates of time of day were due to failures in illuminant perception, rather than failures of applying this percept to extract relevant information about the environment (Granzier & Valsecchi, 2014). The limited evidence on the ability of the human observer to appropriately perceive and exploit the perception of the illuminant in a scene shows the clear demand for more direct tests of illuminant perception.

3.1.5 Illuminant estimation for computer vision

Beyond human vision, illuminant estimation is also a crucial stage of image processing for computer vision applications. There is great interest in developing a model of computational colour constancy. As we know from human vision, colour is crucial for object recognition, and object recognition is one of the key challenges facing computer vision (Dickinson, 1999; Gijsenij et al., 2011). For colour to be useful for object recognition in computer vision, the colour of the objects in the scene must be separated from the colour of the illuminant, and thus one must feed in a specific algorithm to undo the influence of the illuminant into a model of computational colour constancy. Just as with human colour constancy, there have been two broad algorithms proposed by which to achieve this, one requiring an explicit estimate of the illuminant (Finlayson et al., 2001; Hordley, 2006), and the other relying instead on the relative changes in the surface chromaticities between two scenes (Funt & Finlayson, 1995). The additional appeal of achieving an explicit estimate of the illuminant in computer vision is that one can back-render the scene, and then re-render the scene under any given illuminant (Einabadi et al., 2021).

3.1.6 Cues to the illuminant

While it is unclear how an explicit estimate of the illuminant colour may be used by the human observer, the general consensus in the literature is that observers are able to perceive the illuminant colour. How then, from a set of surfaces in a scene can we estimate the illuminant?

There have been many ideas as to how such an estimate of the illuminant may be achieved, most of which come from the colour constancy literature. One major hypothesis is that the brightest surface in the scene provides a cue to the illuminant, assuming the brightest surface is white and thus has a perfectly flat surface reflectance, thus faithfully reflecting the illuminant spectrum (Land, 1977). Another hypothesis, typically referred to as the "grey-world hypothesis", (though perhaps more accurately referred to as the global mean), is that the mean cone signals from the visual scene provide an estimate of the cone signals of the illuminant (Buchsbaum, 1980). There has also been evidence to suggest a temporal mean, based on prior history of illuminant exposure contributes to constancy (Smithson & Zaidi, 2004). It is likely that all of these cues may be used to some extent, with different cues given different weights under different settings (Kraft & Brainard, 1999).

Statistical models of illuminant estimation have also gained traction. The idea behind statistical models is that the gamut of colours available in a scene is constrained by the illuminant (Forsyth, 1990; Finlayson, 1996). The limited gamut available under a particular illuminant is called the canonical gamut. That is to say, if one were to see a very red surface, then it would be impossible for that surface to be under a strong blue illuminant, or else one would not have such a strong red percept. Such statistical models are appealing as they exploit the statistical regularities found in real-world surfaces and illuminants, and put constraints on the underlying unconstrained mathematical problem. Further, there is evidence that our visual system utilises such canonical gamuts to achieve colour constancy (Uchikawa et al., 2012; Morimoto et al., 2021).

In order for such a statistical model to work, there must be statistical regularities in the external world that can be learnt and then exploited by the observer. While the statistical regularities of the real-world are often quoted as fact, they are poorly understood. This is in large part due to our limited understanding of the "spectral diet" of human observers (as described further in the General Discussion in this chapter and Chapter 6 (Webler et al., 2019)). It is further unclear what statistical properties may be useful to achieve the task of illuminant estimation i.e. what features of a scene are most useful for illuminant estimation, and what is the optimal strategy for pooling information from multiple surfaces in a scene?

Machine learning algorithms offer a novel approach to investigate the statistical regularities in

real-world scenes. Indeed, illuminant estimation via convolutional neural networks (CNNs) has gained recent popularity. Most current methods are motivated by computational colour constancy, and thus aim to maximise the performance of such algorithms (Choi et al., 2020). Thus while such CNNs achieve extremely good performance, they are complex and opaque. There has also been some effort in applying such CNNs to evaluate human colour constancy (Flachot & Gegenfurtner, 2018; Flachot et al., 2019).

In this chapter, I employ a more transparent machine learning approach to uncover the statistical regularities in real-world scenes and illuminants that may inform human illuminant perception. The novelty of this work compared to existing applications of machine learning to the problem of illuminant estimation is two-fold. Firstly, while past networks have been trained using the RGB coordinates of the input image (Tominaga et al., 2001), I train my algorithm using the same signals that are available to the human observer: the human photoreceptor signals. This means that my algorithm is only able to learn about the statistical regularities that are available to the human observer. Secondly, past work has focused on training algorithms using complex 2D images, where the algorithm may learn to estimate the illuminant over a range of useful image features (Choi et al., 2020; Flachot & Gegenfurtner, 2018; Flachot et al., 2019). This makes it more challenging to compare feature importance, and it is unclear whether such complexity is necessary for illuminant estimation and constancy. Therefore, I take a bottom-up approach instead, and use a simple combination of single illuminants and single surface reflectances to probe the limits that can be placed on illuminant estimation from only the statistical regularities in real-world radiance spectra as sampled by the photoreceptor signals. This approach thus allows me to evaluate the properties of the surface statistics, illuminant classes, and photoreceptor signals that impact performance of the algorithms to predict illuminant chromaticity. The approach also allows me to quantify the advantage gained from combining signals from multiple surfaces, and the optimal strategy of combination.

3.2 Experiment 1: Base Case

I start with what I will here after call the *base case* to firstly verify that a simple machine learning classification algorithm can learn to estimate the illuminant from a single surface under a single illuminant using just the cone photoreceptor signals. The dataset, algorithm, and metrics described here will be the basis for all subsequent experiments described in this chapter, and in subsequent sections I will describe the differences between the base case described below that are relevant for

each experimental manipulation.

3.2.1 Methods

Illuminant dataset

I start by taking illuminants from the daylight locus (Judd et al., 1964). The colour temperature (CCT) of the illuminants was chosen such that each illuminant spectrum was sampled from along the daylight locus with a colour difference of $5\Delta E$ separating each illuminant from its nearest neighbours. The central illuminant spectrum was chosen to be the D65 standard. This resulted in a series of eight illuminants, with CCTs of 4195K, 4780K, 5365K, 5945K, 6500K, 7170K, 8230K, 10185K. The spectra of the eight illuminants along with their chromaticities as depicted in CIE u'v' space are shown in Fig. 3.2(a) & (b). The intensity of the illuminants was randomized so that the classifier could not learn to associate one particularly strong cone signal with the illuminant chromaticity; the classifier specifically had to pick up on the patterns present in the relative cone signals.

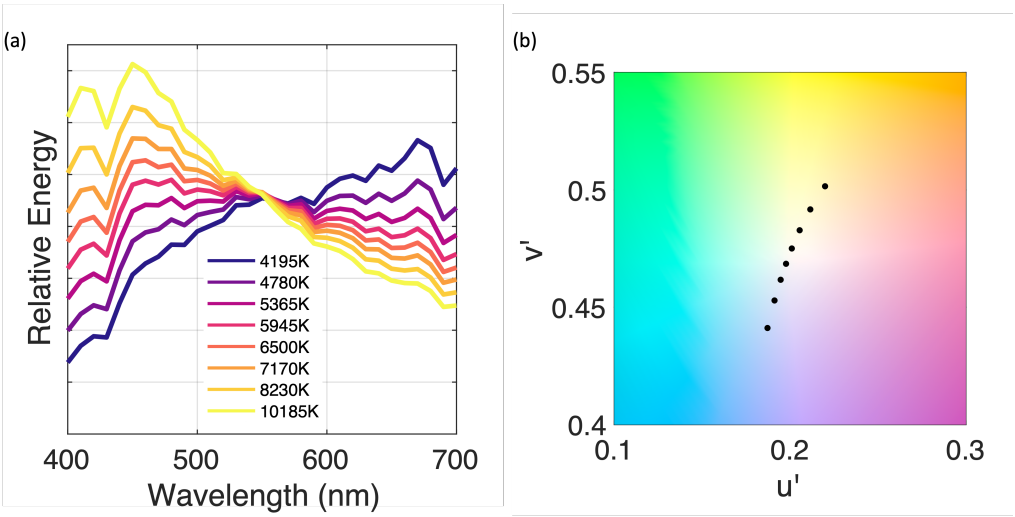


Figure 3.2: The eight illuminants chosen in the base dataset. (a) The energy spectra of the eight daylight illuminants chosen. (b) The u'v' chromaticities of the illuminants. Each illuminant is separated by a colour difference of $5\Delta E$ from its nearest neighbouring illuminants.

Surface reflectance dataset

For the surface reflectance dataset, I use the 49,667 real-world surfaces reflectances from the standard object colour spectra database for colour reproduction evaluation (SOCS; ISO/TR/16066:2003). The chromaticities of the surfaces, as rendered under equal energy white, from the SOCS database are shown in CIE u'v' space in Fig. 3.3(a). The SOCS database is composed of a range of natural and artificial surface samples, falling into the following eight broad categories: skin, flowers, leaves, Krinov

(which itself is a Soviet dataset collected by Krinov, 1947, and contains natural objects of categories (1) trees and bushes, (2) grass, (3) moss and lichen, (4) flowers and vegetables, (5) soil, (6) asphalt, (7) water and snow), photo, paints, printer, graphic. The breakdown of the SOCS database by category is depicted in Fig. 3.3(b). (Note, the majority of skin samples in the SOCS database are from East Asian individuals).

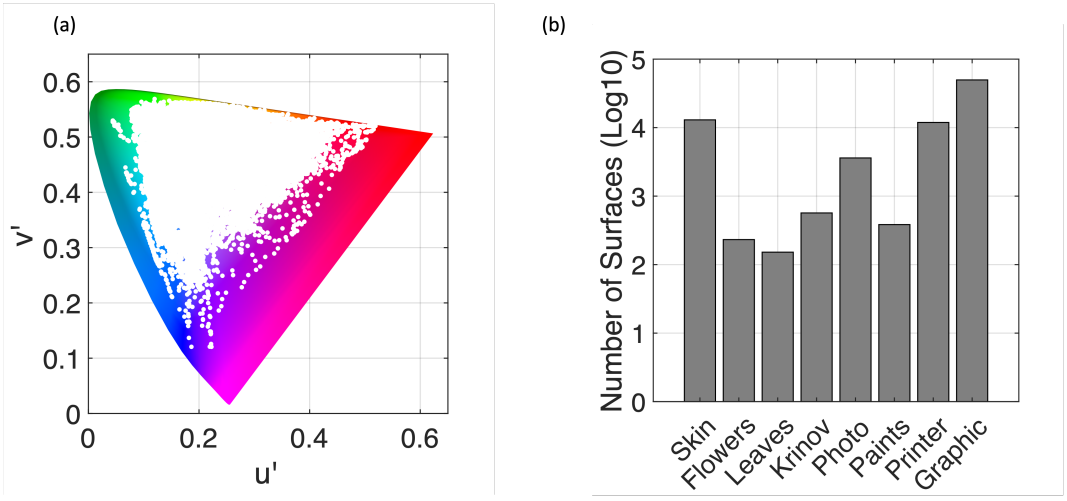


Figure 3.3: The 49,667 surface reflectance samples from the SOCS dataset used in this analysis. (a) The chromaticities of the surfaces in the SOCS dataset shown in u'v' chromaticity space. (b) The breakdown of the categories of surface reflectance data types found in the SOCS dataset. One can see that the SOCS dataset covers the majority of the CIE u'v' chromaticity space. The majority of samples in the SOCS dataset are artificial surfaces from graphic or printer samples, but skin samples also represent a substantial proportion of the reflectance samples.

Random Forest Algorithm

I use a random forest classifier for this task. This classifier earns its name as it is composed of many individual decision trees, working together in a single "forest". Each tree in the forest makes a vote on the output class, given the input data, and the forest then takes the majority of the votes from individual trees to give a single output classification from the forest. Each tree in the forest is only trained on a subset of the training data which prevents overfitting of the model to the training set, something to which simple decision trees often fall victim. A further advantage of the random forest classifier over alternatives such as a decision tree, or k-Nearest Neighbour, is that the experimenter has access to the votes from each of the individual trees in the forest, and thus one can attach confidence to each classification from the forest, and importantly, change the decision rules for classification. This idea will be particularly important for Experiment 4, described later in this chapter, where the optimal strategy of combining information from multiple surfaces is discussed.

Training and Test Data

Firstly, the reflectance samples are randomly split into training and test sets, with the training set containing 80% of surfaces and the test set the remaining 20%. Then, within the training and test sets, each illuminant and reflectance spectrum are paired combinatorically, to produce a single radiance spectrum for each pairing. Each of these radiance spectra are then reduced to the photoreceptor signals that they elicit. I use the 10° Stockman and Sharpe cone fundamentals (Stockman et al., 1999; Stockman & Sharpe, 2000) to calculate the cone signals from the radiance spectra, to model the sampling of the average human trichromatic observer. During training, as depicted schematically in Fig. 3.4(a), the random forest classifier is given the cone photoreceptor signal triplet, along with the CCT of the illuminant that was used to derive the radiance spectrum that elicited those signals. During testing, as shown schematically in Fig. 3.4(b), the random forest classifier is given only the cone photoreceptor signal triplet and is required to classify the illuminant contribution to one of the eight CCTs. It is important to emphasise that at no point does the classifier have access to full spectral information from the surface and illuminant; it has access only to the three scalar cone photoreceptor signals from each illuminant-reflectance pairing.

Evaluation metrics

The classifier output can be evaluated using a range of standard metrics, visualizations, and custom metrics relevant to this particular classification task. These metrics are described below.

Accuracy: Perhaps the most commonly used evaluation metric of classifier performance is accuracy. Accuracy is the proportion of classifications that the classifier got correct, in our case the fraction of times that the random forest correctly identified the illuminant. Accuracy is defined formally in Eq. 3.3,

$$\text{Accuracy} = \frac{\text{Number of correct classifications}}{\text{Total number of classifications.}} \tag{3.3}$$

F1 Score: Another standard metric for classifier evaluation is F1 score: the harmonic mean of precision and recall. While accuracy of a multi-class problem can be turned into a binary outcome as in Eq. 3.3, precision and recall cannot be defined as a binary outcome for a multi-class problem. One must thus turn the multi-class problem here into a series of binary classifications, calculate precision and recall for each binary problem and then take the average. In the case of binary classification with respect to the 6500K category there are four possible outcomes of the classifier. The classifier might

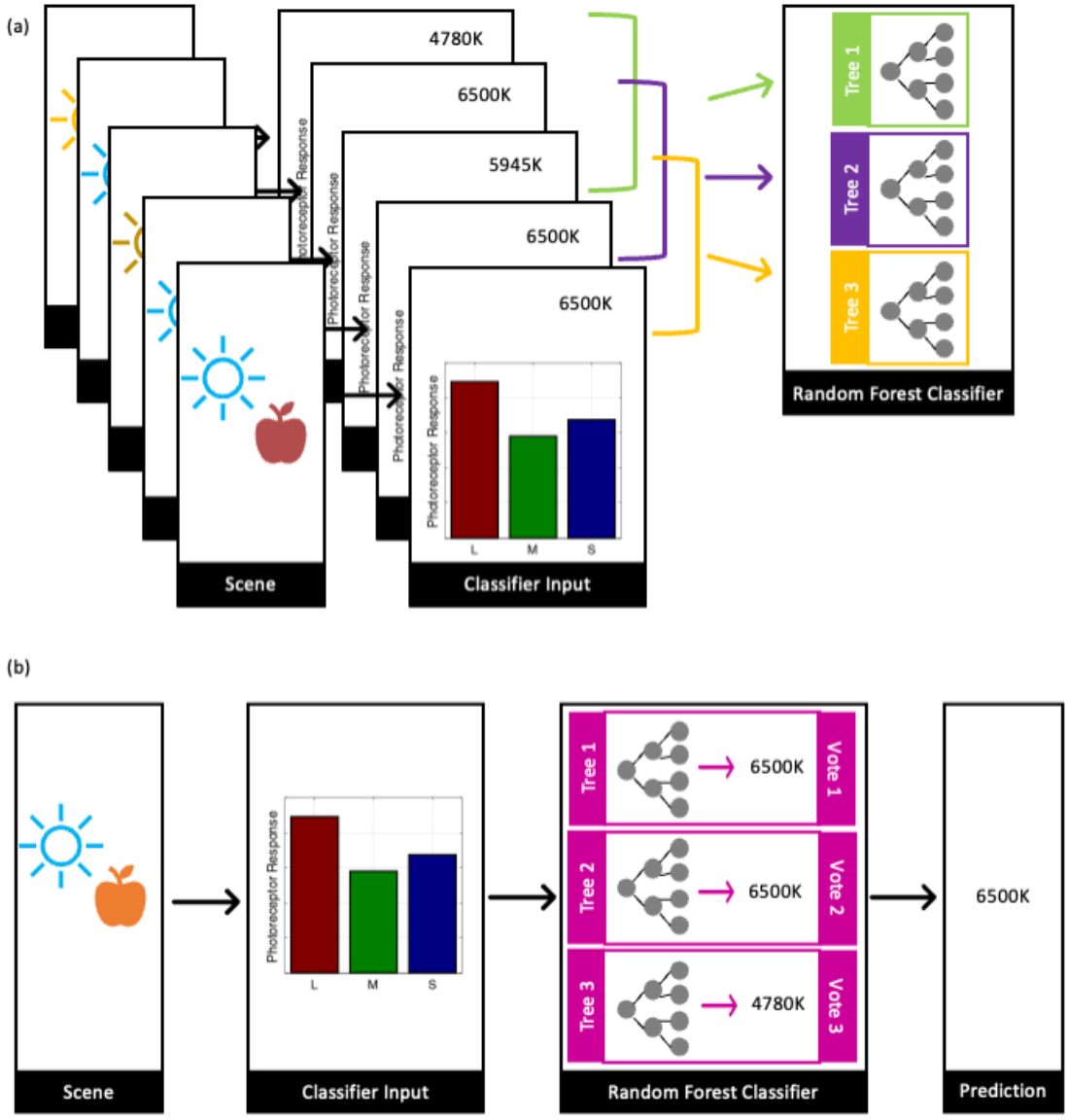


Figure 3.4: A schematic of the pipeline for training (a) and testing (b) the random forest classifier in this analysis. In this schematic I depict a random forest with just three decision trees, while the random forest actually used in this analysis contained 100 trees. First the scenes are created through every possible pairing of each illuminant spectrum with each surface reflectance to produce a single radiance spectrum from every combination. The radiance spectra are then reduced to the three cone photoreceptor signals of the average normal human trichromat, using the 10° spectral sensitivities of Stockman et al., 1999; Stockman and Sharpe, 2000. During training, (a), the cone photoreceptor triplets of 80% of the radiance spectra are given as input to the classifier along with the CCT label of the illuminant from the scene that gave rise to those cone photoreceptor signals. Individual trees in the forest see a subset of these photoreceptor signals and CCT labels during training as the forest learns the features of the input dataset. During testing, (b), only the cone photoreceptor signals elicited from the scenes in the test dataset are presented to the forest. Each tree in the forest votes on the illuminant CCT. The forest pools together the votes from each tree using a majority rule to make a single output classification, or prediction, of the illuminant CCT.

have correctly predicted the illuminant as 6500K, which is called a true positive (TP). Alternatively, the classifier might correctly identify the illuminant as not 6500K, called a true negative (TN). However,

the classifier might also predict the illuminant to be 6500K when the ground truth was in fact a different CCT, and would thus be a false positive (FP). And finally, the classifier may predict the illuminant to be a different CCT when the illuminant was in fact 6500K, a false negative (FN).

Precision is the fraction of positive predictions that were correct, i.e. the proportion of times the classifier identified the illuminant correctly to be 6500K, out of the total number of times the classifier predicted the illuminant to be 6500K, as in Eq. 3.4,

$$\text{Precision} = \frac{TP}{TP+FP}. \tag{3.4}$$

Recall is the fraction of actual positives that were identified, i.e. out of all of the ground truth 6500K illuminants, how many the classifier identified as being 6500K, as in Eq. 3.5,

$$\text{Recall} = \frac{TP}{TP+FN}. \tag{3.5}$$

For an imperfect classifier, there is a trade-off between precision and recall. Thus F1 score is taken as the harmonic mean between the two, as in Eq. 3.6,

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3.6}$$

F1 is calculated for each binary problem and then averaged across each binary case for a single F1 metric for our multi-class problem.

Matthew's Correlation Coefficient: The final standard metric used to evaluate performance here is the Matthew's correlation coefficient (MCC). The MCC is a balanced metric which considers all of the information about the output of a binary classifier i.e. both true and false positives and true and false negatives. The MCC is defined as in Eq. 3.7,

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \tag{3.7}$$

The MCC can be interpreted similarly to a correlation coefficient, where an MCC of 1 indicates perfect classification and 0 random assignment. As with F1 score, the MCC is calculated for each binary problem and then averaged.

Confusion matrix: The possible outcomes of the classifier can also be plotted as a heatmap of frequency of occurrences of each outcome in a confusion matrix. A confusion matrix plots the true class frequencies against the predicted class frequencies, for each possible outcome, as shown for the

simplest case of binary classification in Fig. 3.5(a). In the case of multi-class classification, as we have here, the confusion matrix will extend along each dimension as the number of classes increases. The negative diagonal represents the correct class being assigned to each illuminant, while anything off the negative diagonal represents a classification failure. For multi-class classification problems one can thus use the confusion matrix plot to visualise where failures in performance occurred. In our case, failures in prediction where the classifier predicts the neighbouring CCT from the ground truth are deemed less severe than failures where the classifier predicts a CCT further from the ground truth. Thus, the perfect classifier will plot along the negative diagonal, and the less tightly the confusion matrix plots along the negative diagonal, the worse the performance of the classifier.

ΔE *Confusion:* Finally, one can define a metric specific to the case of illuminant classification. As stated above, classifier errors where the classifier predicts the neighbouring illuminant are less severe in our case than when the classifier predicts an illuminant further away in CCT from the ground truth. Expanding on this idea, one can imagine the case where choosing the wrong, but almost right, illuminant has no perceptual consequence. To put it another way, there may be some set of surface reflectances, which when viewed under similar illuminants, elicit a sufficiently similar set of cone signals such that the two radiance spectra are perceptually indistinguishable. In the extreme case, this would be when the spectra are metameric, but there might also be cases where the two spectra are not metameric, but are still within one unit ΔE colour difference and are thus perceptually indiscriminable to the average observer. I therefore define an additional evaluation metric, which I call ΔE confusion to account for this case. The ΔE confusion is defined as the colour difference in terms of ΔE between the ground truth radiance spectrum, and the radiance spectrum that would have been generated if the true surface reflectance were rendered under the predicted illuminant class. A schematic of the calculation of the ΔE confusion is shown in Fig. 3.5(b). For a full description of the calculation of the ΔE between two spectra see Chapter 1. The ΔE confusion values for each classification may be plotted on a histogram, or a certain threshold value may be applied and the proportion of classifications that fall below the ΔE confusion threshold reported. Here, I choose a ΔE confusion threshold of $5\Delta E$ (which is considered a non-negligible but small colour difference, and generally considered acceptable for many colour rendering applications (Sharma et al., 2005)). A ΔE confusion of 1 would correspond to perfect perceptual classification, i.e. all of the surfaces are rendered within the chosen ΔE threshold, while a ΔE confusion of 0 would correspond to complete failure, i.e. no surfaces are rendered within the chosen ΔE threshold.

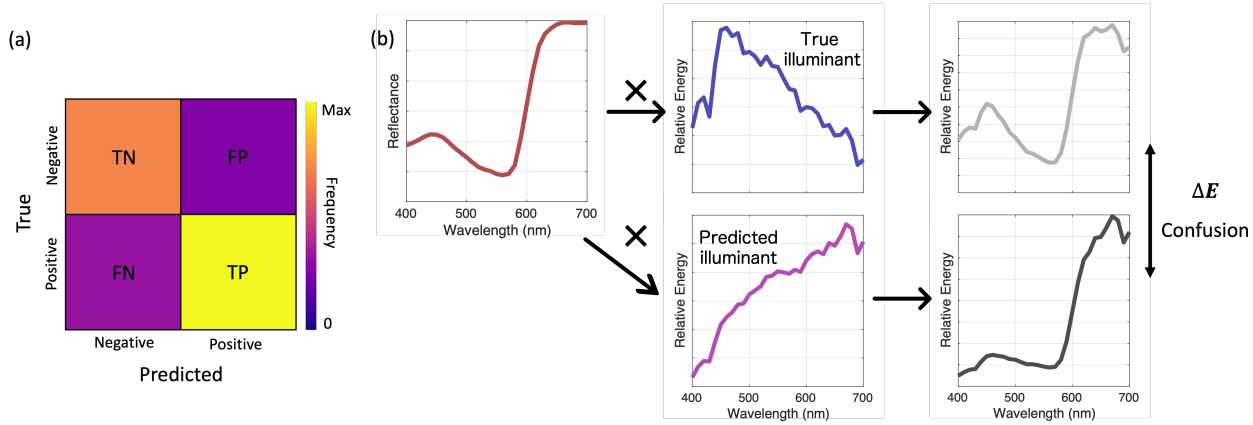


Figure 3.5: Schematic depicting the classifier evaluation visualizations used in this chapter. (a) A diagram of the confusion matrix in the format it is used in this chapter, where the ground truth class is plotted against the predicted class, and the number of classifications that fall into each outcome plotted as a heatmap. The schematic shows the simplest case of binary classification. (b) The custom evaluation metric, ΔE confusion, is calculated as the colour difference in terms of ΔE between true surface reflectance when rendered under the true illuminant vs the true surface reflectance when rendered under the predicted illuminant.

3.2.2 Results

For the base case, the accuracy, F1, MCC, and $<5\Delta E$ confusion, of the random forest algorithm performance on the test data are reported in Table 3.1. While the standard accuracy, F1, and MCC scores may seem relatively low, one must remember that chance performance in this task would be an accuracy of 0.125, thus the classifier is clearly performing well-above chance. Further, the confusion matrix, as shown in Fig. 3.6(a), clearly shows the majority of classifier predictions fall along the negative diagonal, or close to it, showing that when the classifier does predict the illuminant incorrectly, it doesn't get it that wrong. This is further highlighted by the high $<5\Delta E$ confusion metric and the histogram of ΔE confusion values of all outcomes, as shown in Fig. 3.6(b).

Accuracy	0.296
F1	0.293
MCC	0.20
$<5\Delta E$ confusion	0.70

Table 3.1: Table showing the results of Experiment 1, the base case of the illuminant classification problem, with eight daylight illuminants and 49,667 surface reflectance samples in the dataset.

3.2.3 Discussion

The results of Experiment 1 thus show that the random forest algorithm is able to learn to perform the task of illuminant classification on the base dataset of eight daylight illuminants and 49,667 surface reflectance samples when given the cone signals from only a single illuminated surface. As a reminder to the reader, the task given to the classifier is not one that the human visual system is known to

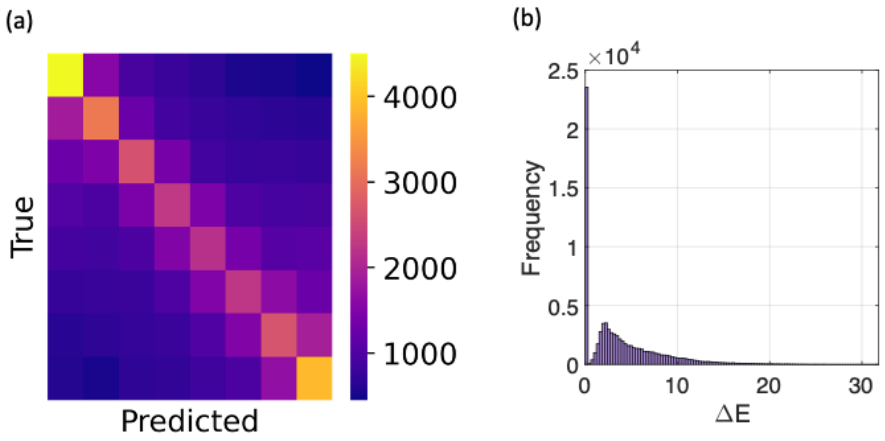


Figure 3.6: Visualizations of the performance of the classifier on Experiment 1: the base case. (a) Confusion matrix showing the performance of the random forest classifier on the test data set. The confusion matrix is ordered such that the lowest CCT is the top row of true class, ascending so the highest CCT is the bottom row, and the predicted classes are ordered with increasing CCT from left to right. A perfect classifier would thus plot entirely down the negative diagonal. The majority of classifications fall along the negative diagonal, or close to it, demonstrating good performance on this task. (b) Histogram of the ΔE confusion values for the classifier predictions for this experiment. The majority of classifications render a ΔE of <1, and very few classifications led to a ΔE greater than 5.

be able to solve without additional constraints: this is an underdetermined problem. The results here show therefore that there are sufficient statistical regularities in the real-world (or at least in the dataset used here to represent the real-world) that the classifier can learn in order to achieve illuminant classification with such accuracy. Such statistical regularities may thus also be used by the human visual system to learn illuminant perception.

3.3 Experiment 2: Varying Illuminant Dataset

In Experiment 2, I investigated what impact changing the illuminants had on classifier performance. There were three questions that were tested in this experiment. The first, tested in Experiment 2a, was whether or not increasing or decreasing the number of illuminants chosen along the daylight locus had a meaningful impact on classifier performance. That is to say with more illuminant classes to choose from, even if the classifier shows a decrease in accuracy, does the classifier also show a large decrease in ΔE confusion i.e. a greater number of perceptually significant misclassifications? The second, tested in Experiment 2b, was whether or not illuminant classification would be better between illuminants along the daylight locus, or illuminants from the daylight orthogonal axis. The final question, as in Experiment 2c, was whether or not there was evidence of a "blue bias" in the classifier performance, mirroring the reduction in sensitivity along the blue chromaticity axis seen by human observers in

illuminant discrimination tasks (Pearce et al., 2014; Radonjić et al., 2016; Radonjić et al., 2018).

3.3.1 Methods

The methods for training and testing the classifier remain as described in Experiment 1. As with the base case, the intensity of the illuminants was randomized in all cases. The only difference is the illuminant database used to render the surfaces. In Experiment 2a, as the total number of illuminants in the dataset increased or decreased compared to the eight used in the base case, there were subsequently a greater or smaller number of resulting radiance spectra, and thus a smaller or bigger dataset fed into the classifier.

I manipulated the illuminant dataset in three ways, in Experiment 2a, 2b, and 2c. In Experiment 2a, the spacing of illuminants along the daylight locus was varied. For the dataset with a smaller number of illuminants tested, I used 5 illuminants along the daylight locus, as shown in Fig. 3.7(a). For the dataset with decreased spacing, i.e. more illuminant classes, I took illuminants from the daylight locus that were separated by $1\Delta E$, which thus should be just above the just-noticeable difference for the average human observer (Fig. 3.7(b)). This resulted in a dataset of 40 illuminants. In Experiment 2b, I took 9 illuminants evenly spaced along the daylight orthogonal axis, such that the extreme green and magenta were separated from D65 by the same ΔE as the extreme daylight illuminants in the base case, as specified in CIELAB space and shown in Fig. 3.7(c). The orthogonal illuminants were derived by taking Kodak filters to generate the magenta and green illuminants, and mixing their spectra with the spectrum of D65 until the desired spacing of the illuminants was reached. In Experiment 2c, I took a mixture of daylight and daylight orthogonal illuminants, taking the 5 daylight illuminants from Fig. 3.7(a) and every other illuminant from the 9 along the orthogonal axis from Fig. 3.7, excluding the central illuminant from the orthogonal locus which was D65 and already captured in the daylight axis. The mixed set of daylight and daylight orthogonal illuminants is shown in Fig. 3.7. To tease apart whether a "blue bias" is something inherent in the structural regularities in the real-world dataset, or something learnt from past exposure, I also consider a case where the mixed dataset is purposefully imbalanced such that three in four surfaces rendered under the orthogonal illuminants are dropped from the classifier input during training, so that the classifier sees four times as many samples rendered under the daylight locus than under the orthogonal locus.

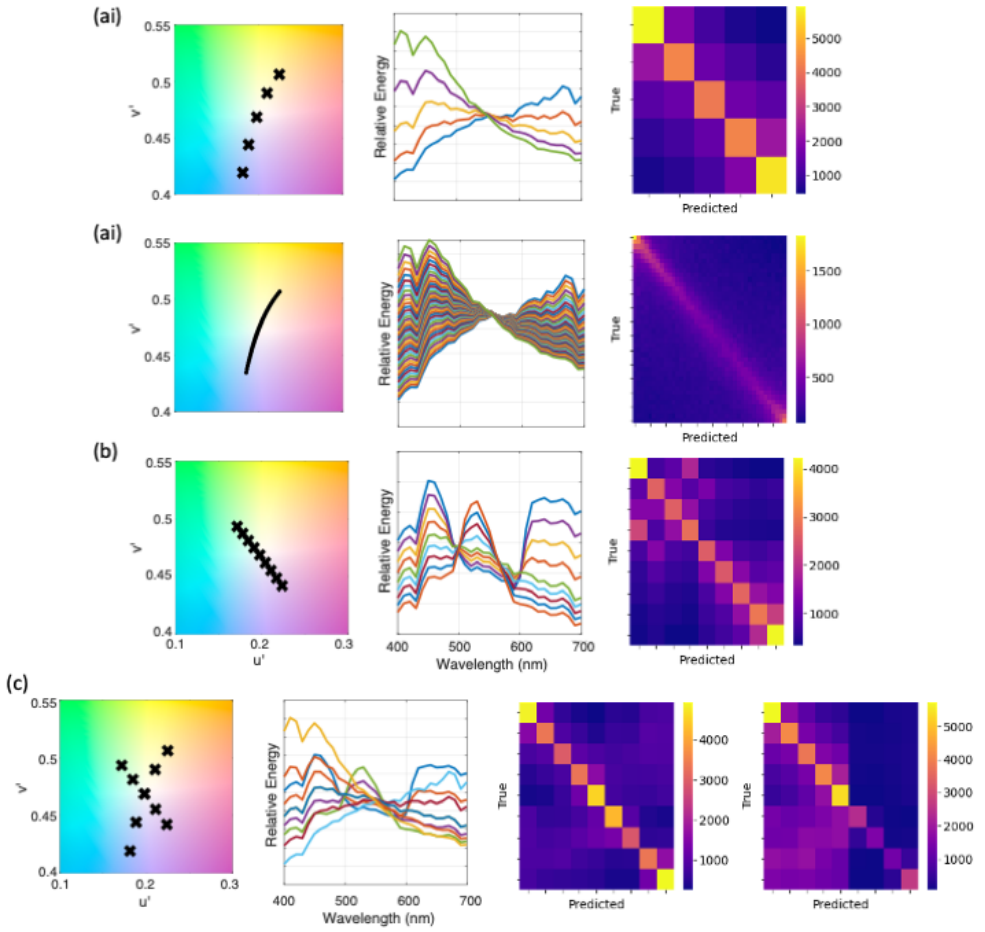


Figure 3.7: Visualizations of the performance of the classifier on Experiment 2. Each row shows, from left to right, the CIE u'v' chromaticities of the illuminants in that particular manipulation, the spectra of the illuminants, and the confusion matrix of the classifier performance on the test set. (ai) The manipulation with five daylight illuminants. The confusion matrix is ordered such that the true class descends in order of increasing CCT, and the predicted class shows increasing CCT from left to right, such that a perfect classifier would fall along the negative diagonal. (aii) The manipulation with $1\Delta E$ separation of the daylight illuminants, resulting in 40 distinct CCTs. The confusion matrix is ordered as in (b). (c) The orthogonal illuminant manipulation. The confusion matrix is ordered such that the top left represents the green illuminants with illuminants becoming increasingly magenta as one moves down and to the right. The central row and column are the D65 class. (c) The mixed illuminant dataset. In this case two confusion matrices are plotted, the left most representing the case where the classes are balanced, and the right most the case where the classes are purposefully imbalanced. In both confusion matrices, the classes are ordered such that the five daylight illuminants are the top five classes along the true class axis, and left most classes along the predicted class axis. The orthogonal classes are the bottom most and right most, with greener illuminants towards the centre of the matrix and the magenta illuminant occupying the bottom row and right most column.

3.3.2 Results

Results for Experiment 2a are shown in Table 3.2. Rows report the accuracy, F1 score, MCC and $<5\Delta E$ confusion metric classifier performance results on the test set for each case, as well as chance performance level. While the standard metrics increase and decrease with the number of classes as expected, the $<5\Delta E$ confusion metric changes by only 2% and 8% compared to the base case when the number of illuminants is increased and decreased respectively. This represents a very small

impact on the performance of the classifier when evaluated in meaningful perceptual terms. This impact can also be seen visually in the confusion matrices that are plotted in Fig. 3.7(a)-(b). The vast majority of classification output continues to fall along the negative diagonal, or neighbouring squares, consistent with the idea that when the classifier does misclassify the illuminant, it does not make large classification errors. Note, due to the relatively large compute time to calculate the ΔE confusion metric, from here on, only the confusion matrices are plotted as they visually depict the same information captured by the ΔE confusion metric, as demonstrated here and in the base case.

	Base case	5 illuminants	40 illuminants
Chance	0.125	0.2	0.025
Accuracy	0.296	0.491	0.067
F1	0.293	0.489	0.066
MCC	0.20	0.36	0.04
<5ΔE confusion	0.70	0.62	0.68

Table 3.2: Table showing the results of Experiment 2a, where the number of illuminants is increased or decreased compared the base case.

For Experiment 2b, where 9 illuminants are taken from the orthogonal axis, the classifier accuracy on the test set was 0.305, the F1 was 0.303, and the MCC 0.22, all of which are very close to the performance results seen in the base case (though note as there were 9 classes to choose from here, chance would be 0.11, instead of 0.125 in the base case, so if anything, these results show a slight increase in performance compared to the base case). However, while the standard metrics show only a small change from the base case, the confusion matrix as plotted in Fig. 3.7(c) shows a much greater spread around the negative diagonal, indicating that when the classifier does make an error, it makes a more perceptually noticeable error than in the base case.

Finally, the results of Experiment 2c show an increase in performance compared to the base case and the orthogonal case alone, with an accuracy of 0.385, an F1 of 0.382, and an MCC of 0.31 on the test set. When daylight is oversampled compared to the orthogonal axis the accuracy drops slightly to 0.347 (F1=0.334, MCC=0.27). However, the confusion matrices for the mixed dataset show the true impact of oversampling daylight in the training set (Fig. 3.7(d). One can see that while the majority of classifier output falls along the negative diagonal for the balanced mixed dataset (though there is more spread seen than in the base case), when the classifier dataset is purposefully imbalanced there is an overtendency to classify along the daylight locus (shown by the clump of misclassifications in the bottom left of the confusion matrix, which represent true orthogonal illuminants misclassified as daylight illuminants).

3.3.3 Discussion

The results presented above support three main conclusions. Firstly, the results from Experiment 2a show there is no meaningful impact of varying the number of classes along the daylight locus on performance. That is to say any benefit gained by restricting the number of categories for the classifier does not translate into an increased performance in terms of perceptually relevant discriminability. In fact, the worse performance in terms of $<5\Delta E$ confusion was the case with the smallest number of classes, which is not surprising, as there are larger intervals between the classes, so incorrect classification is more likely to lead to a larger colour difference.

Secondly, the results from Experiment 2b do not suggest a large difference between illuminant classification along the daylight locus and daylight orthogonal axis. Taken together with the findings Experiment 2c, where the classifier shows a bias towards daylight illuminants only when they are oversampled in the training data, these results suggest that any "blue bias" is not inherent in the statistical regularities in the real-world, but a learned bias. This finding offers further support to the idea that illuminant perception relies on some form of prior which must be learnt (Pearce et al., 2014; Radonjić et al., 2016; Álvaro et al., 2017; Radonjić et al., 2018; Aston et al., 2019; Brainard & Freeman, 1997).

3.4 Experiment 3: Comparing surface statistics

In Experiment 3 I investigate how the nature of the surface dataset influences classifier performance on the illuminant classification task. For the comparison between surfaces I use the base case illuminant set, i.e. the original eight daylight illuminants from Experiment 1. An analysis of the surface statistics that influence illuminant perception is informative as it allows one to ask, hypothetically, what gamut of surfaces are most separable under different illuminants i.e. what gamuts carry the most information to classify within a given family of illuminants.

3.4.1 Methods

I analyse surface statistics in two ways. Firstly, in Experiment 3a, I use the classifier results from Experiment 1 and perform an analysis on the test data to analyse the statistics of the surfaces associated with correctly classified and incorrectly classified illuminants. Specifically, I analyse the surfaces in terms of lightness, and CIE u'v' chromaticity. Secondly, in Experiment 3b, I break the surfaces down by category, and group the 41,008 artificial surfaces (graphic, printer, paints, photo) together and the

8,659 natural surfaces (Krinov, leaves, flowers, skin). I train two new classifiers separately on these two distinct gamuts, as this enables us to probe how the surface gamut impacts the ability of the classifier to learn illuminant estimation. Note, the choice to split the surface data into natural and artificial surfaces is somewhat arbitrary, but rooted in the motivation that the gamut of natural surfaces may be more separable under the daylight locus than an artificial gamut. More importantly, the natural and artificial surfaces provide two clearly distinct gamuts in terms of chromatic extent (as shown in Fig. 3.9(b)), so this division is a neat way of generating two different gamuts (with the bonus of a possible physiological basis) to test the impact of the gamut of surfaces on classifier performance.

3.4.2 Results

The results of Experiment 3a are shown in Fig. 3.8. Fig. 3.8(a) shows histograms of the CIE L^* lightness values of the surfaces (as rendered under equal energy white) for the surfaces for which the illuminant is correctly classified and incorrectly classified. One can see that the surfaces for which the illuminant is misclassified are approximately normally distributed across lightness. However, the surfaces associated with correct classifications show a highly peaked distribution, with a peak at $L^* = \sim 60-70$. This shows that surfaces with a low lightness are less likely support correct classification, and hints that there may be a cluster of similar surfaces with a lightness $L^* = \sim 60-70$ that are particularly favourable as cues to the illuminant. Further, the impact of surface chromaticity is shown in Fig. 3.8(b). To analyse the impact of surface chromaticity, the CIE u'v' chromaticity diagram is broken into a 100 x 100 grid (over the range 0-0.6 in both u' and v') and each surface reflectance is assigned to the grid patch that its chromaticity falls within. Then the percentage of surfaces that supported correct classification within a particular chromaticity patch is calculated (as a percentage of the total number of surfaces that fall within that chromaticity patch), and depicted visually in the heatmap in Fig. 3.8(b). One can see that while surfaces across the full extent of the chromaticity diagram do support correct classifications (see Fig. 3.3 for a reminder of the extent of the chromaticity diagram covered by the surfaces in the SOCS dataset), the classifier performs best for desaturated surfaces.

In Experiment 3b, when the surfaces are broken down into natural vs artificial surfaces and a classifier trained and tested on each separately, one sees a significantly higher performance on natural surfaces than artificial surfaces, as shown in the confusion matrices in Fig. 3.9(a). For natural surfaces, classifier accuracy was 0.637, with an F1 of 0.635, and an MCC of 0.58 (as a reminder, chance would be 0.125 when classifying between the 8 daylight illuminant classes used here). Natural surfaces also showed an extremely tight clustering along the negative diagonal in the confusion matrix. For artificial

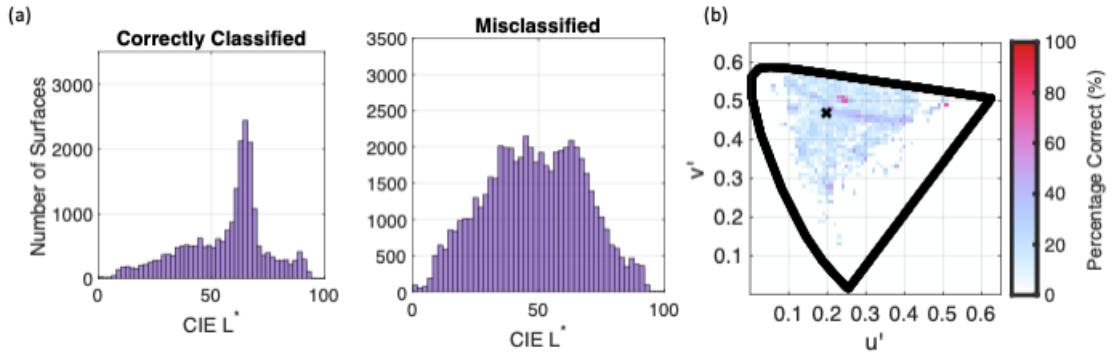


Figure 3.8: The impact of surface lightness and chromaticity on classifier performance. (a) Histograms showing the number of surfaces of a certain CIE L^* lightness that are correctly classified (left) and incorrectly classified (right) (b). Heatmap showing the percentage of surfaces of a particular chromaticity that are correctly classified. The CIE u'v' chromaticity diagram is broken down into a 100 x 100 grid (over the range 0-0.6 in both u' and v'), and surfaces assigned to the grid patch associated with their chromaticity. The percentage of surfaces within each grid patch that have been correctly classified is shown. Note, by comparison to Fig. 3.3(a) one can see that the regions of the chromaticity space that have 0% correctly classified surfaces correspond to the regions of the chromaticity space that were not covered by the surfaces in the SOCS dataset.

surfaces, classifier accuracy was much lower at 0.238, with an F1 of 0.235, and an MCC of 0.13 (though still markedly above chance). Crucially, for artificial surfaces there was not very tight clustering along the negative diagonal in the confusion matrix. The gamut of natural and artificial surfaces is shown in Fig. 3.9(b). One can see that the gamut of natural surfaces is much more clustered and desaturated than the gamut of artificial surfaces.

3.4.3 Discussion

The results of Experiment 3 are perhaps unsurprising. In Experiment 3a, when trained on the whole set of surface reflectances the classifier performs better when classifying the illuminant from more desaturated surfaces than more saturated ones. As one can see from Fig. 3.9(b) desaturated surfaces are more common than highly saturated surfaces. Thus, from a Bayesian perspective, the best guess for a surface colour is desaturated and so classifier performance will be best when the surface matches this prior.

In Experiment 3b, when trained separately on natural and artificial surfaces the classifier performs better in the natural surfaces scenario than artificial surfaces scenario. Natural surfaces form a more limited gamut than artificial surfaces and such smaller, more tightly clustered gamuts are more clearly separable under the different illuminant chromaticities of the daylight locus. To demonstrate this idea of gamut separability, let's consider just the set of natural surfaces. Within natural surfaces, the classifier performance is much higher for skin samples, with an accuracy of 0.64, compared to other

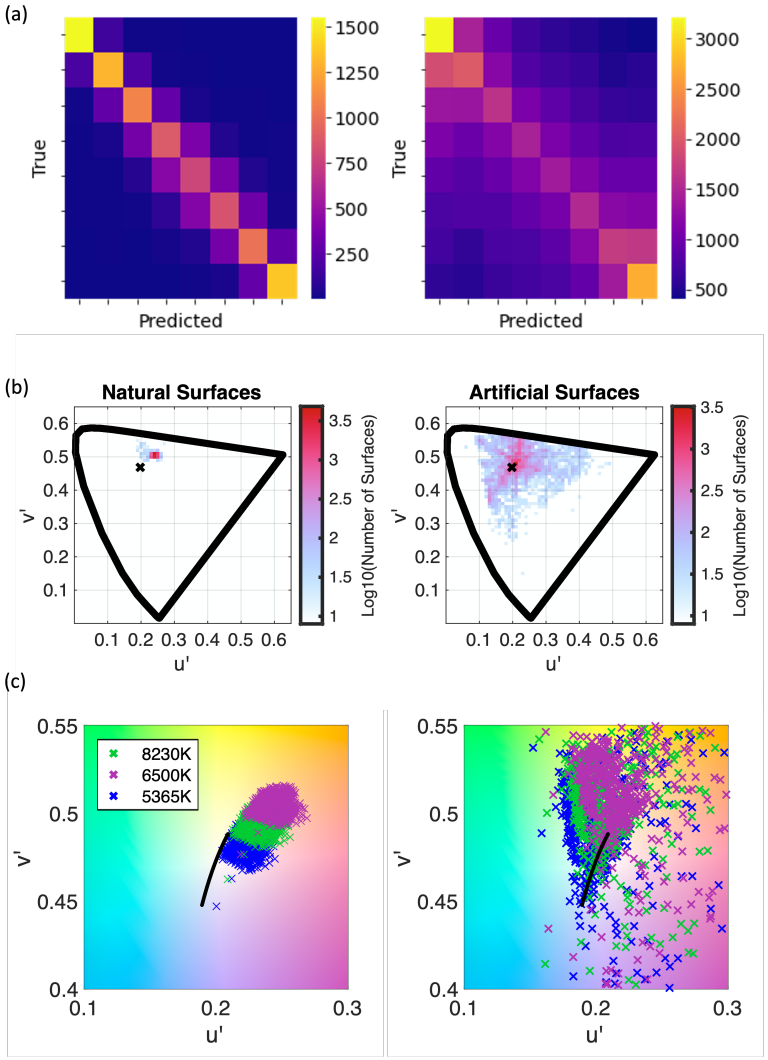


Figure 3.9: Natural vs artificial surfaces on classifier performance. The classifier is trained and tested on either the 8,659 natural or 41,008 artificial surfaces in this case. (a) The confusion matrices of the classifier output when trained and tested on the natural surfaces (left) and on the artificial surfaces (right). As before, the lowest CCT is towards the top-left, and highest CCT class the bottom-right. (b) The frequency of the chromaticities of the natural (left) and artificial (right) surfaces. As in the visualizations in Experiment 2 the CIE u'v' chromaticity diagram is broken down into a 100 x 100 evenly spaced grid (over the range 0-0.6 in both directions) and the total number of surfaces from each category falling into each grid patch counted and depicted on the heatmap. D65 is depicted by the black cross. (c) Demonstration of the separability of certain gamuts under the daylight locus. If one breaks down the natural surfaces in the test set into skin samples vs the other natural samples (Krinov, flowers, and leaves), one finds an increased classifier performance on the skin samples (Accuracy = 0.64), compared to on the other samples (Accuracy = 0.17) in the context of training only on natural surfaces. This figure shows the CIE u'v' chromaticity of the skin samples (left) vs the other natural samples (right) rendered under one of three CCTs (5365K, 6500K, or 8230K). One can see that the skin samples more clearly separate under the three illuminants, and are much more tightly clustered than the other samples. Specifically, one can see that there are regions of the CIE u'v' chromaticity space that are only visited by skin samples under a particular illuminant.

natural surfaces, with an accuracy of just 0.17. To illustrate why the skin gamut has an advantage over the other natural surfaces, Fig. 3.9(c) shows the gamut of skin samples (left) and gamut of the other natural surface samples (right) as rendered under 5365K, 6500K, and 8230K (only three CCTs are chosen

here for readability of the figure). One can clearly see that the skin gamut is more tightly clustered and as CCT increases, the skin gamut translates as a unit along the daylight locus. Comparatively the other natural surfaces form a less tightly clustered gamut, which does not so clearly separate into three distinct clusters, under the three different illuminant chromaticities. Thus, in the context of all natural surfaces, there exists regions of the chromaticity space that are only visited by skin samples under a particular illuminant, so if a sample has a chromaticity that falls in such a region, it must be a skin surface under one particular illuminant. Taken together, the results from Experiment 3 show that the nature of the gamut, i.e. the surface statistics, on which the classifier is trained and tested matters for classifier performance, with increased classifier performance occurring on more restricted gamuts and on surfaces which more closely match any learned prior of surface chromaticity.

3.5 Experiment 4: Strategies for combining multiple surfaces

In Experiment 4 I consider the scenario where multiple surfaces are present in the scene. This thus allows me to ask, what is the optimal strategy for combining information across multiple surfaces to optimise classifier performance? I consider strategies where information is pooled before being fed into the classifier versus strategies where the classifier sees information from each surface and then pools its decisions from each surface to make one single prediction of the illuminant in the scene.

3.5.1 Methods

I use the base set of eight daylight illuminants and the full 49,667 SOCS surfaces in the multi-surface experiment here. I consider four strategies of combination as described below. The first two strategies pool the information before a decision is made. I refer to these strategies as pre-decision pooling. For the first two strategies, during training, 10,000 samples of 1-20 surfaces (depending on the number of surfaces being combined in each case) are drawn from the training data (sampling randomly with replacement). Then, during testing 1000 samples of 1-20 surfaces are taken from the test data (sampling randomly with replacement). The LMS signals that are fed into the classifier during both training and testing are chosen based on one of the two strategies described below.

1. The surface with the highest luminance in the scene is taken as the best cue to the illuminant. Thus the LMS signals from the surface with the highest luminance are fed into the classifier and the LMS signals from every other surface discarded.

2. The mean LMS signals from each surface when rendered under the illuminant are fed into the classifier, pooling information across all surfaces in the scene as the best cue to the illuminant.

Thus, in the above two strategies the information from multiple surfaces is reduced to a single LMS triplet which is all the classifier has access to, regardless of how many surfaces comprised the scene. These scenarios are designed to mirror the "brightest is white" (Land, 1977) and "grey-world" (Buchsbaum, 1980) type hypotheses, both of which assume that information across the scene is averaged before it is used as a cue to the illuminant.

An alternative approach would be to use information from each surface in the scene, and then pool that information into the final decision of the illuminant chromaticity. I refer to this approach as post-decision pooling. The other two combination strategies considered take this approach. For these strategies the classifier is trained as in Experiment 1. Then during testing, 1000 samples of 1-20 surfaces are drawn from the test dataset and rendered under the test illuminant. The classifier then sees the LMS signals from each of these 1-20 scenes, and makes a vote on each scene. Specifically, each tree in the forest makes a vote on each scene, meaning the forest outputs the probability for each illuminant for each of the 1-20 scenes. These probabilities can then be pooled to make a final decision on the global illuminant. I consider two strategies for combining these probabilities as described below.

1. The illuminant class with the single highest probability across all surfaces is taken as the predicted illuminant class. For example, in the hypothetical case of three surfaces in the scene, and just three illuminants to choose from, let's imagine that Surface 1 elicits the following tree votes per illuminant: I1 = 0.75, I2 = 0.05, I3 = 0.2, Surface 2 elicits: I1 = 0.1, I2 = 0.2, I3 = 0.7, and Surface 3 elicits: I1 = 0.05, I2 = 0.25, I3 = 0.7. In this strategy I1 would be chosen as the predicted illuminant as it has the highest single probability from any surface. This is a "winner takes it all" approach.

2. The illuminant class with the highest mean probability across all surfaces is taken as the predicted illuminant. For the above example, the mean probabilities would be I1 = 0.3, I2 = 0.17, I3 = 0.53, and thus I3 would be chosen as the predicted illuminant.

3.5.2 Results

Fig. 3.10 shows the classifier accuracy as a function of number of surfaces in the scene using the four different strategies of combining information from multiple surfaces. There is a clear advantage to

the post-decision pooling strategies over the pre-decision pooling strategies. Within the pre-decision strategies taking the LMS signals from the brightest surface is advantageous compared to taking the mean LMS signals, although with enough surfaces in the scene the two begin to converge. Of note is the initial dip in performance when taking the mean LMS signals pre-decision, presumably due to the fact that you are more likely to have one less informative or highly saturated surface in the scene that skews the LMS signals.

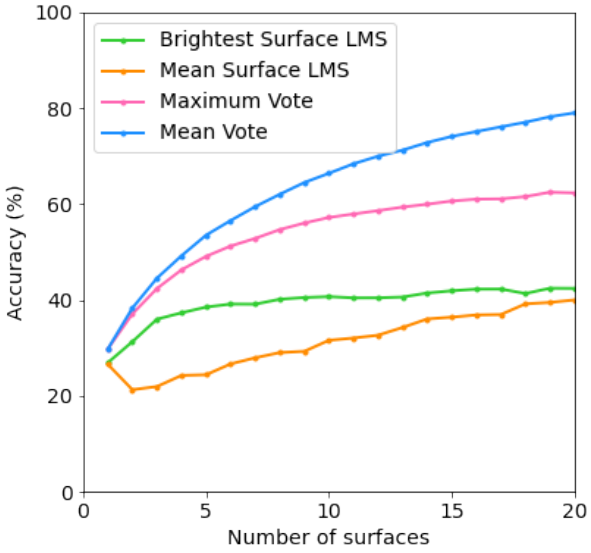


Figure 3.10: Classifier accuracy as a function of number of surfaces in the scene, for each of the two pre-decision pooling strategies (brightest surface LMS, and mean surface LMS) and the two post-decision pooling strategies (maximum vote, and mean vote). Note, the pre-decision and post-decision strategies are tested on classifiers that were trained differently, as explained in the main text. Thus, in this plot the performance of these two separate classifiers is depicted for each of the two testing scenarios for that classifier.

Conversely, for the post-decision strategies the mean vote of the trees in the forest always outperforms the maximum vote. For the post-decision strategies the biggest advantage comes from the addition of the first few surfaces beyond a single surface, and the rate of accuracy increase slows as subsequent surfaces continue to be added.

3.5.3 Discussion

The results suggest that the optimal combination strategy to pool information from multiple surfaces to improve performance in an illuminant classification task is to estimate the illuminant from each surface, average the probabilities for the illuminant estimates made on each surface, and make the single global decision of the illuminant that is most likely from those averaged probabilities. This strategy retains the most information across the four strategies considered here as it does not discard

information from any surface at any stage of the decision process, which every other strategy does at some stage. What is not considered here, but would be an interesting avenue for future work, are Bayesian type combination strategies. For example, given the advantage of skin samples, due to their limited gamut, perhaps the optimal combination strategy would be to predict whether any samples in the scene were skin, and if so, use just the prediction from that sample. Alternatively, higher lightness samples may be given additional weighting than samples lower lightness to achieve maximum performance.

3.6 Experiment 5: Availability of photoreceptor signals

The final experiment in this chapter, Experiment 5, analyses the impact of the photoreceptor signals available to the random forest illuminant classifier. This experiment considers the impact of having access to only the signals from two cone classes, as dichromatic observers do, and the impact having access to an additional rod or melanopsin signal.

3.6.1 Methods

I use the same dataset as in Experiment 1 i.e. 8 daylight illuminants with CCTs of 4195K, 4780K, 5365K, 5945K, 6500K, 7170K, 8230K, 10185K, and the 49,667 surface reflectance samples in the SOCS dataset, with 80% of the surfaces in the training dataset, and the remaining 20% in the test set. However, in this case I calculate all five photoreceptor signals from the resultant radiance spectrum. I again use the 10° Stockman and Sharpe cone spectral sensitivity functions (Stockman et al., 1999; Stockman & Sharpe, 2000), and the rod and melanopsin spectral sensitivity functions as defined in the CIE S026 standard (Wyszecki & Stiles, 1982; Lucas et al., 2014; CIE, 2018). I then consider different combinations of the photoreceptor signals. For each combination the classifier is trained and tested using only the photoreceptor signals from that combination e.g. for the deuteranope case the classifier is trained and tested with only L and S signals.

3.6.2 Results

The accuracy of the classifier performance for each combination of photoreceptor signals is shown in a bar chart in Fig. 3.11. One can see that having access to only the dichromatic cone signals results in decreased performance compared to having access to the trichromatic cone signals. However, if the dichromatic case also includes melanopsin or rod signals, accuracy is essentially equivalent to that in

the trichromatic cone-only case. Additionally, while the LMS + melanopsin or LMS + rod cases do offer a relatively large advantage in performance compared to LMS only, the relative performance gained is less than the performance increase when melanopsin or rod is available in addition to the dichromatic cases. For example the LMS + rod case and LMS + melanopsin have an accuracy 1.10 and 1.14 times greater than the LMS case, respectively, while the LM + rod and LM + melanopsin scenarios have an accuracy 1.47 and 1.44 times greater, respectively, than the LM signals alone.

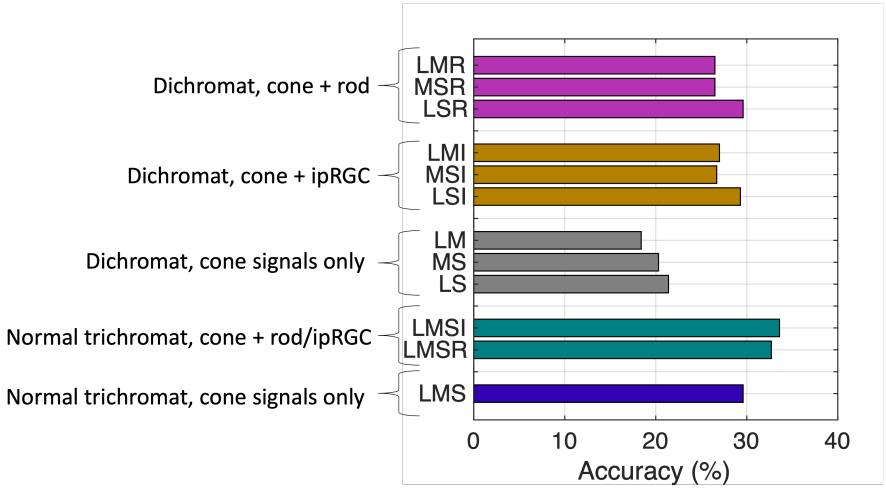


Figure 3.11: Impact of changing the availability of photoreceptor signals, from the three cone types, L, M, and S, rods, R, and melanopsin, I, on classifier performance. The accuracy of the classifier output on the test dataset is shown for the various combinations of photoreceptor signals given as input to the classifier (during both training and testing).

3.6.3 Discussion

It is not surprising that the best performance is possible when four photoreceptor signals are available, as this simply means we are more finely sampling the radiance spectra and thus retaining more information. Though it is interesting to note that the biggest advantage comes from a move from two photoreceptors to three (as opposed to moving from three photoreceptors to four). In particular these results allow one to speculate as to why dichromats may behave similarly to trichromats on illuminant discrimination tasks (Álvaro et al., 2017). Namely, if dichromats are able to utilise the melanopsin or rod signal for illuminant perception then dichromats should have access to similar information as the normal trichromats.

3.7 General Discussion

The results presented in this chapter shed some light onto the statistical regularities that are present in real-world scenes that may be used to constrain the problem of illuminant estimation. The work presented in this chapter contributes to the (limited) field of work on illuminant estimation in that it offers a novel method of addressing what properties of surfaces, illuminants, and the human photoreceptor signals offer the most information to solve an illuminant estimation problem. The statistical regularities in the real-world that allow us to constrain this problem are often quoted as fact, and have been empirically verified (Foster & Nascimento, 1994), yet we still lack a quantification of those regularities and the constraints they bring. The work presented here offers a significant step towards quantifying these regularities.

It is important to emphasise that this work does not allow us to make inferences as to what statistical regularities the human visual system might use to achieve illuminant perception, it merely allows us to put boundaries on what regularities are present in the scene and available to the human visual system. It is also worth noting that if statistical regularities exist in the data, the classifier will learn to detect them. Therefore, one can be quite confident that there are not additional regularities beyond those the classifier has uncovered: if there were additional regularities the classifier would have learnt those too. One may thus think of the classifier as the "optimal" case, rather than as a model of a real human observer. It is possible however that the full extent of the regularities used by the classifier has not been completely characterised by the analysis presented here. Such additional analyses are beyond the scope of the work presented here, though highlighted below as avenues for future work.

3.7.1 Connection to human illuminant perception

One obvious avenue for future work is to make connections between classifier performance and the human ability to estimate the illuminant in a scene. As explained above, one cannot make direct inferences about how human observers perform this task based on the results of the classifier. The classifier instead allows us to speculate as to what statistical regularities the human observer may be able to utilise to best achieve illuminant perception. It thus allows us to identify clear research questions that can then be tested in the human observer.

One such question is whether rod or melanopsin signals feed into illuminant percepts. The results presented here show that, hypothetically, access to such a signal would be advantageous to illuminant perception, particularly for dichromats. To test whether this is the case for human observers, one

could use the MPHDR display (Hexley, Yöntem et al., 2020), as described in Chapter 5, to silence the melanopsin and rod signals while keeping the cone signals constant, and assess the impact this has on illuminant discrimination in human observers. In particular, one could recruit both normal trichromats and dichromats to assess whether the presence of the rod or melanopsin signal offers a greater performance advantage to dichromats over normal trichromats, as predicted here, if at all. Such an experiment would feed into the broader area of literature on whether and how melanopsin and rod signals contribute to visual function that was traditionally thought to be due to cone signalling alone.

3.7.2 Limitations

While the results presented here offer a significant step towards quantifying the limits that can be placed on illuminant classification, there are of course limitations to the current approach. While we are able to place boundaries on the statistical regularities found in the surfaces we use to model the real-world, the SOCS dataset, while arguably the most comprehensive set of surface reflectance data currently available, is not a perfect representation of the real-world. Crucially, the SOCS dataset does not consider the frequency with which we are likely to encounter certain surfaces - something that would be informative in a prior for human illuminant perception. We lack an understanding of the "spectral diet" of the human observer (Webler et al., 2019) that we would ideally use as our real-world dataset here.

There is additionally a limitation in choosing the appropriate way to split the data into train and test sets. Here, to avoid overfitting the classifier, I reserve 20% of the surfaces for the test set, meaning the classifier has never seen the surfaces on which its performance is evaluated. However, it is likely the case for human illuminant perception in the real-world that the observer will have seen the surfaces before, as the external world we interact with is familiar to us over time. Thus it may be more appropriate to allow the classifier to see all of the surfaces from the test data during training, though it is not obvious how to do this while avoiding overfitting. One way may be to show each surface under seven of the eight illuminants during training, though this still does not fully capture the experience most observers are likely to have: namely that at some point they have interacted with the majority of surfaces from the real-world under the majority of real-world illuminants.

Finally, the surface samples analysed here as a first demonstration of the potential of this approach were simple, flat, matte surfaces; yet many surfaces in the real-world have texture, shape, and shading. The "scene" modelled here is the simplest case, and does not include an account of these surface

properties or the inter-reflections that can occur when one has multiple surfaces in a scene (Morimoto et al., 2019). Thus, there is scope for future work is to use this approach to assess what textures, highlights, and shape of surfaces are most informative to illuminant classification. Additionally, there is often more than one illuminant seen in a scene. Thus, another avenue for future work would be to test how well the classifier performs when it has to judge multiple illuminants, rather than just one.

3.7.3 Further directions with the machine learning approach

The results presented here are really the beginning of the analysis that could be done using a machine learning approach to place boundaries on the statistical regularities in photoreceptor signals from surfaces for illuminant estimation. A major contribution of this chapter is in providing a new tool to the vision science community to address the problem of illuminant classification. I outline a few additional promising strands of future research using the approach presented in this chapter in the paragraphs below.

One obvious line for future work is to consider more complex strategies for combining multiple surfaces. For example, there may be some sort of confidence threshold placed on the probabilities associated with each surface, such that one surface that may provide a particularly strong cue to the illuminant is given additional weighting in the averaging over surfaces than a surface that assigns nearly equal probabilities to each class. One way in which this might be realised is to have the classifier identify both the surface and the illuminant. If the classifier is able to correctly classify a surface into a specific category, this might help the classifier narrow-in on what region of the gamut of possible surfaces the test surfaces falls in, and thus how that surface is likely to change under different illuminants. This would be an implementation of a "memory colour" type approach in the classifier (Witzel et al., 2018). More broadly, it would be interesting to test how well the classifier can identify the surface, rather than the illuminant, chromaticity, if trained on surfaces labels as opposed to illuminant labels. The difficulty with such an implementation is choosing how to discretise the expansive set of surface chromaticities in the real-world.

The machine learning approach used here can also be used to assess whether or not the photoreceptor spectral sensitivities are optimised for the task of illuminant discrimination in the real-world. Instead of using the spectral sensitivity functions of the human photoreceptors one could use any arbitrary set of spectral sensitivity functions to reduce the radiance spectra to a few scalar signals and address how classifier performance changes when an arbitrary set of spectral sensitivities are used. One could also feed in the signals accessible to the post-receptoral pathways, i.e. L+M, L/(L+M), and

S/(L+M), as opposed to just the pre-opponent cone photoreceptor signals to see whether the same statistical regularities can be extracted from the post-receptoral pathways as is available to them.

3.8 Conclusion

In this chapter, I describe the results of using a random forest classifier to perform an illuminant classification task using only the cone photoreceptor signals from simple surfaces. The classifier performs above chance, and typically correctly identifies either the correct illuminant or an illuminant that gives rise to a perceptually indistinguishable scene. By probing the successes and failures of the classifier under different conditions, such as varying the illuminant set, varying the gamut of surfaces, varying the number of surfaces in the scene, and varying the availability of the photoreceptor signals, I am able to begin quantifying the limits that can be placed on illuminant estimation from photoreceptor signals and characterise the statistical regularities in a real-world dataset that may be exploited by the human observer to achieve illuminant perception.

3.9 Acknowledgements

The work presented in this chapter was a collaborative project with Dr Takuma Morimoto (University of Oxford and Justus Liebig Universität Giessen), Prof. Keiji Uchikawa (Kanagawa Institute of Technology), and Prof. Hannah Smithson (University of Oxford). The collaborative project was led by ACH. The results presented here were presented by ACH at the 43rd European Conference on Visual Perception (August 2021), for which ACH was awarded the Colour Group (GB) WD Wright Award. The results have also been presented by ACH at the Colour Group (GB) Student Awards Meeting (December 2021). I would thus like to thank all of my collaborators on this project for their contributions to the analysis and interpretation of the results. I would also like to thank Dr Danny Garside (Postdoctoral Visiting Fellow at National Institutes of Health (NIH)) for his thoughtful comments on this project following my presentation of this work at ECVP.

Chapter 4

Online visual psychophysics: is it feasible to study photoreceptor signals with online experiments?

Contents

4.1 Introduction 138

4.2 The stages of display calibration and their impact on stimulus control 140

4.2.1 Linearity of display output with input drive-signal 140

4.2.2 Chromaticity of stimuli 144

4.2.3 Illumination levels 152

4.2.4 Spatial characteristics 154

4.2.5 Timing 158

4.2.6 Further Considerations 160

4.2.7 Display calibration checklist 161

4.3 Review of existing methods of perceptual calibration 162

4.3.1 Linearity of display output with input drive 163

4.3.2 Chromaticity of stimuli 169

4.3.3 Illumination levels 174

4.3.4 Spatial characteristics 176

4.3.5 Timing 178

4.3.6 Further Considerations 179

4.3.7 Perceptual Calibration Checklist 180

4.4 Discussion 181

4.5 Conclusion 184

4.6 Acknowledgements 184

4.1 Introduction

With the closure of many labs around the world and a pause on traditional "in-person" human participant research during the COVID-19 global pandemic (March 2020-ongoing), many visual psychophysicists and cognitive scientists have sought solutions to move their experiments online. While the move towards online or remote testing has been brought to the forefront of many researchers minds as a result of the COVID-19 situation, the idea of online or remote psychophysics testing is not new. Further, the advantages of online testing go beyond the ability to test participants when access to the lab is restricted. For instance, online testing offers the possibility of larger participant sample sizes than is realistically achievable in a lab setting, a more diverse sample of participants (who would hopefully be more representative of the population being studied) than those who are typically recruited for in-person experiments, and more rapid data collection than is possible in the lab as many participants can be tested in parallel without the need for researcher intervention (Woods et al., 2015).

While the prospect of online testing offers many exciting opportunities for data collection beyond the scope of what is possible in a lab environment, online testing has its disadvantages (Woods et al., 2015). One challenge of online testing is compliance. A lack of interaction between the participant and the experimenter, presents a greater risk of participants misunderstanding instructions or not paying attention to the task. Furthermore, it is possible that the same participant may try to complete the remote experiment multiple times, especially if a reward is offered for completing the study, and due to the complication of ethics and data protection (which in themselves are difficult to implement for remote testing), it can be difficult to ensure data are coming from unique participants. Arguably though, the biggest difficulty facing online testing for vision experiments is the challenge of remote display calibration. In this chapter I explore the techniques for calibrating displays for online visual psychophysics, and whether or not such calibration is sufficient for experiments that assess photoreceptor signals to be carried out remotely.

It is worth remembering that calibration is an issue for both in-person experiments and online experiments. Many standard techniques, such spectroradiometric or colorimetric measurements,

have long been common practice for calibrating lab-based displays. Methods for calibrating displays remotely, where one does not have access to specialized and expensive equipment for calibration, however, are less standardized and often subject to more uncertainty when pitted against accepted "gold standard" instrument-based calibration methods. Historically, this has made many visual psychophysicists wary of moving their experiments online, out of the traditional, controlled in-person environment to the less controlled remote setting.

It is perhaps, worth reframing the discussion away from in-person vs remote display calibration, and refocusing on the goal of display calibration, which is the accurate presentation of visual stimuli. While remote display calibration methods may never be able to achieve the gold standard calibration that can be achieved in the lab, if one understands what aspects of a display can be calibrated remotely, and more importantly what features of a stimulus can therefore be controlled remotely, one can begin to get a handle on the scope of stimuli that can be accurately presented in online testing. It is therefore perhaps better to think about designing experiments *for* online testing, within the realms of what is possible for stimulus control, rather than trying to adapt an in-person experiment to work online.

To understand the scope of experiments that are possible online, one needs to understand two things: 1) the stages of display calibration, and what each stage is doing in terms of precise stimulus delivery, and 2) which stages of display calibration can be achieved remotely. I address these two topics in this chapter. In the first half of this chapter, I present an overview of the various stages of display calibration that one needs to consider for any vision science experiment, both online and in-person. I discuss the various implications each stage can have on the reproduction of visual stimuli, in terms of contrast, luminance, chromaticity, spatial characteristics, and timing of the stimulus, and holistically what that means for conducting experiments to understand photoreceptor inputs to vision. In the second section of this chapter, I discuss the existing methods for remote display calibration, and discuss possible extensions of existing techniques. These techniques typically rely on exploiting properties of the human visual system to calibrate displays, and thus I term these methods perceptual calibration. I conclude by discussing the potential applicability and shortcomings of current methods of remote display calibration to the application of online, remote experiments using visual stimuli, and discuss what type of experiments might or might not be feasible to conduct online, with a particular focus on the types of experiments discussed in the rest of this thesis: those that assess the photoreceptor signals and their consequences for vision.

4.2 The stages of display calibration and their impact on stimulus control

All visual stimuli are subject to distortion when presented on an uncalibrated display and the precision of calibration required depends on the experiment. For instance, in a higher-level visual cognition experiment using complex image stimuli, it may not be necessary to ensure the colour of the images is reproduced exactly across two different displays, though it may be important to ensure the stimulus timing is consistent. For lower-level contrast threshold measurements, precise reproduction across all stimulus characteristics and thus precise display calibration at all stages may be essential. Regardless of the level of calibration that is necessary, it is critical that the experimenter is aware of how their stimuli will be transformed if presented on an uncalibrated display.

In this section of the chapter, I discuss the many aspects that must be considered when calibrating a display system for stimulus presentation and demonstrate the problems that can occur if one overlooks certain important factors. I provide a summary at the end of the section with a bullet-point display calibration checklist, containing the key points discussed.

4.2.1 Linearity of display output with input drive-signal

The first aspect of display calibration one must consider is the non-linear relationship between driving input, v, to display luminance output, L. Typically, this non-linearity is described by a power law, typically referred to as a gamma, γ, function, as in Eq. 4.1:

$$L(v) = L(v_{max})\left(\frac{v}{v_{max}}\right)^{\gamma}. \tag{4.1}$$

The gamma function is a specific case of an electro-optical transfer function (EOTF) (Mantiuk, 2016). The gamma function describes the EOTF of cathode-ray tube (CRT) displays and remains a rough approximation of the EOTF of many liquid-crystal displays (LCDs) and organic light emitting diode (OLED) displays. This is because LCD and OLED display driver electronics have been designed to make the response of LCDs and OLEDs mimic the response of CRTs to ensure backwards compatibility, despite the differences in the underlying physics driving the different display technologies (Mantiuk, 2016). Yet, despite attempts to make the response of LCDs mirror the response of CRTs, the EOTF of LCDs may be better approximated with an S-shaped curve than with a gamma function (Glasser, 1997; Kwak & MacDonald, 2000).

One can use the inverse of a display's EOTF to linearize display output with input drive signal. In the case where the EOTF is described by a gamma function, this is often referred to as gamma

correction. If one fails to linearize a display, one will fail to reproduce (i) the desired contrast of an image or stimulus, (ii) the desired mean luminance of the stimulus, and (iii) alter the spatial frequency spectrum, as discussed below. Additionally, non-linearities will distort colour transformations, as outlined in Section 2.2.

One may characterize the EOTF of each channel of the display (R, G, B) independently, or characterize the EOTF of the grey-levels. One way to characterize the EOTF of a display is to generate a look-up table (LUT) between the display input values and the luminance output, which does not need to assume an underlying functional form of the EOTF. If using a photometer, it is possible to measure the output luminance for each input value and produce a direct one-to-one mapping of display input values and output luminance. One could use interpolation or curve fitting to reduce the influence of measurement noise and refine the LUT. Interpolation and curve fitting can also reduce the number of measurements needed to generate the EOTF of the display. In the case where the functional form of the EOTF is known, such as for a CRT, one may only need a small number of measurements to fit the EOTF. One caveat is that for these calibration methods, the EOTF must remain stable and reproducible between measurements. Whilst most commercial displays are stable, this may not be true whilst they warm up and for an initial period after powering up the display, or under conditions of fluctuations in supply voltage or temperature.

Consequences of non-linearity

What is the error associated with ignoring the non-linearity of the display, or incorrectly estimating this non-linearity? To illustrate these errors, the influence of non-linearity on stimulus profiles is simulated, using the Gabor patch as an example case, with an intensity profile I, which consists of a sinusoidal grating with horizontal and vertical spatial coordinates, x, y, spatial frequency f, contrast c, phase ϕ, and rotation θ, seen through a Gaussian contrast window, with standard deviation σ, as in Eq. 4.2:

$$I(x,y) = 1 + c\cos(2\pi f[y\sin\theta + x\cos\theta] + \phi)e^{-\frac{x^2+y^2}{2\sigma^2}}. \tag{4.2}$$

It is first assumed that the display in question can be perfectly described by an EOTF with a simple gamma function with a gamma value of 2.2. Then, a gamma correction is applied to a Gabor image based on four different assumed gamma values: (a) 2.2, (b) 1.0, (c) 2.0 and (d) 2.4 which are chosen to express (a) perfect correction, (b) no correction, (c) under-correction and (d) over-correction, respectively.

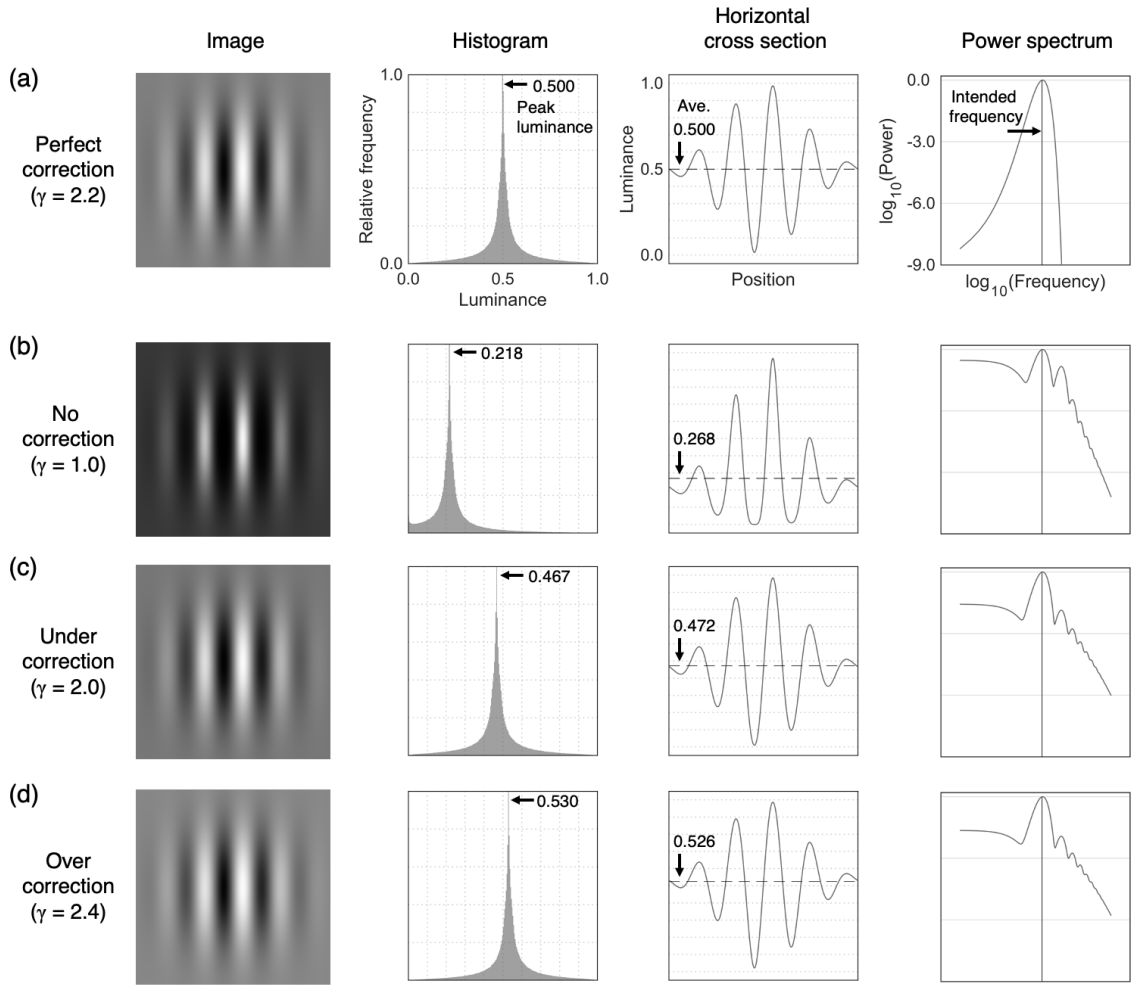


Figure 4.1: Simulation of the effects of misestimating the monitor gamma value on stimulus properties of a Gabor patch. (a) The case where perfect gamma correction ($\gamma = 2.2$) was performed. Sub-panels indicate from left to right the Gabor image, the histogram of output pixel intensities of the image, horizontal cross section at the vertical centre of the image, and power spectrum obtained by Fourier transformation (FFT). (b-d) The case of no gamma correction, under-correction and over-correction, respectively which introduce undesired artifacts to luminance profiles and to frequency components. The whole image comprised 500 × 500 pixels. A sine grating component was set to have the maximum contrast possible (i.e. mean and amplitude were both set to 0.5), a frequency of 5 cycles per image and presented in sine phase with respect to the centre of the Gaussian. The Gaussian component peaked at the image centre, and its standard deviation was set to 100 pixels. A sufficiently large image (not shown here) was used so that intensity modulation converges to the level of DC component at the image edge. Note, the image of the Gabor shown in the left column may not be an accurate reproduction depending on the reproduction process of the image on a particular display or in print. Also, note that the DC component was subtracted from the original image before FFT. The *fft2* and *fftshift* functions provided in MATLAB R2020a (Mathworks, Natick, MA, USA) were used to implement the FFT.

Fig. 4.1 shows the simulation results. Fig. 4.1(a) indicates the case where perfect gamma correction was performed, which thus replicates the ground-truth properties of the desired Gabor image. From left to right, sub-panels show (i) the image of the Gabor patch, (ii) the histogram of output intensity values, (iii) the intensity profile of a horizontal cross section at the vertical image centre and (iv) the power spectrum of the horizontal cross section of the image computed via fast Fourier transformation (FFT).

Fig. 4.1(b) indicates the case of ignoring linearization entirely, which thus introduces the largest errors within the tested gamma values. The image of the Gabor patch appears much darker than the original image, the peak of the intensity histogram shifts towards lower values, and the horizontal intensity variation is somewhat distorted, lowering the average value. Moreover, the power spectrum clearly indicates that frequency components that are not present in the original image are introduced over a wide frequency range.

Fig. 4.1(c),(d) shows the results of under- and over-correction, which result in a less apparent distortion than no correction, as expected. The direction of shift is opposite for under- and over-correction as shown in the histograms and horizontal cross-sections. The right most panel shows the various unintended frequency components present in the image, despite the relatively small error in the gamma value of only 0.2.

It is worth noting that the effect of incorrect gamma estimation depends on the intensity values within the stimulus. The shape of the gamma curve tells us that high and low intensity values are less affected than the middle range of intensities. Thus, in an extreme case where only very high or low intensity values are used, incorrect estimation of the EOTF may not lead to as severe distortions to the stimulus (or in the case where only black and white are used, no distortions). Unless the range of intensities used in a stimulus is actively limited, one must be sure to estimate the non-linearity correctly.

The simple simulation above demonstrates the crucial role that linearization plays in accurate stimulus presentation in the case of the Gabor patch. Failure to apply the correct linearization introduces mismatches between the desired and displayed intensity profiles in addition to artifacts such as the introduction of unintended frequency components. Such errors should also be anticipated with more complex images.

Additionally, the chromaticity transformation of the display, as discussed in Section 2.2, relies on the display being linearized, and thus failures in linearization also result in failures in chromaticity reproduction.

Dynamic range reduction

It is worth noting that an unfortunate consequence of display non-linearity is an inherent reduction in the effective dynamic range of the display. For instance, consider a display where the first 20 input values, 0-20, all map onto the same measurable luminance. One cannot code any meaningful information using those first 20 input values as they all compress to a single light output level. Even

though most displays may have 8-bit control of display input values (i.e. 256 input values), after linearization it becomes apparent that some displays may have a lower output range (e.g. 6-bit) than the input control would suggest (To et al., 2013). Often display drivers with greater than 8-bit control such as the Cambridge Research Systems VisaGe MKII Stimulus Generator that offers 14-bit colour and luminance control (Cambridge Research Systems, Rochester, UK) are used for psychophysical experiments in lab settings to increase the dynamic range of the display output. Some commercial displays offer 10-bit resolution, but these are rare, and also require specialized drivers and image formats.

Saturating and cut-off non-linearities

Saturation non-linearity at the highest output levels and cut-off non-linearity at the lowest output levels are specific aspects of linearization, which while captured in an EOTF, are often worth considering prior to linearization (To et al., 2013). The black level, i.e. the display output when the display is driven at the 0 setting, is never actually perfectly black, i.e. the output is not 0 cd/m^2. At the minimum driving input value displays will output a low, but non-zero, luminance. This means the simple gamma function model, or S-shaped curve model, used to describe the EOTF of displays must be modified to include this offset. The black level offset can reduce the contrast of an image from the desired contrast, as specified assuming a zero black-level.

For LCDs there may be an additional non-linearity in the output luminance between the 254 setting and 255 setting on an 8-bit display. At the highest input value of 255, all of the liquid-crystal cells are switched "off" to allow full transmission of the LCD backlight, which is a different physical arrangement of the transmissive element than at any other input value, often resulting in a jump in the luminance output between 254 and 255 (To et al., 2013).

4.2.2 Chromaticity of stimuli

Another crucial aspect of display calibration is colorimetry. In fact, the traditional model of CRTs developed by Berns, 1996, suggests a two-stage calibration: linearization and colorimetric calibration. Colorimetric calibration refers to the process of transforming a display's native and device-dependent RGB colour space to a device-independent colour space, such as CIE 1931 XYZ (CIE, 1932).

Images are rendered as RGB triplets. The stimuli associated with particular RGB coordinates, however, are device-dependent and thus RGB coordinates are not appropriate for describing colour between different displays. The CIE 1931 XYZ colour space was developed by the *International*

Commission on Illumination (CIE) as the first device-independent colour space, which acts as a standard, reference colour space, that does not depend on the characteristics of the display system on which colour is being rendered (CIE, 1932). For complete colorimetric calibration one needs to calculate the RGB to CIE 1931 XYZ transformation matrix for a display to accurately reproduce a colour of a known CIE 1931 XYZ on the display. I note that for a lot of visual psychophysics one may prefer LMS coordinates over XYZ as an independent colour space, but here I describe the method for XYZ as, without giving away the conclusion to this chapter, there is no way to approximate the RGB->LMS transformation as there is the RGB->XYZ.

In order to determine the transformation matrix that maps RGB to CIE 1931 XYZ, one only needs to know the chromaticities (CIE 1931 xy coordinates) of each primary and the luminance of each primary (CIE 1931 Y coordinate, specified in $\mathrm{cd/m^2}$) (as described in full in Chapter 1). If one has access to a spectroradiometer or other colorimetric device, it is easy to measure the CIE xy chromaticities of the primaries and luminances of the primaries.

When the CIE 1931 XYZ coordinates of the primaries are known, one can then calculate the XYZ coordinate of any arbitrary RGB triplet (assuming linearity correction as discussed in the previous section has already been applied) as in Eq. 4.3:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = M \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \tag{4.3}$$

where

$$M = \begin{bmatrix} X_R & X_G & X_B \\ Y_R & Y_G & Y_B \\ Z_R & Z_G & Z_B \end{bmatrix}.$$

One can then calculate the RGB triplet needed to reproduce a specific CIE 1931 XYZ tristimulus value on a display using the inverse of matrix M, as in Eq. 4.4:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = M^{-1} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \tag{4.4}$$

It is worth noting that there are two ways in which one can consider the absolute luminance of the display in chromaticity reproduction. One can either consider the CIE 1931 xyY triplet, where Y

is the absolute luminance, or one can consider a relative CIE 1931 xyY triplet, where Y of the white point is one, and the absolute luminance, Y_{peak}, is factored out of the above equations (as the absolute luminance is simply a scale factor in the conversion between CIE 1931 xyY and CIE 1931 XYZ). If one was to use the relative CIE 1931 xyY coordinates, and factor absolute luminance out, then one can convert between RGB and CIE 1931 XYZ as in Eq. 4.5 and Eq. 4.6:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = Y_{peak} \cdot M \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{4.5}$$

and

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \frac{1}{Y_{peak}} \cdot M^{-1} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{4.6}$$

where

$$M = \begin{bmatrix} X_R^{\text{rel}} & X_G^{\text{rel}} & X_B^{\text{rel}} \\ Y_R^{\text{rel}} & Y_G^{\text{rel}} & Y_B^{\text{rel}} \\ Z_R^{\text{rel}} & Z_G^{\text{rel}} & Z_B^{\text{rel}} \end{bmatrix}.$$

If one cannot determine the CIE 1931 XYZ coordinates of the primaries of a display, then one can only derive an approximate device-specific chromaticity transformation matrix.

sRGB (standard RGB) colour space provides an approximate solution in the case where one cannot measure the CIE xyY values of the primaries of a display (International Electrotechnical Commission, 1999). Designed in a collaborative effort between Microsoft and Hewlett-Packard in 1996, sRGB colour space is an attempt to standardise colorimetry across displays by defining a universal native RGB coordinate system. Many manufacturers opt to produce displays that span the sRGB gamut. The sRGB gamut is defined by specific CIE 1931 xyY coordinates for each primary. The idea behind sRGB colour space is that a standard CIE 1931 XYZ to RGB chromaticity transformation matrix can be applied to all sRGB displays. However, as many different primary combinations are used to produce sRGB displays, in practice one often finds displays are almost sRGB, but with slight deviations from the specified standard (see Section 2.2.3).

For illustrative purposes, I ask what errors would be introduced if one incorrectly assumes that a display conforms exactly to the sRGB chromaticity transformation and render to the display us-

ing the sRGB chromaticity transformation matrix, when the display was not a perfect sRGB display. For the following simulations I investigate the colour reproduction of the 24 colour samples of the MacBeth ColourChecker under Illuminant C (McCamy et al., 1976; Poynton, 2008). The CIE 1931 xy chromaticities of these samples are depicted in Fig. 4.2(a), and I refer to these 24 colour samples as the ColourChecker samples. I model the colour distortions produced on displays that are almost, but not perfect, sRGB displays, and for which the CIE 1931 XYZ coordinates of the primaries are known. I render the ColourChecker samples onto a particular (either a hypothetical or example) display using the CIE 1931 XYZ to sRGB transformation matrix to calculate the sRGB triplet for each colour sample. I then calculate the actual CIE 1931 XYZ coordinates realized on the display when driven with the sRGB triplet, as one can derive the actual native RGB to CIE 1931 XYZ transformation for the display in question using the known primaries. I take the error measure, which I call the colour reproduction error, to be the CIE 2000 Colour Difference, ΔE, between CIE 1931 XYZ coordinates of the ColourChecker samples and the CIE 1931 XYZ coordinates rendered on the display, transformed to CIELAB coordinates for the ΔE calculation, where the white point is defined using the CIE XYZ coordinates of the white square on the MacBeth colour checker under Illuminant C (X=88.3, Y=90, Z=106.5). A ΔE below 1 is typically considered to be below the JND for a human observer, while a ΔE of around 5 or below is often considered an acceptable colour error for commercial applications (Sharma & Bala, 2017).

In this simulation, I ignore the effect of absolute luminance, and factor out the absolute luminance as in Eq. 4.5. For the hypothetical displays I consider two alternative perturbations from the sRGB standard A) the case where the relative luminances of the primaries deviates from the specified standard and B) the case where the CIE xy gamut is contracted or expanded compared to the specified standard. Finally, in case C, I also ask how common such perturbations may be in practice, and examine the colour differences that would arise if sRGB colour space was used to render stimuli on a set of eight common displays for which I have full colorimetric specification. The displays I used were the MacBook Pro 2009, MacBook Air 2012, NEC 2490WUXI, NEC PA272W (in its sRGB, not Adobe RGB, mode), Retina MacBook Pro 2014, Surface Pro 3, ThinkPad T440s, and iPad Pro 2015. The colorimetric data for all of these displays was obtained from f.luxometer ('f.lux Software LLC', 2011-2015).

A) Chromaticities of the primaries

sRGB colour space defines the CIE 1931 xy chromaticity coordinates of each of the primaries, as:

$$R : x = 0.6400; y = 0.3300,$$

$$G : x = 0.3000; y = 0.6000,$$

$$B : x = 0.1500; y = 0.0600.$$

In practice, most displays have a gamut that covers almost the same area in CIE 1931 xy chromaticity space as the sRGB gamut, but which deviates slightly from the sRGB gamut. In this simulation, I consider only the case where the gamut expands or contracts around its white point. I define relative gamut size as the percent shrinkage/expansion of the distance between each primary and the white point, and simulate the colour reproduction error as a function of relative gamut size. Fig. 4.2(b) shows an example of a hypothetical display gamut in CIE 1931 xy chromaticity space which has been contracted or expanded by 5% compared to the sRGB gamut. Fig. 4.2(c) shows the colour reproduction error as a function of relative gamut size across the 24 MacBeth ColourChecker samples, and one can see that relatively small contractions or expansions can result in relatively large colour differences.

B) Relative luminance of the primaries

sRGB colour space defines the luminance ratio, that is the relative CIE 1931 Y coordinate, across the primaries as:

$$R : 21.26\%$$

$$G : 71.52\%$$

$$B : 7.22\%.$$

As with the chromaticities of the primaries, the relative luminance of the primaries deviates from this sRGB specified ratio in practice. Fig. 4.2(d) shows the colour reproduction error as a function of the percentage perturbation from the sRGB standard luminance contribution for the (i) R, (ii) G, and (iii) B primaries for a hypothetical display. When varying the ratio of one primary, I assume the other two primaries maintain the same relative ratio to each other. One can see that surprisingly small changes in the percentage contribution to the total luminance for each primary can result in relatively

large errors in colour reproduction.

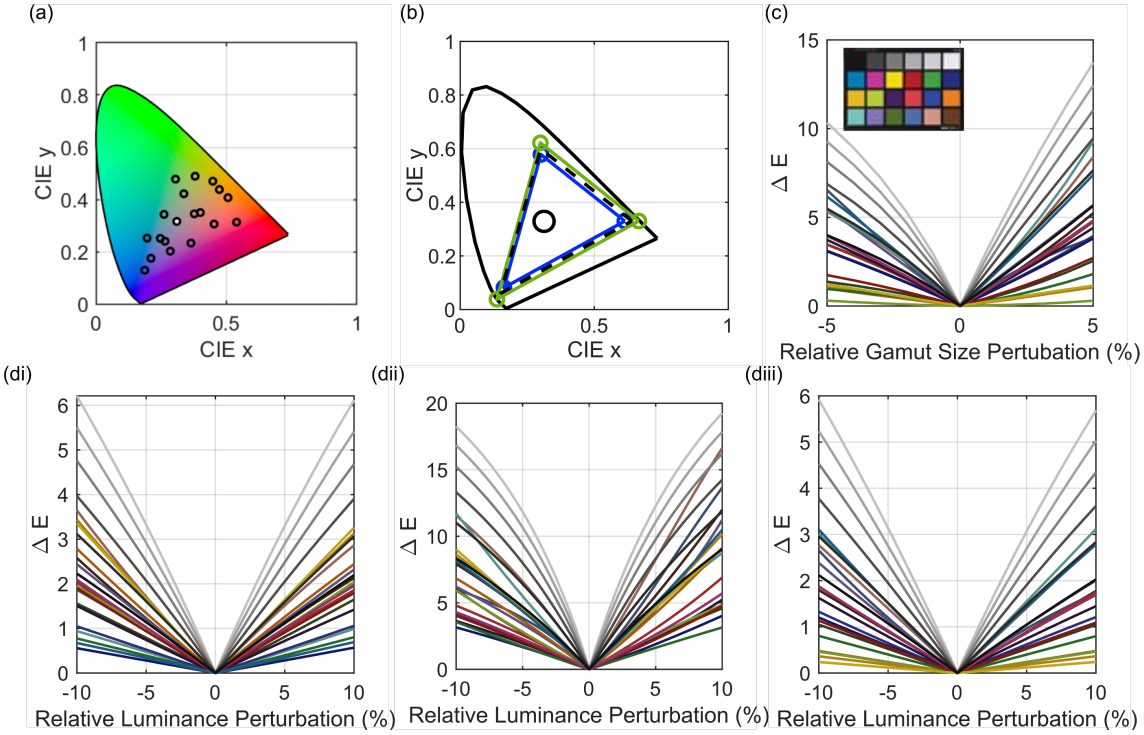


Figure 4.2: Simulation of the effects of ignoring colorimetric calibration and rendering to a hypothetical display assuming it is a perfect sRGB display, when the display in fact differs from this standard. (a) Plot of the 24 MacBeth ColourChecker samples in the CIE 1931 xy chromaticity diagram. (b) Simulation of the effects of the CIE xy chromaticity coordinates of each of the primaries varying in distance from the sRGB specified white point (x=0.3127, y=0.3290) compared to the sRGB standard. The CIE xy chromaticity gamuts of the sRGB display (black dotted line) is shown compared to a gamut where the distance between the primaries and the white point is reduced by 5% (blue), and increased by 5% (green). (c) The colour reproduction error, ΔE, of the 24 MacBeth ColourChecker samples, as a function of the relative gamut size, defined as the percentage difference in the distance from the primaries to the white point compared to the sRGB gamut, where the white point is fixed. The colour reproduction error for each of the samples is depicted separately, where the colours of the lines on the plot represent the colour of the samples on the MacBeth ColourChecker (depicted in the inset). (d) Simulation of the effects of the relative luminance of each of the primaries for some hypothetical display varying compared to the specified luminance ratio in the sRGB standard (such that the percentage perturbation is of the percentage contribution from that primary in the standard, e.g. a 5% increase for G reflects a change from 71.52% to $1.05 \times 71.52\% = 75.10\%$). The colour reproduction error, ΔE, of 24 the ColourChecker samples, as a function of the percentage difference in the relative luminance contribution of each primary (i) R, (ii) G, and (iii) B from the sRGB specified luminance contribution of each primary.

C) Colour reproduction error in common displays

The CIE 1931 xy chromaticity gamuts of the eight example displays compared to the sRGB standard gamut, are depicted in Fig. 4.3(a). One can see that the CIE 1931 xy gamuts of the displays overlap with each other, and in most cases, with the sRGB gamut, but do show slight deviations. When plotted in CIELAB coordinates, Fig. 4.3(b), for a more perceptually uniform representation, one can see the relative chromaticities of the primaries of the example displays deviate quite noticeably from each other. In addition to deviations in the chromaticity coordinates from sRGB, the eight example displays

also show deviations in the relative luminance ratios of the primaries, as shown in Fig. 4.3(c). I simulate what the colour reproduction error for the ColourChecker samples would be on eight common displays. Note, the colour reproduction here encompasses deviations in both the chromaticities of the primaries and the relative luminances of the primaries compared to the sRGB standard.

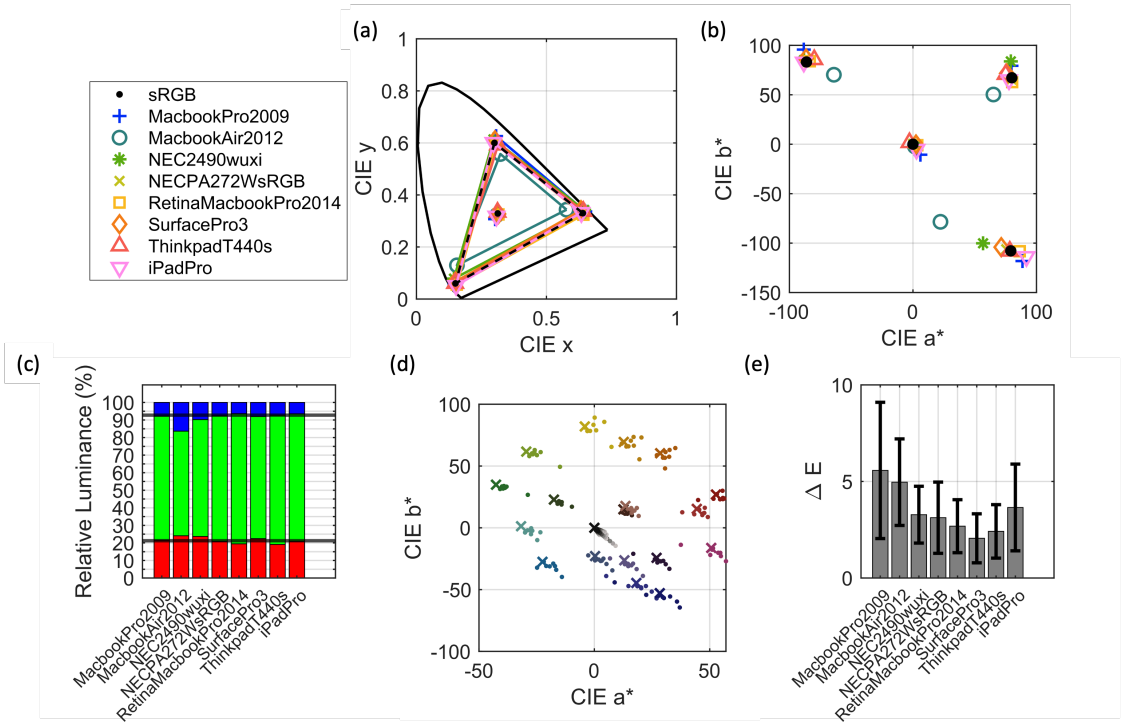


Figure 4.3: Simulation of the effects of ignoring colorimetric calibration and rendering to a real example display assuming it is a perfect sRGB display, when the display in fact differs from this standard. (a) The CIE 1931 xy chromaticity gamuts of the eight example displays compared to the sRGB CIE 1931 xy gamut. (b) The CIELAB coordinates of each of the primaries of the eight examples displays compared to the CIELAB coordinates of the primaries of an sRGB display, where the white point used to calculate the CIELAB coordinates is the white sample of the ColourChecker under the reference illuminant (Illuminant C). (c) The relative luminance ratio seen between the primaries across the eight example common displays. The two black horizontal lines indicate the relative luminance ratio in the sRGB specifcation. (d) The CIE $a^*, b*$ coordinates of the 24 MacBeth ColourChecker samples as rendered across the eight example displays (dots) compared to rendering on a perfect sRGB display (crosses). (e) The colour reproduction error, specified as the mean colour difference, ΔE of ColourChecker samples, with error bars showing the standard deviation of the colour difference across the ColourChecker samples, for the eight example displays.

First, I visualize the variance in the CIELAB coordinates of each of the ColourChecker samples when rendered on these example displays. Fig. 4.3(d) shows the CIE a^*, b^* coordinates of each sample of the ColourChecker under the reference illuminant (Illuminant C) along with the CIE a^*, b^* coordinates of the reproduced samples across the eight displays. I then summarize the colour differences across the displays by plotting the colour reproduction error, along with error bars indicating the standard deviation in the ΔE colour difference across the 24 samples of the MacBeth ColourChecker, for each example display in Fig. 4.3(e). There is significant variance in the colour reproduction error across common displays, with some displays showing relatively smaller average colour reproduction errors

while other displays have quite large colour reproduction errors. The error seen across these displays is well within the limits of what is considered acceptable for commercial applications. For experiments requiring subjective specification of colour the sRGB model is a reasonable approximation, but for experiments requiring precise colour specification, the sRGB model is likely to fall short.

It is worth emphasising that this colour reproduction error is not an inherent problem with the displays, but rather with using the sRGB standard as a calibration approximation for the displays. Many low-end consumer displays will simply use the sRGB model as the on-board colour management for the display, and thus would be susceptible to the errors described above, while some high-end consumer displays come equipped with more complex colour management profiles, such as an ICC profile (International Color Consortium), which uses the specific chromaticity transformation for that display derived by colorimetric measurements for colour management. However, even in the case of displays with sophisticated colour management, the chromaticity transformation matrix is kept "under the hood" and is not accessible to the end-user.

Colorimetric calibration is not (yet) physiologically relevant

I have shown the colour differences that can arise as a result of chromatic differences between the primaries of a display and the sRGB standard. So far, I have considered chromaticity only in terms of CIE 1931 XYZ (or transformations thereof e.g. CIELAB) coordinates, as this is the device-independent colour space typically used as a reference for displays. However, the CIE 1931 XYZ colour space does not have a physiologically relevant basis. Thus, one is unable to determine the physiologically relevant cone responses of a particular RGB setting on an arbitrary display, even if one can calculate the device independent CIE 1931 XYZ coordinates, as there is no linear transformation between the CIE 1931 XYZ colour matching functions and the CIE 2006 physiologically-relevant LMS cone fundamentals (Stockman et al., 1999; Stockman & Sharpe, 2000; CIE, 2006; Stockman, 2019). Therefore if one is interested in colour reproduction in terms of the CIE 2006 LMS responses, i.e. for using displays to test the human photoreceptor signals, then a colorimetric calibration to convert device-dependent RGB space to device-independent CIE 1931 XYZ space is insufficient. One would instead need to know the full spectrum of each of the primaries to calculate the LMS to RGB transformation matrix, or one could use the new physiologically-relevant CIE 2006 XYZ functions, which are defined to be linear transforms of the cone fundamentals, as a physiologically-relevant device-independent chromaticity space (CIE, 2015). At the time of writing this thesis the CIE 2006 XYZ functions have not yet been widely adopted by the display industry (and most, if not all, colorimetric devices still operate using the CIE 1931 XYZ

functions).

4.2.3 Illumination levels

In addition to knowing the relative luminance levels of each of the primaries of a display for colorimetric calibration, one also needs to know the absolute luminance level of the display. In this section I discuss the impacts of the absolute illumination levels, of both the display and ambient light, on display calibration.

Absolute luminance level

As described in the section on linearization, one needs to know the minimum absolute luminance output of a display to characterise the black level offset. The absolute, or peak, luminance output of a display can vary quite widely. Across the eight example displays I consider, I find the peak luminance output varies between 175.9 cd/m^2 to 390 cd/m^2 with a mean peak luminance of 265.7 cd/m^2, as shown in Fig. 4.4(a).

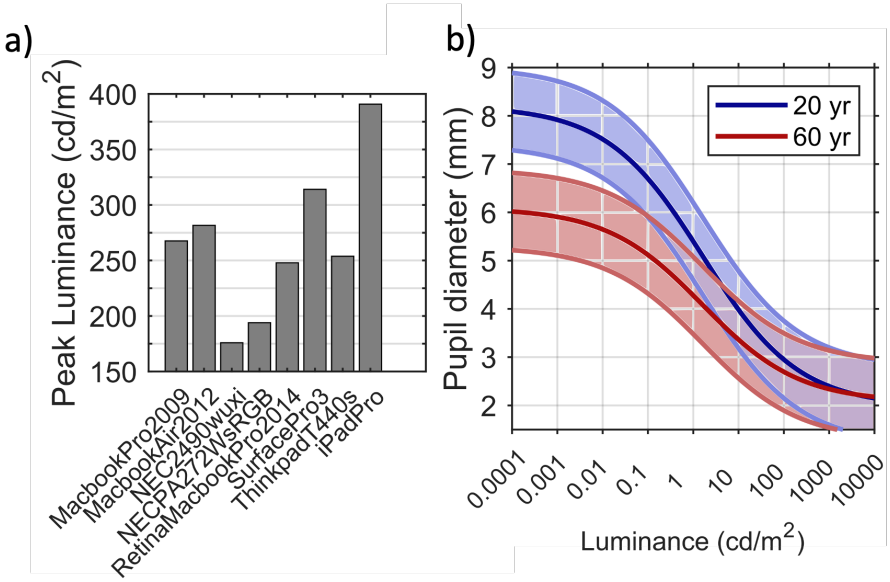


Figure 4.4: Figure showing the variety of luminance ranges that exist in common displays and the effect of the absolute luminance on pupil size. (a) The peak luminance output, taken as the luminance output when the display is set to the chromaticity of D65 white, of the eight example displays. (b) Plot showing the effect of luminance on pupil diameter for a 20-year-old and 60-year-old observer (data taken from the Watson and Yellott, 2012, model). As luminance increases the pupil constricts. There is large uncertainty in pupil diameter estimates between observers of a particular age at a particular luminance.

The absolute luminance level of a display is critical for determining the adaptation state of the observer. As the mean light level of the stimulus changes, so to does our spatial and temporal contrast sensitivity (Kalloniatis & Luu, 2007b, 2007a; Hart, 1987; Laming, 1991). As absolute luminance increases,

spatial contrast sensitivity increases, with a specific relative increase in sensitivity of higher spatial frequencies compared to lower spatial frequencies (Kalloniatis & Luu, 2007b; Laming, 1991), i.e. our peak spatial frequency sensitivity is at a higher spatial frequency for bright light conditions than low light conditions. Similarly, as absolute luminance increases, temporal contrast sensitivity increases, with a specific relative increase in sensitivity of higher temporal frequencies (Kalloniatis & Luu, 2007a; Hart, 1987) i.e. our peak temporal frequency sensitivity is at a higher temporal frequency for bright light conditions that low light conditions. This means that different observers who are adapted to different conditions will see a Gabor of a specific physical contrast to be a different contrast perceptually. Thus it is important to know the absolute luminance level of the display in order to understand the way in which the observer's spatiotemporal vision is operating.

Adaptation state is also tightly linked to the photoreceptor control of visual function: with the rods driving scotopic vision at low-light levels, the rods and cones working together in the mesopic range, and the cones alone driving photopic vision at bright light levels. Some of the differences in the spatiotemporal contrast sensitivity described above can be explained by the switch from cone to rod vision, but this is only part of the explanation for this change: there is also sensitivity adjustments within the photopic and scotopic ranges. While there is no clear point in the intensity range that separates the switch from scotopic to mesopic to photopic vision, it is important to know roughly where one falls along the intensity axis to know what photoreceptor signals might be contributing to the perception of a particular stimulus (Hood & Finkelstein, 1986; Stockman & Sharpe, 2006). In, other words if one wishes to run an experiment to quantify the contribution of the different photoreceptor signals to a particular phenomenon, one needs to know the absolute light level output of a display.

Additionally, the absolute luminance level of a display contributes to the control of pupil size of the observer (Watson & Yellott, 2012). Fig. 4.4(b) shows the relationship between photopic luminance and pupil diameter for a 20-year-old and a 60-year-old observer. As photopic luminance increases, pupil diameter decreases. Changing pupil size can impact the amount of light reaching the retina, and thus change the photopic and scotopic retinal illuminance. Further, a smaller pupil diameter can lower the limits of spatial vision, and reduce the maximum spatial frequency that an observer can resolve.

Ambient light

The light output of the display is not the only factor contributing to the absolute light level in the viewing environment. Ambient light also affects the overall light intensity. Ambient light can cause reflections from a display which can result in decreased contrasts and increased luminance of the

image or stimulus displayed (Devlin et al., 2006). Thus if ambient light is not controlled a stimulus will not have the desired contrast and luminance due to the reflections of the ambient light.

Ambient light also affects the adaptation state of the observer prior to viewing the display. Different observers may have been exposed to different levels of ambient light prior to starting a particular experiment and thus be adapted to different conditions.

For the above reasons, when conducting a psychophysics experiment in the lab, the level of ambient light is often minimised. If ambient light cannot be eliminated in the environment, then the ambient light level should be measured so that its impact can be quantified. It is also common practice in in-person testing to require the observer to undergo some period of adaptation to the light-level used in the experiment, to control for adaptation state of the observer.

4.2.4 Spatial characteristics

When presenting stimuli on a display, it is crucial that the spatial characteristics of the stimuli and the display are matched. This is particularly important for stimuli that have high spatial frequency components in them. In this section the two main display considerations – spatial resolution and viewing distance – and two main observer considerations – refractive errors and chromatic aberration; that impact spatial vision are discussed.

Spatial resolution

The types of displays that are likely to be used for online testing have a wide range of spatial resolutions. For example, the displays listed in Figure 4.2 range from 94 DPI (Dots Per Inch) (24-inch NEC2490WUXI, 1900×1200 pixels) to 264 DPI (11-inch iPad Pro, 1668×2388 pixels). According to the Nyquist–Shannon sampling theorem (Nyquist, 1924; Shannon, 1948), the highest spatial frequency that a discretely sampled system such as a display can reproduce is equal to half its sampling frequency. If one considers a display with a resolution of 94 DPI, which has a pixel size of 0.26 mm, the highest spatial frequency that could be displayed would be 2 cycles per millimetre. A 264 DPI monitor has a pixel size of 0.1 mm and could display spatial frequencies up to 5 cycles per mm. Not only does the DPI limit the highest spatial frequency that can be displayed, but any spatial frequencies in the stimulus that are above the Nyquist limit will be aliased. This creates unwanted power at lower spatial frequencies below the Nyquist limit, which will appear on the display as aliasing artefacts. It is therefore important that stimuli are low pass filtered before rendering to remove power at spatial frequencies above the Nyquist limit. Typically, one is interested in the spatial frequency content of stimuli defined in angular

units for a specific viewing distance.

Viewing distance

A consistent viewing distance is crucially important to maintain the desired visual angle or spatial frequency of the stimulus. The pixel size of a display can be converted from spatial units, s, to angular units, θ, using a simple geometric relation, as in Eq. 4.7,

$$\theta = 2 \arctan \frac{s}{2d}. \tag{4.7}$$

where d is the viewing distance. The sampling frequency in angular units is simply the reciprocal of the angular size of the display's pixels. The highest spatial frequency, sf, in cycles per degree (c.p.d), that can be reproduced by the display is therefore specified as in Eq. 4.8:

$$sf = \frac{1}{4 \arctan \frac{s}{2d}}. \tag{4.8}$$

This relationship is shown in Figure 4.5(a) for a range of displays, which are described in Table 4.1.

Display	Screen size (inches)	Screen resolution (pixels)	Screen resolution (DPI)
NEC2490WUXI	24	1920×1200	94
NECPA272WsRGB	27	2560×1440	109
MacBook Pro 2009	15	1440×900	113
MacBook Air 2012	13	1440×900	131
ThinkPad T440s	14	1920×1080	157
Surface Pro 3	12	2160×1440	216
Retina MacBook Pro 2014	13	2560×1600	232
iPad Pro	9.7	1536×2048	264

Table 4.1: Screen specifications for the eight example displays considered in these simulations.

Next I consider the reproduction of gratings on the retina after passing through the eye's optics, which is quantified by the optical transfer function (OTF) of the eye. Briefly, the imaging quality of an optical system is quantified by its point spread function (PSF), which describes how a point of light is reproduced. The Fourier transform of the PSF is the optical transfer function (OTF) and this describes how spatial frequency information is reproduced. Being complex-valued, one can consider the amplitude (the modulation transfer function, MTF, indicating contrast reproduction) and the phase (the phase transfer function, PTF, indicating relative phase shifts) of the OTF. For an eye that is optically perfect, the MTF is a decreasing function whereby higher spatial frequencies are reproduced with lower contrast, as shown in Figure 4.5(c) due to diffraction at the eye's aperture or pupil. The PTF

for such a diffraction-limited system is uniformly equal to zero – there is no shift in the phase of any spatial frequency component. The decrease in contrast is due to diffraction and the rate of decrease is determined by the wavelength of light, λ, and the diameter of the eye's pupil, d_p. The highest spatial frequency that can be reproduced is the optical cut-off frequency, the frequency at which contrast is reduced to zero, and is equal to d_p/λ for a circular pupil. Figure 4.5(a) shows the cut-off frequency for a diffraction-limited eye with a pupil diameter of 2 or 5 mm over the visible range of wavelengths. This shows that for typical pupil sizes, a diffraction-limited eye would be able to resolve gratings of up to 50 cycles per degree, but the viewing distance must be at least 150 cm to achieve this with a 94 DPI monitor or 55 cm to achieve this with a 264 DPI monitor. It should be noted that this analysis does not consider chromatic aberration, which is discussed in a later section.

Refractive errors

One must also consider the additional effect of any refractive errors that the participant may have that are not corrected for. One may reasonably expect up to ±0.25 D (dioptres) of spherical (focus) refractive error in the population as this is considered the range of emmetropia and so would not typically need a spectacle prescription (Langenbucher, 2015). For shorter viewing distances (less than 2 m) it is expected that a young participant would likely be able to accommodate and reduce their spherical refractive error to within 0.1 D at best (Charman & Heron, 1988). However, one should also consider that some participants may have larger spherical refractive errors, or may not be able to fully accommodate to the display due to presbyopia (which will be the case for a high proportion of older participants). Figure 4.5(c) shows an example MTF for an eye with 1 D of spherical refractive error, a pupil diameter of 4 mm and a stimulus wavelength of 555 nm. The MTF falls off with increasing spatial frequency and reaches the first minimum at 10 cycles per degree, which is considered to be the effective cut-off frequency. The MTF increases again between 10 and 20 cycles per degree, however the polarity is inverted, as indicated by the signed MTF in Figure 4.5(c). This inversion is evident on inspection of the PTF, since spatial frequencies between 10 and 20 cycles per degree are shifted by 180° and are therefore in anti-phase. This means that these spatial frequency components are inverted and are, therefore, not reproduced faithfully.

Figure 4.5(c) considers the effective cut-off frequency (the first contrast minimum) of an eye that has spherical refractive error, over a range of pupil sizes and at the extreme ends of the visible wavelength range (ignoring chromatic aberration). For example, with a spherical refractive error of 0.25 D, one should not expect spatial frequencies above 50 cycles per degree to be reproduced on the

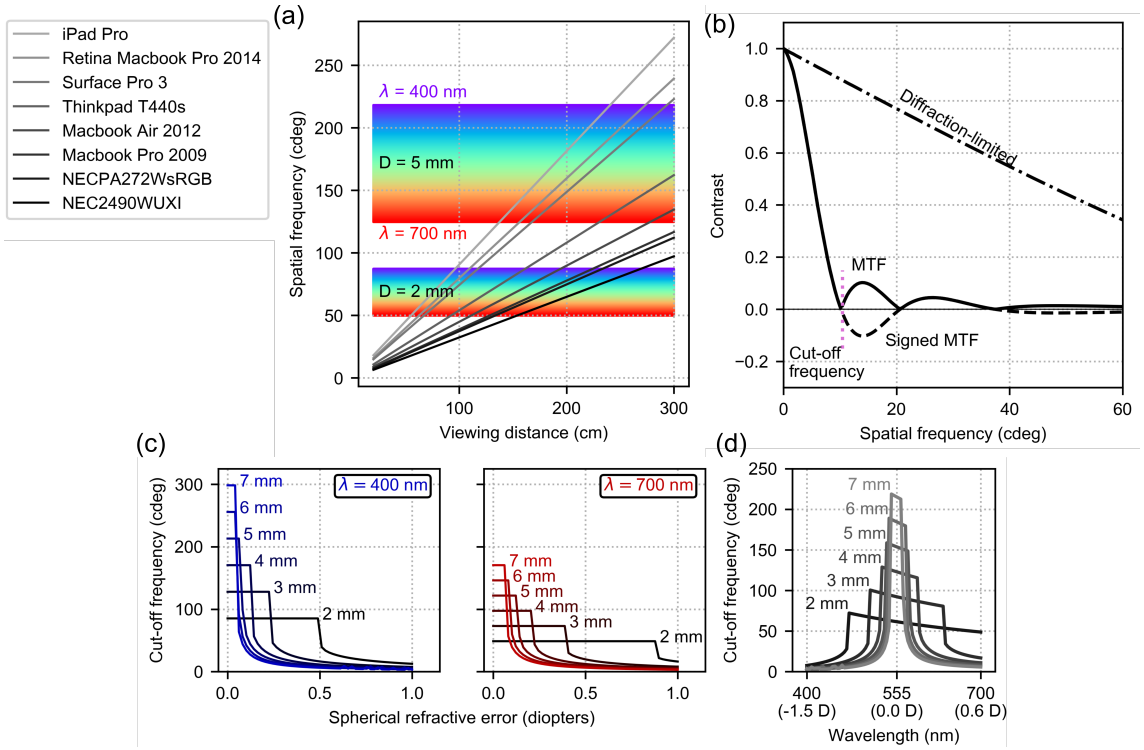


Figure 4.5: Spatial resolution considerations for stimuli displayed on a monitor. Panel (a) shows the maximum (Nyquist limited) spatial frequency that can be displayed on a range of monitors with resolutions from 94 to 264 DPI as a function of viewing distance. As an example, to accurately display a 50 cycles per degree (c.p.d) grating to a participant one would need a viewing distance of at least 150 cm for a 94 DPI monitor, or 55 cm for a 264 DPI monitor. The coloured regions show the highest spatial frequency that could be reproduced on the retina for an optically perfect (diffraction-limited) eye, which depends on the wavelength of light (λ) and the pupil diameter (d_p). Higher spatial frequencies can be reproduced by the eye for shorter wavelengths and with larger pupil sizes. For example, if the participant's pupil diameter is 2 mm, a 400 nm 87 cycles per degree grating could be reproduced on the retina. However, a 700 nm 87 cycles per degree grating could not be reproduced on the retina - in this case diffraction limits the cut-off frequency to 50 cycles per degree and altering the viewing distance will not change this. Panel (b) shows for a wavelength of 555 nm the modulation transfer function (MTF) of an eye with a pupil diameter of 4mm and 1 D (dioptre) of spherical refractive error (solid black line), the signed version of the MTF (dashed black line), indicating phase changes causing a change in contrast polarity associated with negative values, and the MTF of the equivalent diffraction-limited eye (dash-dot black line). The effective cut-off frequency is taken to be the first zero crossing in the MTF, indicated by the magenta dotted line, beyond which spatial frequency reproduction is compromised. Panel (c) shows the cut-off frequency for a range of spherical refractive errors and pupil diameters (indicated by line colour) at the extreme ends of the visible wavelength range. Note that a spherical refractive error of 0.25 D is not considered to be clinically significant and would not require a spectacle prescription. Panel (d) shows the cut-off frequency in the presence of chromatic aberration for a range of pupil diameters (indicated by grey lines). Chromatic focus shifts (in D) have been calculated under the assumption that the eye is in focus and therefore diffraction limited for a wavelength of 555 nm. It is useful to note that the sharp drops in cut-off frequency in panels (c) and (d) are due to the introduction of phase changes in the MTF (as demonstrated in panel (b)), which dramatically change the effective cut-off frequency.

retina for pupil diameters down to 2 mm and for stimuli with a wavelength of 700 nm. One should also note that the contrast of spatial frequencies up to these values will also be reduced.

Even in an eye that is correctly focused, residual refractive errors remain, including higher-order optical aberrations that cannot be corrected with lenses. These errors result in MTFs and PTFs with more complicated profiles, and so an effective cut-off frequency is difficult to define. Metrics such

as the visual Strehl ratio, which is a measure of the summed contrast weighted by the neural contrast sensitivity function, would be more appropriate (Thibos et al., 2004). While not discussed in detail here, it reinforces the point that caution should be taken when presenting stimuli with high spatial frequency components.

Chromatic aberration

In the previous sections the reproduction of stimuli on the retina, considering the wavelength of light, has been discussed. However, stimuli displayed on a monitor are broadband and are therefore subject to chromatic aberration in the eye. This effect is common to all experiments using broadband stimuli, not just those performed online, and is the 'normal' viewing condition of the observer. If one considers the eye to be best-focused for a particular wavelength, λ_f, one can estimate the chromatic focus shift, R_λ in dioptres, for other wavelengths present in the stimulus as in Eq. 4.9:

$$R_\lambda = q\left(\frac{1}{\lambda - c} - \frac{1}{\lambda_f - c}\right), \tag{4.9}$$

where $q = 0.63346 \times 10^{-6}$, and $c = 0.21410 \times 10^{-6}$ (Thibos et al., 1992). For grey scale stimuli, one can estimate $\lambda_f = 555$ nm as the peak wavelength of the photopic luminosity function. The reduction in cut off frequency associated with a particular chromatic focus shift, relative to 555 nm, is shown in Figure 4.5(d) for a range of pupil diameters. It is clear that chromatic aberration in the eye makes it impossible to reproduce high spatial frequencies on the retina simultaneously across the full visible spectrum without wavelength-specific correction.

4.2.5 Timing

Many visual psychophysics experiments require precise control over the timing of the stimulus. Almost all visual psychophysics experiments in the lab are conducted on either CRT monitors, or research-grade LCD monitors such as the CRS Display++ LCD Monitor (Cambridge Research Systems, Rochester, UK). One will rarely, if ever, find visual psychophysics experiments run on the typical LCD monitors that one might have access to at home for online experiments. A large reason for the persistence of CRT monitors for in-lab psychophysics is they offer precise stimulus timing in a way consumer grade LCDs do not.

Fundamentally, CRTs and LCDs use different driving mechanisms to produce a given luminance output on a single frame. CRTs, which rely on sending a raster scanned beam of electrons through a

phosphor, have a pulse luminance output signature, whereby the desired luminance output is delivered almost immediately after the phosphor is excited and the luminance then returns exponentially, (generally within the course of a single frame), back to its resting state. LCDs, however, rely on a different underlying technology, where the liquid crystals are physically realigned to allow more or less light to pass through. This results in a steadier luminance signature, but a much slower response time to reach the desired luminance after the frame is sent to be displayed. Then the luminance output is maintained at the desired level as the crystals maintain the same physical position, until the LCD is switched back to black, again requiring a shift in the position of the crystals and thus a slow decay back to black level. A schematic showing the typical luminance profiles from an LCD and CRT presenting four consecutive white frames followed by a black frame is shown in Fig. 4.6.

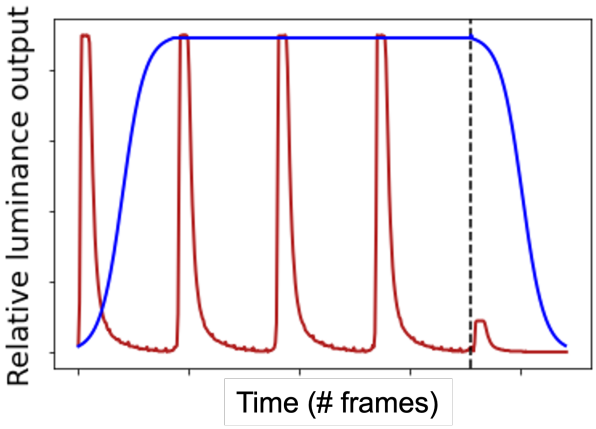


Figure 4.6: Temporal considerations of displays for stimulus presentation. The typical luminance output profile over time for a CRT (red) and LCD (blue) is shown. In this schematic the displays are sent four consecutive white frames, switching from a black frame to white at time 0, followed by a black frame (indicated by the black dotted line). The temporal profiles shown here demonstrate the large differences in the temporal profiles of LCD and CRT displays. The CRT luminance profile data are adapted from Rohr and Wagner, 2020 and the LCD luminance profile is simulated to match the typical luminance output profile of LCDs (based on the typical profile depicted in Elze and Tanner, 2012).

One obvious limitation of LCDs compared to CRTs is the much slower rise and fall times. This is typically worse when transitioning from a white to black frame and vice versa as opposed to between two mid-level grey frames (Ghodrati et al., 2015). In the schematic shown in Fig. 4.6 the LCD is depicted as fully transitioning from black to white and vice versa within a single frame, but this might not necessarily always be the case for some lower-end LCDs. This means, that unlike in CRTs where each frame can be considered independent from the previous, the output of the LCD on a given frame will depend on the output of the previous frame (Rohr & Wagner, 2020). This may result in persistence and artefacts when presenting moving stimuli (Elze & Tanner, 2012; Ghodrati et al., 2015). In particular, if one considers presenting a moving Gabor patch on an LCD, some of the stripes in the Gabor patch

will reach their target output luminance sooner than others, meaning intermediate artefact images will appear during stimulus presentation (Elze & Tanner, 2012).

Due to the temporal limits of consumer grade LCD displays, in-person experiments tend to prefer CRTs or research grade LCDs. However, with the continued improvement of LCD technology, it is possible the rise and fall times and persistence issues will improve as the technology does, and in fact, some studies have found less persistence in high-end LCDs compared to CRTs (Lagroix et al., 2012). One final issue one must consider is the stability of the temporal response of the display, and whether or not the display requires a warm-up period to reach this stability. For example, in the lab, I have found that the CRS Display++ does require about 20 minutes to reach peak luminance for a white frame. Fortunately, after a warm-up period, most displays do offer a stable luminance response profile (Ghodrati et al., 2015), though of course this will depend on the particular individual units.

4.2.6 Further Considerations

In this section I highlight some of the additional factors that an experimenter needs to consider when setting up a visual psychophysics experiment. It may not be possible to correct for these aspects, in the lab or for remote testing, but they do need to be acknowledged by an experimenter during experimental set-up.

Spatial homogeneity: When calibrating stimuli for psychophysical experiments, one ought to check for spatial homogeneity across the monitor. That is to say a stimulus presented at any spatial location of the monitor should appear the same regardless of its spatial location.

Viewing angle: Despite advances in LCD technology, a large angular viewing dependence persists. One must thus ensure the observer is always viewing the display from the correct angle, typically perpendicularly, so that errors are not introduced due to the observer's viewing angle.

Unknown observer characteristics: In the section on spatial resolution, I discussed how there may be persistent small refractive errors between individual observers that could cause individual differences in visual acuity even among observers who are considered to have normal or corrected-to-normal visual acuity. One must be aware of these differences that can persist among observers. One could simply have participants perform an acuity task or contrast-sensitivity function (CSF) measurement before completing an experiment to screen and/or correct for these individual differences between observers. Beyond acuity differences, individual observers may also have differences in colour perception. One should consider screening for colour deficiency if an experiment relies on an observer having normal colour vision. More subtly, as discussed in Chapter 2, there are large individual

differences in colour vision in the colour normal population, due to individual differences in the cone spectral sensitivities, meaning identical stimuli may appear differently to different observers (Asano et al., 2016; Webster, 2015). Such differences are of interest to the vision scientist studying the photoreceptor signals, but also must be considered if one is trying to control for these differences.

Channel and pixel independence: In modern LCDs it is often assumed that the three primary channels are independent from each other i.e. changing the output of one of the primaries has no impact on the output of the other two primaries. Further, neighbouring pixels are also often assumed to be independent i.e. the output of one pixel has no impact on the output of the neighbouring pixels. One should check if these assumptions hold, though often even in the lab these assumptions are not verified for each display as these assumptions are typically violated to some extent for modern LCD and OLED displays.

4.2.7 Display calibration checklist

In this first half of the chapter, I have reviewed the aspects of display calibration and experimental set-up for psychophysical experiments that an experimenter should be aware of before undertaking any data collection. In summary, those aspects are:

- The EOTF (which describes the relationship between input driving value and display light output), which is necessary for linearization. Failure to estimate the EOTF correctly can result in incorrect reproduction of contrast, mean intensity, spatial frequency power spectrum, and chromaticity of a visual stimulus.

- The CIE 1931 xy chromaticities and relative luminances of a display which are necessary to define the device-specific chromaticity transformation. Failure to derive a device-specific chromaticity transformation can result in errors in colour reproduction and colour appearance.

- The absolute luminance output of a display and of the ambient light in an environment. Failure to do so means that the adaptation state of the observer is uncontrolled, and can result in other artefacts.

- The spatial characteristics of a display and how they interact with the optical properties of the eye of the observer. Failure to do so can result in an incorrect reproduction of the spatial frequency components of a visual stimulus.

- The constraints of stimulus presentation timing.

- Additional considerations such as spatial and temporal homogeneity, channel and pixel independence, viewing angle, and unknown observer characteristics.

4.3 Review of existing methods of perceptual calibration

The above section details the many physical properties of displays and the observer-display interactions that must be considered when calibrating a display for visual psychophysical experiments. In a lab setting a highly accurate calibration can be achieved with a spectroradiometer or similar device. By simply measuring the full spectral radiance output of the display, at as many input driving values as necessary, one can generate both an EOTF for the display and determine the chromatic and spectral properties of the display for chromatic calibration. In a lab setting, one can also control for observer characteristics, by screening for visual acuity and colour vision deficiency prior to the experiment, and one will have tools such as a head-rest to keep the observer at a fixed viewing distance and a fixed viewing geometry. It is also possible to create a dark room with no or minimal ambient illumination, for example.

It is much more challenging to control all of these aspects when calibrating a display in a less controlled environment, such as the home, without access to specific (and expensive) calibration devices such as a spectroradiometer. This poses a major challenge for the move to online testing for experiments using visual stimuli. How can low-tech tools available to a naive participant with a display at home be used to produce sufficient calibration for visual psychophysical experiments? In this half of the chapter, I review some of the existing methods that rely on exploiting the capabilities of the human visual system to calibrate displays perceptually, along with other tricks that a researcher may be able to use to create a controlled environment for remote visual psychophysical experiments.

To identify papers that present perceptual calibration and/or remote display calibration for visual stimuli methods, I, along with five of my collaborators on this project, conducted a systematic literature review. The process for the literature search is depicted as a flow chart in Fig. 4.7. Two independent database searches were conducted and combined, with a forward and backward citation search on the relevant articles identified from the search and a supplementary Google Scholar search, which identified a total of 26 articles to review that presented visual calibration methods.

At the end of this chapter I provide a summary with a "perceptual calibration checklist", summarizing the main approaches described in this section along with their limitations and some possible extensions.

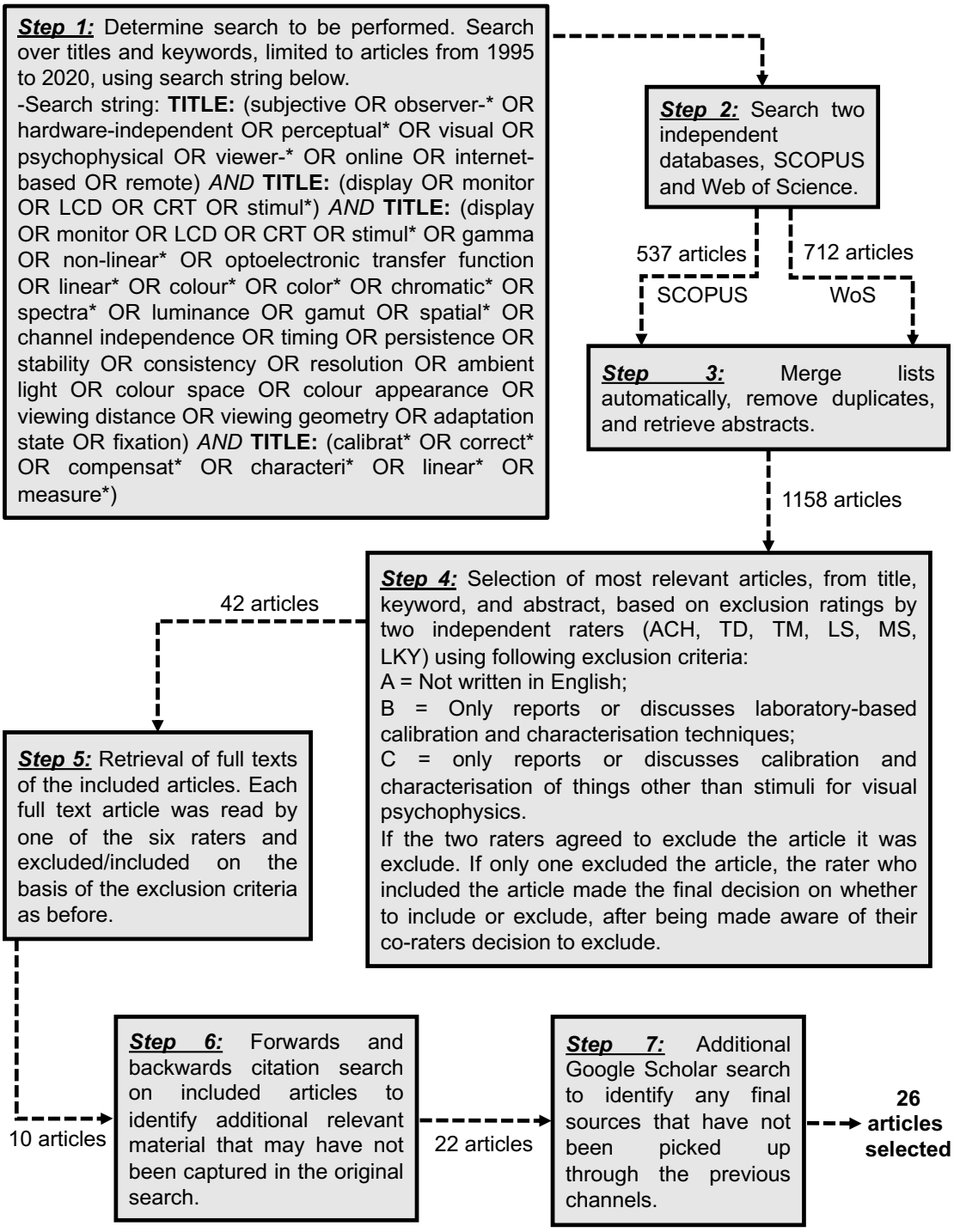


Figure 4.7: A flow chart of the systematic literature review process to identify articles relating to remote and/or perceptual calibration of visual stimuli. The rater ACH is the author of this thesis. The other raters (TD, TM, LS, MS, LKY) are all collaborators on this project, and co-authors on the OSA Webinar presentation where this work was presented, and will be co-authors on any publication resulting from this work. Note this flow chart does not identify all references in this chapter, just the articles describing perceptual/remote calibration methods.

4.3.1 Linearity of display output with input drive

The first aspect of display calibration considered here is the non-linear response of display output with input drive. Fortunately, a perceptual approach to estimate the EOTF of a display based on brightness

matching judgements gives a good approximation to the EOTF derived photometrically.

The basic principle behind perceptual linearization methods is to have an observer match a region on the display with a uniform input driving value to a region on the display that is filled with a half-tone pattern (Peli, 1992; Daly & Lee, U.S. Patent 5,754,222, May 1998; Edgar & Kasson, U.S. Patent 5,298,993, Mar 1994). The half-tone pattern may be, for example, a checkerboard such that the output of every other pixel is maximal while alternate pixels are set to minimum output, or a set of alternating raster lines where every other raster line was set to maximum output while the remainder were set to minimum output. An example brightness match stimulus, where the half-tone pattern is an alternating horizontal raster line pattern, is shown in Fig. 4.8(a).

The observer would then be asked to adjust the input value of a uniform grey patch until it appeared the same brightness as the half-tone pattern. At the point of a brightness match, the average luminance of the half-tone pattern and the luminance of the uniform patch should be equal. One can then use the known (relative) average luminance of the half-tone pattern to determine the (relative) luminance of the uniform patch that corresponds to the input value of the uniform patch set by the observer. In the example where the half-tone pattern is a checkerboard or alternating raster lines pattern, the output luminance of the half-tone pattern should be half of the maximum luminance of the display. At the point of a brightness match, the output luminance of the uniform grey-patch should also be half of the maximum luminance of the display, and one can then record the input setting for the uniform grey patch that produced a uniform grey of half the maximum luminance. This determines a single point on the EOTF, and in the case where a gamma function can be used as a good approximation for the EOTF of a display, e.g. for a CRT monitor, one can simply fit a gamma model to this match to estimate the EOTF (To et al., 2013; Peli, 1992; To et al., 2009). As an alternative to the brightness matching paradigm, one could use a different perceptual experiment that relies on an equiluminance match between half-tone patterns and a uniform colour patch, such as a motion-nulling technique (Mulligan, 2009), though brightness matching has emerged as the more common technique for perceptual linearization.

There are a variety of ways in which this perceptual brightness matching task can be implemented. One decision the experimenter must make is whether to estimate the EOTF of each channel independently, or whether to just measure the EOTF of channels combined (i.e. grey levels). This decision will depend on the type of stimuli to be displayed in the experiment, or the likely variability in the EOTF between channels for a particular display type. For example, if one wishes to use colour stimuli, then one must derive the EOTF for each colour channel.

A very quick way to implement perceptual linearization on displays (where the EOTF is assumed

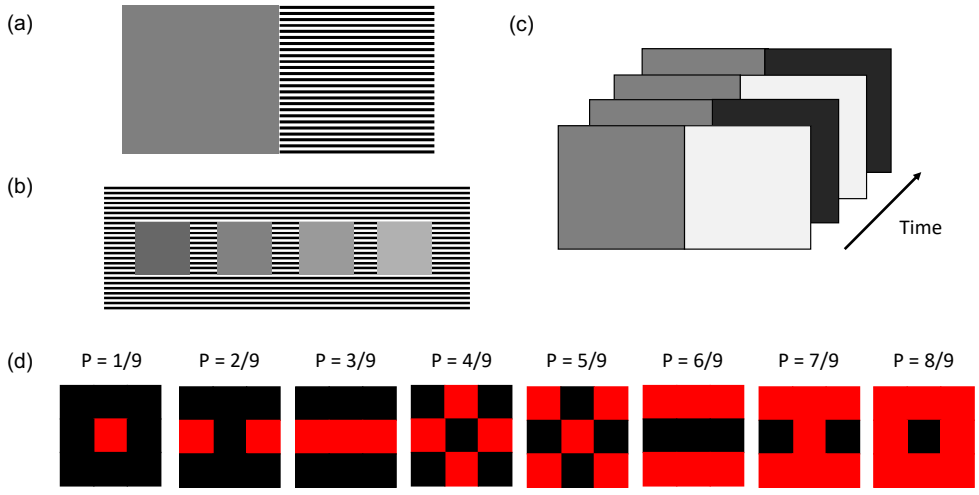


Figure 4.8: Examples of the type of stimuli that are used in perceptual linearization methods. (a) A schematic depicting the basic principle of the brightness matching task for perceptual linearization. In this schematic, the half-tone pattern (right) is an alternating horizontal raster line pattern where each alternating horizontal line is set to the maximum input value with every other line set to the minimum input value. The observer would vary the input value of the uniform grey patch until the brightness of the uniform grey (left) matched the brightness of the half-tone pattern. This would then give the input value corresponding to half of the peak luminance of the display, which would allow for a gamma function to be fit to estimate the EOTF. The set-up of the brightness matching task shown here depicts the set-up of the early implementations of perceptual linearization (Peli, 1992; Daly & Lee, U.S. Patent 5,754,222, May 1998; Edgar & Kasson, U.S. Patent 5,298,993, Mar 1994). (b) A quick implementation of the brightness matching task when a gamma model is assumed to represent the EOTF of the display. In this example, each grey square in the foreground is set to the input value that would produce half of the peak luminance of the display according to a different gamma value. The observer is asked to select which square best matches the background and the EOTF of the display is thus approximated by the gamma value associated with that square (Lavin et al., 1999). (c) A schematic of the temporal-dithering version of the brightness matching task, whereby the half-tone pattern is created by temporally alternating between frames of maximum input value and minimum input value. The observer would control the output of the uniform grey until the brightness of the uniform grey and the half-tone patch appeared uniform. (d) Examples of the eight half-tone patterns often used in brightness matching tasks today (shown here for estimating the EOTF of the red primary) (Mantiuk, 2016; Xiao et al., 2011). Each pattern consists of a 3x3 pixel grid which would be repeated to create a background of the half-tone pattern to which a uniform foreground square would be brightness matched. These patterns allow one to estimate an arbitrary EOTF of a display, without relying on assumptions about the shape of the EOTF.

to be described by a gamma function) would be to have the observer pick between a brightness match between one of a handful of uniform patches, derived from different gamma estimates, set against a background of the half-tone pattern (Fig. 4.8(b)) (Lavin et al., 1999).

For LCD and OLEDs, it is unlikely that the EOTF is well approximated by a gamma function, and thus one brightness match will not be sufficient to fit an estimate for the EOTF of the display. One, therefore, needs an observer to make multiple brightness matches to fit an EOTF for LCDs or OLEDs (Xiao et al., 2011; Neumann et al., 2003; Neumann et al., 2006). In fact, even in CRTs, multiple brightness matches may improve the perceptual estimate of the EOTF if the CRT is better described by

a gamma-offset-gain model, though there is conflicting evidence as to whether this is really the case (Besuijen, 2007).

To define half-tone patterns with varying mean luminance, one could either use half-tone patterns where a pre-determined proportion of pixels are switched to either minimum or maximum brightness output (Xiao et al., 2011; Neumann et al., 2003; Neumann et al., 2006; Besuijen, 2007; Colombo & Derrington, 2001), or one could use a recursive method, whereby the maximum brightness pixels are replaced with pixels set to produce say half of the maximum brightness (as determined by the brightness match from the checkerboard/alternating raster line pattern) (To et al., 2013; To et al., 2009). The advantage of using predetermined pixel proportion patterns is to avoid cumulative error that may occur in the recursive method (though it is unclear if such cumulative errors actually occur or create problems (To et al., 2013; To et al., 2009)). The disadvantage of using predetermined pixel proportion patterns is that it means using patterns with a spatial structure that may be visible (depending on pixel size and viewing distance), while a simple checkerboard or alternating raster line pattern may not be resolved (though it is unclear if the spatial structure of a simple checkerboard or alternating raster line pattern will be completely invisible (Colombo & Derrington, 2001) (see Section 4.2.4 for a detailed discussion of the factors influencing reproduction of spatial detail)).

The choice of the spatial structure of the half-tone patterns is dependent on the type of display. Checkerboard patterns are typically avoided in CRTs due to pixel dependencies along horizontal raster lines (To et al., 2013; To et al., 2009; Besuijen, 2007; Colombo & Derrington, 2001). Pixel dependencies between adjacent horizontal pixels are less of a concern for LCD and OLED displays and thus checkerboard patterns should be accurately reproduced (Xiao et al., 2011; Neumann et al., 2003; Neumann et al., 2006). Patterns of a predetermined proportion of the maximum luminance using alternating horizontal raster line half-tone patterns would typically lead to visible spatial structure in the half-tone pattern (Colombo & Derrington, 2001). Checkerboard patterns with a predetermined proportion of pixels set to the maximum luminance may be less visible. It is worth noting that even with visible patterns, most observers are still able to make brightness match settings that give a good estimate of the EOTF (Colombo & Derrington, 2001).

Different techniques to reduce the visibility of half-tone patterns have been employed, though there does not seem to be a definitive way to reduce the visibility of the half-tone patterns on LCDs or OLEDs. In CRTs, one may avoid the use of spatial half-tone patterns altogether, and instead use a temporal half-tone technique, whereby the half-tone pattern is created by alternating between frames of maximum output and minimum output, as shown in Fig. 4.8 (Miller & Yang, 2003). Most

LCDs and OLEDs do not have sufficiently high frame rates and sufficiently low persistence for such a temporal-dithering method, though it is worth keeping such an approach in mind as it may be a possible approach in future displays with higher frame rates and good temporal independence between successive frames. It is worth noting that the failure of the temporal-dithering method is not simply in the perception of the flickering frames (or the perception of persistence of previous frames), but the fact that the display will not be able to reach the maximum output driving value during the white frames, nor the minimum output driving value during the black frames, resulting in a different realised luminance for the half-tone pattern than desired (Besuijen, 2007). This highlights the fundamental limitation with the use of these half-tone methods: they only work if the assumption that the black frames are 0, and the white frames are 1, holds, and any display characteristic that violates this assumption compromises this method.

For LCDs, one approach to reduce the visibility of the traditional spatial half-tone patterns is to simply have the observer sit further away from the screen. In situations where the maximum viewing distance is limited (as in remote testing), mirrors were used to attempt to increase the viewing distance, but were found to cause the brightness-matching task to fail due to the strong angular dependency of LCDs (Besuijen, 2007). At present, there does not seem to be a solution to produce half-tone patterns on LCDs such that the spatial structure is completely invisible to the observer (though again, there is not sufficient evidence to suggest such patterns are always visible to observers or that any visibility impacts brightness matches or the EOTF estimate).

A commonly-used set of half-tone patterns, proposed by Xiao et al., 2011 that have been validated for perceptual linearization on an LCD are depicted in Fig. 4.8(d). This set of eight patterns contains half-tone patterns where in a 3x3 pixel grid, a set proportion, P, of the pixels are set to maximum brightness output. Brightness matches are made between a uniform patch and a background made of the half-tone patterns for each half-tone pattern and for each colour channel in turn. Once brightness matches have been made, the experimenter can fit the EOTF with a functional form such as a polynomial curve (Xiao et al., 2011), or solve an optimization problem to fit the EOTF curve (Neumann et al., 2003; Neumann et al., 2006).

For online testing a limitation of the published implementation of this method is that estimates of the EOTF improve as brightness matches from multiple observers are averaged (Xiao et al., 2011). It remains to be validated how reliable such a method may be for the case of a single observer making brightness matches to estimate the EOTF of a single display, as would be the case in online testing. Perhaps a cascade of repetitions, possibly with additional half-tone patterns interspersed between

these eight would be sufficient to estimate the EOTF from one observer. Further, one could imagine recruiting only participants where two or more individuals have access to the same display to combine brightness matching data from multiple participants estimating the EOTF of a single display. Combining brightness matching data from just two observers can improve the estimate of the EOTF, so there does seem to be scope in combining data from a relatively small number of participants who all have access to the same display (e.g. a household) (Xiao et al., 2011). Alternatively, the human observer could be replaced, or supplemented, in the brightness matching tasks with a camera, such as a web-cam, to remove concerns of individual observer anchoring (Thomas & Gerhardt, 2012). This camera must be spatially uniform but need not be calibrated because it is only needed to make a relative measurement between two patches on the screen.

Despite the promise of perceptual linearization methods, limitations of currently verified methods remain (Parraga et al., 2014). One obvious limitation for the case of online visual psychophysics is how to get a reliable estimate of the EOTF of an LCD or OLED with brightness matching data from just one observer. There is also scope to make the brightness matching task proposed by Xiao et al. 2011, a quicker and easier task for a naive participant to complete (Parraga et al., 2014). Another major, and perhaps debilitating, concern of these methods is the robustness of these techniques across different display technologies (Parraga et al., 2014). LCDs do not use a universal technology, but instead employ one of three display technologies: the Twisted Nematic LCD, the Vertical Alignment LCD, or the "In-Plane Switching" LCD. Parraga and colleagues show that while the perceptual linearization of LCD units of the Twister Nematic and Vertical Alignment technology produce good estimates of the photometrically derived EOTF, the perceptually derived and photometrically derived EOTF of the particular LG "In-Plane Switching" display they tested differed (Parraga et al., 2014). Particularly they found that observers did not perform as expected in the mid-luminance brightness matches. It is possible that this is the result of a particular extreme unit, but it highlights the need for caution when applying perceptual linearization methods across different display technologies (and indeed different units with the same technology) without verification of the efficacy of perceptual linearization for the particular display.

Dynamic range reduction

Following a successful perceptual linearization using the half-tone brightness matching task, it is likely that the participant's assumed native 8-bit display resolution would be reduced in resolution to within the region of 6-bit (To et al., 2013). Such resolution is unlikely to be sufficient to display just-noticeable

difference contrast stimuli. Bit-stealing is a method to increase the dynamic range of grey levels on the 8-bit display (Tyler, 1997). Bit-stealing shares bits between colour channels to increase the dynamic range of achromatic stimuli. Bit-stealing relies on the principle that a participant cannot discriminate the difference between a perfect grey level, e.g. $R = G = B = n$ where n is an arbitrary n-bit input value, from a grey level that contains an extra-bit or two for one colour channel than in the other two, e.g. $R = n$, $G = n$, $B = n + 1$. To implement bit-stealing one needs to know the relative luminances of the primaries in order to determine how many bits need to be "borrowed" from each colour channel to produce grey levels at intermediate levels of the luminances produced from perfect grey levels alone. Perceptual methods to estimate the relative luminances of the primaries are discussed in Section 4.3.2.

Saturating and cut-off non-linearities

Saturation of the highest or lowest input driving values can be relatively easily calibrated perceptually, as long as an observer is able to complete the half-tone brightness matching task. In terms of saturation towards the black level, one could simply use half-tone patterns that produce luminances in this region of the EOTF to get a better estimate of the shape of the EOTF at the very low luminance end (Neumann et al., 2003; Neumann et al., 2006). Alternatively, one could use a discrimination task where a range of the lowest input values were displayed on a background of the minimum (zero) input value. By selecting the first visible square then one can determine the minimum input value that produces a discriminable difference from the minimum output luminance (black level) (To et al., 2013). The same principle can be applied for the highest input values to determine the saturation non-linearity on both ends of the EOTF.

On 8-bit LCDs there may be a large increase in the output, departing from linearity in the interval between 254 and 255 due to the different physical arrangement of the liquid crystals. To account for this, one could simply exclude the estimate for an output of the 255 point, either in their fit of the EOTF or in the creation of the half-tone patterns for brightness matching, or both, and use 254 as the maximum value.

4.3.2 Chromaticity of stimuli

Perceptual colorimetric calibration remains a challenge for online psychophysics. The two key aspects of a display that one must measure to determine the chromaticity transformation between the native RGB colour space of the display and device-independent CIE 1931 XYZ colour space are 1) the relative luminances of the primaries and 2) the chromaticities of the primaries. In this section I review existing

techniques to estimate the relative luminance of the primaries, and discuss why it remains a challenge to estimate the chromaticities of the primaries perceptually. I also highlight workarounds that can be used to estimate the chromaticity transformation of the display when the chromaticities of the primaries remain unknown.

Estimating the relative luminance of the primaries

Estimating the relative luminances of the primaries is necessary for colorimetric calibration and can also allow one to employ bit-stealing to increase the dynamic range of a participant's 8-bit display (To et al., 2013; To et al., 2009). If one is able to estimate the relative luminance of the primaries for a particular display, then one can replace the relative luminances of the presumed sRGB model with the measured relative luminances of the display in question. Therefore, even without modifying the CIE 1931 xy chromaticity coordinates of each primary from the sRGB standard, one can reduce the colour reproduction error of the display.

One simple technique to estimate the relative intensities of the primaries is by making a white point adjustment. This could be done by simply asking the observer to make a patch on their display appear white, though such a white-point setting may be susceptible to chromatic adaptation (as achromatic settings are known to depend on the illuminant condition in colour constancy tasks (Brainard, 1998)). Alternatively, an observer could make a match between their display and a real white object, such as a white piece of paper (though white paper may emit light through fluorescene which could impact the match setting) (Sugiyama & Kudo, 2006, 2005). Introducing real objects and illuminants potentially recruits colour constancy mechanisms, which means that there may be substantial influences of the "mode of viewing", which might not be easy to control and may be different between the reference item and the display (Fairchild, 2013). One would either need to assume the observer is adapted to the ambient light in their environment such that the white paper appears white, or ask the observer to make the match under a known illuminant, assuming the observer is discounting the influence of the illuminant in the same manner for both the real object and the display. Another difficulty in making such a white point setting is when the luminance of the real-world object and display white differ it can be challenging to make a chromaticity match (Sugiyama & Kudo, 2006, 2005).

The luminance ratios of the primaries could also be estimated by a heterochromatic flicker task (Braun, 2003). In such a task frames of a single colour channel are alternated temporally, and the observer is asked to adjust the intensity of one of the channels until the perceived flicker between frames was eliminated or minimal, as shown in Fig. 4.9(a). The point at which the flicker was eliminated

or minimal corresponds to the equiluminance point for the two primaries. One could then compare the digital input values required to make an equiluminance match across all possible combinations of primaries to determine their relative luminance ratios (Braun, 2003). While this method will work well on displays with sufficiently high-frame rates such that one cannot perceive the colour switch between the frames of different primary colours, displays with low-frame rates will result in a noticeable flicker as the frames switch, making it difficult for the observer to achieve a flicker-null percept.

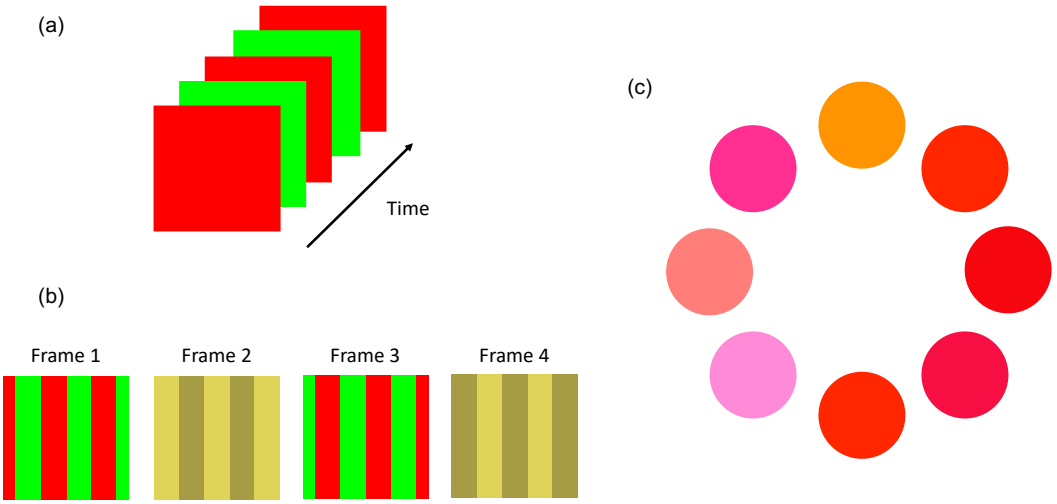


Figure 4.9: Examples of the type of stimuli that could be used to achieve colorimetric calibration perceptually. (a) Schematic of a heterochromatic flicker photometry task that could be used to estimate the relative luminances of the primaries (Braun, 2003). In this schematic, the display is depicted as alternating between frames of uniform red and uniform green over time. The observer would be able to adjust the relative luminance of one of the primaries, for example green, while the luminance of the red channel is fixed. When the relative luminance of one of the channels is greater than the other, the observer would perceive the display to flicker. When the relative luminances of the channels is equal then the observer would perceive minimal or no flicker. The display would need to have a sufficiently high frame rate such that the observer is able to temporally integrate between successive frames so that the colour of the display appears uniform throughout the alternating frames. (b) Schematic of the motion nulling task that could be used to estimate the relative luminances of the primaries (To et al., 2013; To et al., 2009; Mulligan, 2009; Anstis & Cavanagh, 1983). The four-frame sequence depicted here would be repeated during the task. In frames 1 and 3, alternating bars of red and green are shown, and the observer would have control of the relative intensity of one of the bars, for example red. In frames 2 and 3, alternating bars of brighter and darker yellow are shown, which are produced by a half-tone pattern of red and green. Between successive frames there is a $\frac{\pi}{2}$ phase-shift of the spatial pattern. If the red bars are brighter than the green bars in this schematic, then the observer would perceive rightward motion. If the red bars are dimmer than the green bars, the observer would perceive leftward motion. The observer would adjust the intensity of the red bars until the motion was nulled or minimal. At the motion null point, the relative luminance of the red bars and green bars would be equal. (c) A schematic showing an example of the unique hue selection task (Karatzas & Wuerger, 2007). The observer would be asked to pick the circle that appears "neither blue nor yellow" in this task. The task is repeated at progressively narrower hue angles until an estimate of the RGB setting required to generate a unique red is determined.

An alternative method to estimate the luminance ratios of the primaries takes advantage of a motion illusion (To et al., 2013; To et al., 2009; Mulligan, 2009; Anstis & Cavanagh, 1983). The stimulus

consists of four frames each shifted $\frac{\pi}{2}$ in spatial phase, presented on a continuous loop, as shown in Fig. 4.9(b). In the case of the green and red primaries, for example, when the green bar has a higher luminance than the red bar then the observer would perceive the stimulus as moving to the left. When the red bar has a higher luminance than the green bar, the percept would be the stimulus moving to the right. When the green and red bar have equiluminance then the motion would be eliminated or null. The observer is asked to indicate whether the stimulus is moving to the right or to the left, and based on the observer's responses the intensity of the red bar relative to the green bar is changed until the null point is found and the display input values corresponding to equiluminance for those primaries are determined. This can be repeated across the different primary combinations to achieve an estimate of the relative luminance ratio across all three primaries. It is worth noting that in their implementation of this method, To and colleagues suggest keeping constant the intensity of the primary with the lower relative luminance (i.e. red in the case of red and green, and blue in other comparisons) to ensure a null point is achievable within the 8-bit range for each primary (To et al., 2013; To et al., 2009).

One problem with perceptual methods to estimate the relative luminance ratio of the primaries based on equiluminance matches is inter-observer variability in equiluminance settings. The equiluminant point, particularly for the red and green primaries, depends on the relative ratio of L to M cones a particular observer has, and L:M ratio is known the vary widely across individuals in the colour normal population (with estimates ranging from 0.37:1 to 16.5:1) (Carroll et al., 2002; Hofer et al., 2005), as explained in more detail in Chapter 2. Therefore any calibration using equiluminant matches to determine the relative luminance ratio of the primaries will be observer-specific. For some applications this is exactly what is needed since it achieves the goal of equating the stimuli in terms of their effect on the luminance pathway of the individual participant.

Estimating the chromaticities of the primaries

While there are a variety of approaches to estimate the relative luminance ratio of the primaries of a display perceptually, perceptual methods for estimating the chromaticities of the primaries remain sparse. One possibility may be to extend the white-point setting technique, but to make chromatic matches in addition to white-point matches between household real-world stimuli, such as a dollar bill or a can of Coca-Cola, and the display (Woods et al., 2015; Mulligan, 2009). For such an approach, one must carefully pick objects that have a universal chromaticity that does not vary between units, or degrade over time, a chromaticity that will be within gamut of most displays (approximately within the sRGB gamut) and which are commonly and universally accessible. One must also carefully consider

and control the illumination under which the object is shown, as the chromaticity of a can of Coca-Cola under daylight would be different than the chromaticity in dim-office lighting. One would therefore need to not only have a reference object with a known chromaticity but a known chromaticity under a known illuminant, and an illuminant that could be easily obtained by a participant (e.g. daylight at noon on a clear day facing a particular direction). It is unclear, however, how much colour constancy mechanisms would affect the match, as this is an example of a classic asymmetric matching task, where the "mode of viewing" differs between the real object and the display (Fairchild, 2013), and the illuminant is generally considered to be discarded by the observer to some extent. This means such matches can be difficult for the observer to make. In some cases the chromaticities of the real-world objects may fall out of the gamut of the display making matches impossible. Another issue with such an approach is that any match, and thus any calibration, will be observer-specific due to differences in the cone spectral sensitivities between individual observers (Asano et al., 2016; Webster, 2015). It may be possible to replace the observer with an uncalibrated device, say a mobile phone camera, to make the match between the display and the real-world chromaticity under the same illuminant. Most cameras have white-point correction and any local white-point correction applied by the camera software would need to be disabled for reliable results from this method.

Alternatively, one may turn to the unique hues to achieve perceptual colorimetric calibration (Hering, 1964), as proposed by Karatzas and Wuerger, 2007. Unique hues are hues that appear to not contain any other hue. Unique red and green are the red and the green which appear "neither blue nor yellow", and unique yellow and blue are the yellow and the blue which appear "neither red nor green", and white (or grey) is the colour that appears "neither red nor green nor yellow nor blue". Most observers are able to make such unique hue settings, and these are usually stable for an individual observer, though they differ between observers (Webster et al., 2000). To use unique hues to perform a perceptual colorimetric calibration, an observer is asked to make unique hue settings for unique red, yellow, green, blue, and grey, on a display whose calibration profile is unknown (i.e. the display to be calibrated) and on a reference display (whose calibration profile is known). A schematic of the observer's unique hue selection task is shown in Fig. 4.9. By comparing the RGB settings corresponding to the unique hues on both the test display and the reference display, one is able to derive a chromaticity transformation matrix for the test display using the known chromaticities of the unique hues of the reference display.

The unique hue selection task uses similar principles to the idea of colour-matching between a real-world object of known chromaticity and a colour patch on the display. While the unique hue task

eliminates the need for reference objects, it introduces the need for a reference display on which the observers can first make the unique hue settings. It is not feasible to expect participants to have access to a photometrically calibrated reference display at home, which perhaps renders the unique hue task obsolete for the purposes of online testing. One could use average unique hue settings from a reference display instead to match the participant's unique hue settings to (Hu & Luo, 2019), but that relies on the assumption that unique hue settings are consistent across observers. While inter-observer differences in unique hue selection do not correlate to individual differences in the cone spectral sensitivities, there still exists large inter-observer differences in the hue angle of the unique hues (Webster et al., 2000). This again means that a display can at best, only be calibrated to suit the observer who performed the calibration and our present understanding of unique hue mechanisms does not allow this to be mapped to a known physiological condition (e.g. equiluminance). Further, the unique hue match can at best only calibrate hue, and does not provide a calibration for saturation and lightness.

Chromaticity of some common displays

It might be possible to develop some form of universal look-up table with display profiles. In fact, f.luxometer goes a large way to providing such a database ('f.lux Software LLC', 2011-2015). If one were to characterize the complete range of displays that one would anticipate participants might use for online testing, then one could simply ask the participant to enter the model number of their display and look up a previously derived photometric calibration profile. Experiments could even have inclusion criteria whereby only participants with displays that the experimenter has photometrically-derived calibration profiles for could participate. A major problem for such an approach, is the unknown variability between individual display units within a particular display model (Parraga et al., 2014). There is no guarantee, for example, that every MacBook Air 2012 has the same chromaticities and relative luminances. If such a database were to be developed, the variability between individual units would need to be explored. But if the variability was found to be minimal, such a database could be the key to improving colorimetric calibration for at-home psychophysics experiments. As display technologies and tolerances continue to improve, it is likely that variability between units will continue to be reduced in future.

4.3.3 Illumination levels

Estimating the absolute luminance output of a monitor is impossible without a photometric device. However one may be able to take advantage of some of the physiological responses to light to narrow

in on the absolute light level and estimate a range of maximum and minimum luminance outputs of a monitor. Further, there are ways in which one could control and estimate the ambient light in the at-home viewing environment. I discuss these methods in this section.

Absolute luminance level

One consequence of not controlling the absolute luminance level of a monitor is that it drives changes in pupil size as shown in Fig. 4.4(c). If one simply rotates this graph by 90° then one may be able to exploit these changes in pupil size to estimate the absolute luminance. If one could measure pupil diameter using, for example, a webcam-based tracking method, one could estimate the range of absolute luminances of the environment in which an observer of a known age may plausibly be in (the issue of receiving ethical approval for webcam-based tracking for online experiments is a separate issue that I do not discuss in this chapter). There are unfortunately large individual differences in pupil size. Pupil size estimates within the 100-400 cd/m^2 range, which corresponds to the range of maximum outputs of most displays, are largely overlapping and may not help to narrow the range of plausible absolute luminances further than the known range that exists in common display units. However, this technique may be helpful in estimating the black level (minimum luminance output) of the display as pupil size estimates are less overlapping in the lower luminance range.

One could imagine a similar database as the chromaticity database previously proposed that contained information on the minimum and maximum luminance output of a range of common displays that a participant is likely to use for online testing (again, f.luxometer offers a good starting point for this ('f.lux Software LLC', 2011-2015)). Indeed, many manufacturers will quote the maximum luminance output of a display, though whether or not these marketed luminance values are met in practice is unclear. There is also likely to be unknown variability between individual display units. Such marketing honesty and variability between units needs to be assessed before such a database could be a useful tool for the remote experimenter. Further, custom user settings and automatic adjustment of brightness of the display must be accounted for, and switched off before the participant begins an experiment.

Ambient light

Perhaps one of the simplest ways to control for ambient light for online testing is simply to ask the participant to complete the experiment in a completely dark room (i.e. a room with no windows, or with black-out curtains). Participants may have access to such a room (e.g. a bathroom without

windows, or a large closet) at home. A difficulty with this may be comfort for the participant during the experiment, and the ability to create the desired set-up within the constraints of the space the participant has available (for example, one may struggle to sit a sufficient distance from the display, or create a set-up where the display is viewed perpendicularly at eye level). Further, it is a challenge to verify that the participant followed instructions and did in fact complete the experiment in a dark room.

For some experiments, it may be possible to estimate the relative amount of ambient light and adjust the display settings such that a desired contrast can be displayed under the estimated ambient light conditions. Devlin et al., 2006, suggest a method to estimate the amount of ambient light reflected from a display by comparing the just-noticeable difference (JND) contrast thresholds in the dark and in office-lighting. By comparing these thresholds they estimate the relative intensity of the reflected (ambient) light in the office-lighting condition, and suggest subtracting this relative intensity from the input driving values when the display is viewed under ambient light conditions, to reproduce the contrasts achieved in the dark. This method, however, assumes that the participant is able to make a reliable JND setting in a completely dark room, at which point the participant may simply just complete the entire experiment in the completely dark room and thus the need for ambient light estimation is eliminated.

4.3.4 Spatial characteristics

Less work has been done to develop methods to perceptually calibrate spatial characteristics of the display. This is partially because some of the issues are simple to correct for, while others are very tricky (or in fact impossible). It may not be necessary to control these trickier issues directly, but rather it may be sufficient to simply be aware of the limitations they impose on stimulus design.

Spatial resolution

It is relatively straightforward to get estimates of display resolution and physical size for online testing. To obtain screen resolution, one could simply instruct their participants to retrieve the screen resolution from their native display settings and submit it via an input box during the experiment. One could do something similar for physical screen size, and have a participant use a ruler or tape measure to measure their physical screen size and submit the results. Alternatively one could ask the participant to match the size of a rectangle on the display with a common household object of a standard size, such as a £5 note or a credit card (Yung et al., 2015).

Viewing distance

Viewing distance is likely to be relatively easy to control for remote testing. It is reasonable to assume that most participants will have access to either a tape measure or a ruler and thus could measure the distance between their display and their eyes. A webcam-based tracking method may be able to detect if an observer's viewing distance changes throughout the experiment, though it may be hard to discern small differences in viewing distance via this method. Yet, small differences in viewing distance may be important, as a 10% error in viewing distance results in a 10% error in spatial frequency (as explained in Section 1.2.4), which could easily change the stimulus from being above to below threshold at the acuity limit where the contrast sensitivity function is steep. Participants may be able to use their own marker to maintain viewing distance. One of the challenges with viewing distance for online testing is, unlike in the lab where an observer could sit any arbitrary distance from a display and be able to respond to a stimulus using a Bluetooth keyboard for instance, the maximum viewing distance achievable may be limited if participants only have access to the inbuilt keyboard on a laptop for example. For experiments where a larger viewing distance than what could be reasonably and comfortably achieved using a laptop to both display stimuli and input responses, one may wish to only recruit participants with access to a keyboard that can be moved a set distance away from the screen.

Refractive errors

While it is expected that observers would have the appropriate correction for the viewing distance, it is possible that small refractive errors remain either because the observer does not have the correct prescription (or has not recently had an eye exam), or they cannot fully accommodate to the viewing distance. In these instances residual focus errors and uncorrected refractive errors including higher-order aberrations, would limit the highest spatial frequency that could be reproduced on the retina. In a lab setting, observers would be screened for refractive errors and optical correction could additionally be used. For online testing with stimuli that contain high spatial frequency components it would be prudent to test the observer's visual acuity with any visual correction they require for viewing the display. This could be a simple Snellen or LogMAR acuity test. If very accurate spatial frequency representation is required, the observer's contrast sensitivity function could be measured, for example by the Quick CSF method (Lesmes, 2010). This information could then be taken into account in the data analysis, or the stimuli could be adapted to compensate.

Chromatic aberrations

Chromatic aberrations cannot be compensated for outside a lab setting as this would require an achromatising lens. However, such a correction is unlikely to be desirable for most experiments that expect observers to view stimuli with the chromatic aberrations normally present in their eyes. In experiments where chromatic aberration is of concern, this would ordinarily be minimised in a lab setting using filters to reduce the wavelength bandwidth of the stimuli, but this would not be practical for online testing.

4.3.5 Timing

Unfortunately, the vast majority of displays participants are likely to have access to at home are typical consumer-grade LCD displays, which do not offer as precise control of stimulus timing as the sorts of displays used in psychophysics labs. This means the timing of the stimuli will typically be less well controlled for online testing than in lab-based testing. One must therefore be careful with the temporal properties of the stimulus they choose for online visual psychophysics. If high temporal control is needed, then one may consider only recruiting participants who have access to high-end gaming monitors which may offer sufficient temporal control for presenting moving stimuli, and avoid the issue of persistence (Ghodrati et al., 2015; Lagroix et al., 2012).

Beyond the limits of the physical display, if experiments are to be run via a web browser (as is likely the case in remote/online testing, unless the experimenter were to design a specific application for an experiment), there is the additional issue of synchronizing stimulus presentation in the browser with the refresh rate of the monitor, as the browser does not know when the display refreshes (Woods et al., 2015). Therefore, if a stimulus is to be presented at the frame rate of the display via a browser it is possible the stimulus will not be presented at all, should the presentation occur during a refresh.

It is typically possible for one to look-up or even set the frame rate of most displays. One could therefore specify the stimulus timing in terms of refresh rate and have the participant input, or programmatically obtain, the refresh rate of their display to better control stimulus timing. However this does not provide a solution for the asynchrony of the browser and the monitor refresh, or of the limits of the temporal response profile of LCDs. The safest way to ensure accurate timing is to present stimuli at a lower frame rate than that maximum that would be theoretically achievable based on the display refresh rate alone.

Temporal consistency and the required warm-up period could be tested by repeated perceptual linearization at multiple time points, though it is perhaps sufficient to assume temporal consistency,

as it is likely to hold true except for some extreme individual units, and estimate a warm-up period to wait before starting an experiment.

As with any calibration issue, to some degree it doesn't matter whether or not one can "solve" the problem, as long as one can *characterise* the problem. This argument was elegantly raised by Mollon and Polden in their paper on the time constant of electronic tachistoscopes, which reported that the time constants of such devices were much lower than widely assumed (Mollon & Polden, 1978). The fact that so many experimenters had not correctly calibrated their tachistoscopes, led to many experimental results being invalidated. Yet, Mollon and Polden do not argue that one should abandon the tachistoscope due to its poorer temporal performance than assumed, simply that one must not conduct experiments that were so reliant on temporal precision on the tachistoscope. The same is true for modern-day displays: as long as one is aware of the temporal limits of the display they are using then one can design their experiments appropriately. For in-person psychophysics, this would typically mean designing experiments to use CRTs or research-grade LCDs, while for online psychophysics this may mean designing experiments without moving stimuli, or where persistence would not confound the conclusions drawn from the experiment.

4.3.6 Further Considerations

In this section I review the miscellaneous factors that also need to be considered when setting up a display viewing environment for psychophysical experiments and ways in which these factors may be controlled remotely.

Spatial homogeneity: Spatial homogeneity may often not be verified even in lab settings. If one is particularly worried about spatial homogeneity one could restrict stimuli to a specific spatial location on the display, or repeat perceptual calibration methods across multiple spatial locations.

Viewing angle: In many modern-day LCDs, the major viewing angle dependence exists along the vertical axis. To control for the angular dependence along the vertical axis of modern-day LCDs in a lab, one is likely to ask their participants to view a display while their head is fixed using a head rest or bite bar. It is, of course, unlikely that participants have head rests for remote testing. One may, however, be able to employ a webcam-based tracking method to monitor the participant's head position remotely.

One point of particular note is that many remote participants may be completing online experiments on laptop screens, which often sit below eye-level. This may be particularly problematic for vertical viewing angle dependence. The experimenter should therefore be careful to instruct the participant to view the screen perpendicularly, whether that be tilting the screen to exactly the right

angle, or perhaps more simply, placing the laptop on a pile of books so that the centre of the display is at eye-level. The experimenter may also wish to bear in mind this viewing angle dependency along the vertical axis in LCDs when designing their stimuli, or their placement on the screen.

Unknown observer characteristics: Two unknown observer characteristics that are often tested for before conducting an experiment in the lab are visual acuity and colour vision deficiencies. Just as one could use remote methods to screen for visual acuity, one could employ online colour vision deficiency tasks, such as the City University Online Colour Vision Test, to screen for colour deficient observers. The effectiveness of such online tasks for screening for colour vision deficiency, however, is still debated.

Channel and pixel independence: For most modern LCD and OLED displays that participants are likely to have at home, it is probably safe to assume channel and pixel independence. If one was particularly worried about failures of these assumptions, one could design stimuli that avoid relying on channel and pixel independence entirely.

Default settings: It is crucial to note that any default settings of a participant's display that may impact the display's stability or override perceptual calibration settings must be switched off. For example, most Apple devices have a night mode feature and automatic brightness adjustment, both of which could nullify the results of a perceptual calibration and cause instability over time. Any such default settings should be switched off before perceptual calibration and only turned back on after data collection has been completed. The experimenter may give explicit instructions to their participants before an experiment on how to switch off these settings, or may be able to programmatically control them on a participant's display.

4.3.7 Perceptual Calibration Checklist

In this half of this chapter, I reviewed some of the existing methods of perceptual calibration. I discussed limitations of existing methods and suggested what further verification would be necessary for these methods to be applied for remote display calibration for online visual psychophysics testing. I also suggest some possible extensions of existing methods, some of which could be relatively easily implemented without additional verification, while other suggestions will require further validation and research. An itemized list of the main discussion points is given below.

- Brightness matching tasks between a uniform colour patch and a half-tone patterned patch can be used to perceptually linearize a display. However, it remains to be verified how universal some techniques are across different types of display technology.

- Perceptual colorimetric calibration remains a challenge. While one can use a perceptual equiluminance technique to estimate the relative luminances of the primaries, there is no solution to determine the chromaticities of the primaries, and thus achieve a full device-specific colorimetric calibration, perceptually.
- Absolute luminance cannot be determined without a light measurement device.
- An experimenter can gather information on the spatial characteristics of a participant's display relatively easily by having the participant input the required information.
- When considering timing, one must consider both the maximum frame rate of the display and the asynchrony between browser and display refresh rates for online experiments.
- Large viewing distances, and perpendicular eye-level viewing angles may be difficult to achieve given the practical constraints one may expect a participant to have at home. Further it may be challenging to verify a constant viewing distance and viewing angle are being maintained, though webcam-based tracking methods may be helpful here.
- Default monitor settings that may override perceptual calibration should be switched off.
- Perceptual calibration methods always give calibration profiles tied to a particular observer.
- In the future, a database providing photometric calibration profiles for a variety of display makes and models may be helpful to circumvent the need for perceptual calibration as a proxy for photometric calibration, though variance between individual display units needs to be assessed before such a database could be deployed.

4.4 Discussion

The move from face-to-face in-person experiments in the lab to remote online testing for visual psychophysics is perhaps both exciting and daunting. Possibilities of larger and more varied participant pools await the experimenter who is ready to venture into the world of online experiments. One of the biggest challenges the experimenter faces is that of remote display calibration without the use of expensive, specialist equipment.

The first half of this chapter should provide the reader with a thorough overview of the aspects of display calibration and how the stages allow for accurate stimulus control. The emphasis here is primarily on the stimulus. I am not claiming that every experimenter using visual stimuli should

be undertaking a thorough display calibration with a spectrophotometer and participant screening before embarking on any "in person" testing. There may be some experiments where it doesn't matter if the participant is colour deficient, or has remaining optical aberrations, or if the display is only approximately colorimetrically calibrated, for instance in a face perception task. Rather, I am advocating for every experimenter who uses visual stimuli to be aware of the ways their stimulus is distorted through an uncalibrated display and to make an informed choice of which calibration procedures are appropriate to ensure controlled delivery of their particular stimulus. For some more basic colour experiments, such as those concerned with characterizing the human photoreceptor signals, one may need a very thorough calibration procedure, while for some more subjective colour memory tasks, some aspects of calibration may be relaxed. One must always go back to the stimulus and ask what features of the stimulus are important for the particular experiment.

In the second half of this chapter, I review the existing literature of perceptual or remote calibration methods for presenting visual stimuli. The structure of the second half of this chapter, mirrors the structure of the first half, as I return to the stages of display calibration that I identified and discuss existing methods, their limitations and possible extensions, to achieve each stage remotely. Some aspects of display calibration can be achieved remotely through perceptual calibration, while others cannot. For some of the stages of display calibration, there is still some remaining uncertainty as to the applicability of current perceptual methods for the case of online testing.

Perceptual linearization methods, for example, give a good estimate of the photometrically derived EOTF of a display, but it is yet to be determined how reliable the EOTF estimated from just one observer's brightness matches and how well the perceptual method works across different display technologies. These concerns may mean that the EOTF derived perceptually by an arbitrary single observer on an arbitrary single display, as in the case of calibrating a display for online testing, may not match the EOTF derived photometrically. It is again worth emphasising that this may not be a concern in all applications. In some cases, one may not want to untangle the observer from the display, such as in personalized display imaging (Mantiuk, 2016) or medical imaging (Belaid et al., 1997; Li et al., 2002), where the goal is not to linearize the display such that the physical light output is linear but such that the visibility of features of a stimulus is maximised for a particular observer. Further, one may only need a rough approximation of the EOTF if exact reproduction of contrast, mean intensity, and spatial frequency are not crucial for an experiment.

In the case of colour, one must be aware that any colorimetric calibration achieved perceptually will be highly observer-dependent. Estimates of the relative luminances of the primaries will be affected

by the relative L:M cone ratio of a particular participant, and estimates of the chromaticities of the primaries (if a method to perceptually make these estimates is developed) will be affected by colour vision differences. Again, in the case of relative luminance estimates, this may not be a problem if one wishes to design equiluminant stimuli for a particular experiment. Additionally it has been a long-sought goal of personalized display imaging to achieve personalized colour imaging, so there may be use cases where personalized colour imaging is desirable. Unfortunately personalized or depersonalized perceptual colorimetric calibration remains a challenge with existing techniques as, to my knowledge, there is currently no method to measure or estimate the chromaticities of the primaries, which is crucial to determine the colorimetric calibration of a particular display without the use of a colorimetric device. Though again, I emphasise, precise colorimetric calibration may not be necessary for many experiments using chromatic stimuli, for many cases approximate colour reproduction using the native sRGB colour space for the display may be sufficient and importantly Section 1.2.2 puts limits on these errors. Though lack of colorimetric calibration for online experiments does render online experiments insufficient for characterizing cone photoreceptor contributions to the particular visual processing function being measured. Further, it is worth pointing out the obvious fact that at the time of writing this thesis, no consumer grade displays with more than three primaries are available to purchase. Thus it is impossible to use online testing, with traditional three-primary displays to study the function of all five photoreceptors. The need to utilise five-primary displays for such experiments is expanded on in Chapter 1 and Chapter 5.

Finally, I note that the methods of perceptual calibration reviewed here are likely to not be the limits of what can be achieved as a remote display calibration standard. As interest in online testing continues to rise I expect interest in improving methods for perceptual calibration to also rise. I speculate as to some of these improvements in this chapter and identify the limitations of existing methods that need to be addressed. One possibility of particular promise may be a calibration "database" where the calibration profiles of a large number of displays that participants are likely to have access to are characterized and made available to the visual psychophysics community. If such a database is to be developed, the variability between individual display units of a particular make and model must be investigated. One could imagine such a database, combined with some perceptual verification, as a future standard.

4.5 Conclusion

This chapter reviews the process and limits of stimulus calibration for online testing. In the first half of the chapter, I discuss the aspects of display calibration that one needs to consider to ensure accurate stimulus presentation, both in the lab and for online testing. In the second half of the chapter I present various ways in which these calibration steps can be achieved remotely, without access to a spectrophotometer or colorimetric device, by reviewing the existing literature on perceptual and remote calibration and discussing the limitations and possible extensions of past approaches. There is remaining scope for validating and improving perceptual calibration methods for online visual psychophysics testing. While there is room for improvement for perceptual calibration methods, there is still promise for the move to online testing both in the future and at present, as current methods offer sufficiently precise calibration for a range of visual psychophysics experiments.

It is perhaps helpful to reframe the discussion of in-lab vs online experiments, and refocus on stimulus control, considering instead the limits remote calibration places on stimulus delivery. One can then begin to design experiments and stimuli that can be accurately presented within those limits. Sadly, for the topic of interest of this thesis - human photoreceptor signals - current techniques for perceptual display calibration render such experiments beyond the limits of online testing.

4.6 Acknowledgements

The work presented in this chapter was part of a collaborative project, led by ACH, alongside Tsvetomira Dumbalska (TD), Dr Takuma Morimoto (TM), Li Shiwen (LS), Dr Manuel Spitschan (MS), Dr Laura Young (LKY), and Prof. Hannah Smithson. The work presented here was adapted from a conference presentation delivered by ACH at the OSA Technical Group Special Webinar: Display Calibration for Internet and At-home Human Vision, Visual Perception and Color Research (Hexley, Dumbalska et al., 2020) for which ACH was awarded the Best Paper Prize. This work is also currently being prepared as a manuscript for publication, from which the text in this chapter is largely adapted, on which ACH will be first author. The majority of the original text for the manuscript was prepared by ACH, though the section and simulation of linearity correction (Section 4.2.1) was written and performed by TM and TD, the section and simulation of spatial resolution (Section 4.2.4) was written and performed by LKY, and the figure showing pupil size vs luminance was created by ACH, using an adapted version of code, originally written by MS. For the literature review, ACH completed the database searches to generate the list of 1158, which were then split between six of the seven collaborators (ACH, TD, TM, LS,

MS, LKY) for inclusion/exclusion ratings. Once the 42 relevant articles were identified, 7 articles were read by each of these six raters, and short bullet point/paragraph summaries of the relevant articles written by the rater of the article in question. The final 26 articles were read in full by ACH and the summary text for each article as it appears in this chapter is the original work of ACH. All authors have had editorial input in the text being prepared for a manuscript, and MS and HES have made editorial contributions to the text as it appears in this chapter.

Chapter 5

Demonstrating a multi-primary high-dynamic range display for photoreceptor isolating experiments

Contents

5.1 Introduction 188

5.1.1 Melanopsin isolating experiments 188

5.1.2 Cone and rod isolating experiments 190

5.1.3 The MPHDR display 191

5.2 Methods 191

5.2.1 Hardware overview 191

5.2.2 Calibration 197

5.2.3 Display algorithm 198

5.2.4 Designing photoreceptor isolating stimuli 201

5.3 Results 203

5.3.1 High-dynamic range and high luminance output 204

5.3.2 Expanded colour gamut 205

5.3.3 High-contrast photoreceptor-isolating stimuli 207

5.4 Discussion 209

5.5 Future development 211

5.6 Conclusions 214

5.7 Acknowledgements 214

5.1 Introduction

As described in detail in Chapter 1, the first stage in the human light response is the absorption of photons of light by the photoreceptors: the three classes of cones (long-wavelength-sensitive [L], medium-wavelength-sensitive [M], and short-wavelength-sensitive [S]), the rods, and the melanopsin-containing intrinsically-photosensitive retinal ganglion cells (ipRGCs). There is a lot we now know about the function of these five photoreceptors, thanks to the many years of effort of countless researchers. Yet, as detailed in Chapter 1, there is still much we don't know. Crucially, much of our understanding of ipRGC function, comes from mouse models (Mure, 2021), yet there is a limit to how much we can learn about the human light response from mouse models, due to the stark differences between human and mouse retina (Huberman & Niell, 2011). Additionally, much of our understanding of cone and rod function has come from human psychophysics under the assumption that these were the only photoreceptors influencing perception, though it is now less clear whether this is the case (Zaidi et al., 2007; Zele, Feigl et al., 2018; Spitschan et al., 2017; Cao et al., 2018; Brown et al., 2012; Zele, Adhikari et al., 2018; Stockman & Sharpe, 2006; Zele & Cao, 2015).

5.1.1 Melanopsin isolating experiments

Ultimately, to fully understand the ipRGC contributions to vision and non-visual light mediated neural responses in humans, one needs to study the role of the photoreceptors in the awake, behaving human observer i.e. through human psychophysics. To achieve this, one needs a five-primary stimulus display system with which to implement the method of silent substitution (as presented in Chapter 1) to independently modulate the stimulation of one photoreceptor class, whilst holding constant the modulation of all other classes (Estévez & Spekreijse, 1982; Spitschan & Woelders, 2018; Estévez & Spekreijse, 1974).

Such melanopsin isolating psychophysics experiments have been fruitful to date in contributing to our understanding of melanopsin signals in humans. For example, the silent substitution method was used by Spitschan et al., 2014, to identify the opponent contributions of S-cone and melanopsin signals to the human pupillary light response. In this experiment, Spitschan et al., 2014, presented the melanopsin isolating stimuli via diffusing glass which the observer views, while the light is projected onto the glass via a fibre optical cable supplied by a custom-built digital spectral integrator. In a

separate experiment, Spitschan et al., 2019, utilised a five-primary display system, in which two projectors are stacked on top of each to offer five primary control, to silently modulate L, M, and S cones, and melanopsin while observers viewed short movie-clips, and demonstrated photoreceptor specific control of the pupil size of observers in this set-up. Allen et al., 2019, used a 4-primary display to design melanopsin isolating stimuli to investigate form vision from melanopsin. As a final example, Zele, Feigl et al., 2018, use a five primary photostimulator (Cao et al., 2015) to demonstrate a potential melanopsin contribution to perception of colour.

The difficulty with melanopsin isolating psychophysics is that in order to achieve melanopsin isolation, one needs a display with as many primaries as photoreceptors they wish to control, i.e. five. As no commercial display to date is available with five primaries, this requires the development of custom displays to be built by the experimenter before they may carry out their experiments. Display design always requires compromise. For example, perhaps the most widely used method for delivering melanopsin-isolating stimuli is through custom-built multiple primary Maxwellian view systems, like that of Cao et al., 2015 (and indeed the earlier instances of that of Sun et al., 2001a and Pokorny et al., 2004, if one wants control of just four of the photoreceptors). Such Maxwellian-view systems, however, are typically limited in their spatial control, and depending on the originating light source may be limited in temporal control. This imposes limits on the types of stimuli that can be displayed on such systems, and subsequently what we can learn about ipRGC function from such systems. For example, we don't have a good quantification of the spatial distribution of ipRGCs, or how ipRGC characteristics vary across retinal eccentricity, and such display systems would not enable us to study this easily.

An alternative display design is the approach of e.g. Spitschan et al., 2019 and Allen et al., 2018, who use a projector-based design. Such projector-based multi-primary display systems allow the experimenter to generate photoreceptor isolating stimuli that are spatially controllable across the visual field (Packer et al., 2001; Ajito et al., 1999). However, such display systems can currently only generate low luminance, low-dynamic range stimuli. Thus if such multi-primary projector-based display systems are used to study the function of individual photoreceptor classes, one cannot assess their function at high luminance levels or over a high-dynamic range (HDR). This is particularly limiting for studies into the role of melanopsin, which is known to have a relatively low sensitivity, and to operate over a large dynamic range (Do & Yau, 2010; Dacey et al., 2005). Thus such display systems do not allow us to probe the operating range of ipRGC contributions to various mechanisms, which is of interest as it is likely that the nature of photoreceptor integration varies over operating range. The limitations of current display systems that achieve melanopsin isolation in practice thus mean that the

experimenter always has to compromise on experimental design.

5.1.2 Cone and rod isolating experiments

Additionally, a critical perspective on the limitations of current display systems in conjunction with our growing understanding of ipRGC function, highlights the need for caution over traditional understanding of rod and cone function. Unlike our understanding of ipRGCs, much of our understanding of cone and rod function has come from human psychophysics. However, such psychophysics has traditionally been implemented on traditional three-primary displays such as CRTs (Brainard et al., 2002; Carpenter & Robson, 1998), under the assumption that a certain photoreceptor class is responsible for the observed behavioural response. Such assumptions are typically made by controlling the adaptive state of the cone or rod subsystem. For example, it has been long believed that rods saturate in bright light, and thus rod function is often ignored when considering the response of the observer to a stimulus displayed on a CRT. However, recent work suggests that rods may not be saturated at light levels that were previously assumed to be well into the photopic range, and CRTs almost certainly have brightness outputs that fall in the mesopic region (Lall et al., 2010; Frederiksen et al., 2021; Shapiro et al., 1996).

Four-primary systems have thus enabled much of our understanding of the interplay between rod and cone function (Sun et al., 2001a; Shapiro et al., 1996; Pokorny et al., 2004). For example, it was via a four-primary photostimulator that Sun et al., 2001b, demonstrated how rods and cones sum into the postreceptoral pathways. However, to date most of the work on rod and cone function has largely overlooked the possible contributions from ipRGCs, as it is only recently that ipRGCs have been implicated with such potential function (Zaidi et al., 2007; Zele, Feigl et al., 2018; Spitschan et al., 2017; Cao et al., 2018; Brown et al., 2012; Zele, Adhikari et al., 2018; Stockman & Sharpe, 2006; Zele & Cao, 2015). There is therefore also a need to utilise five-primary stimulus delivery systems to assess cone and rod function independently of ipRGC activity and more specifically to study the interactions of the signals from all five photoreceptors. One such example of a case where it would be prudent to utilise five-primary display systems to look for rod and ipRGC inputs into traditional cone-mediated function would be an experiment investigating a potential role of rod and melanopsin signals in illuminant estimation, as discussed in Chapter 3.

5.1.3 The MPHDR display

Thus, there is clear scope for improved custom-display design to enable necessary tests of melanopsin, cone, and rod function that are unachievable on current devices. Here, I present such a system: a multi-primary high-dynamic range (MPHDR) display for photoreceptor isolating experiments to study the human photoreceptor signal inputs to vision. The MPHDR display that is presented in this chapter is the first system to my knowledge that allows for both spatio-temporal and high luminance, high-dynamic-range control of stimuli to conduct photoreceptor isolating experiments without the compromises imposed by currently available systems. The MPHDR display offers five independent parameter spatial control of six effective primaries over a high-dynamic range (see Section 4.2.3 for details on the display algorithm).

5.2 Methods

5.2.1 Hardware overview

The hardware system design is inspired by earlier work that used a combination of a Digital Light Processing (DLP) projector and a Liquid Crystal Display (LCD) (Seetzen et al., 2004) to produce a HDR display. The DLP projector was used to replace the static backlight of a commercial LCD monitor such that the backlight is dynamically controlled to create a dual-modulation display system. In the MPHDR setup, two DLP projectors are used to modify the primary colours of this backlight, and improve the image quality with respect to the original design using updated components. The main components of the hardware are given in Table 5.1.

The additional primaries needed for the MPHDR setup, depicted in Fig. 5.1(a), are derived by placing different filters in front of the two projectors. The filters were selected to provide maximum available melanopsin contrast and maximum available colour gamut (details on which are provided later in this section). The spectra of the individual components used in the set-up are shown in Fig. 5.1(b)-(e).

Modification of the LCD panel

The first step in construction of the system is to modify the most delicate component, which is the LCD panel. In the MPHDR setup, a high-resolution LCD display panel is used, as listed in Table 5.1. Modern LCD panels contain several light enhancing layers, which are stacked to improve the light efficiency of the display (Seo et al., 2012). These layers are used to improve brightness and screen uniformity

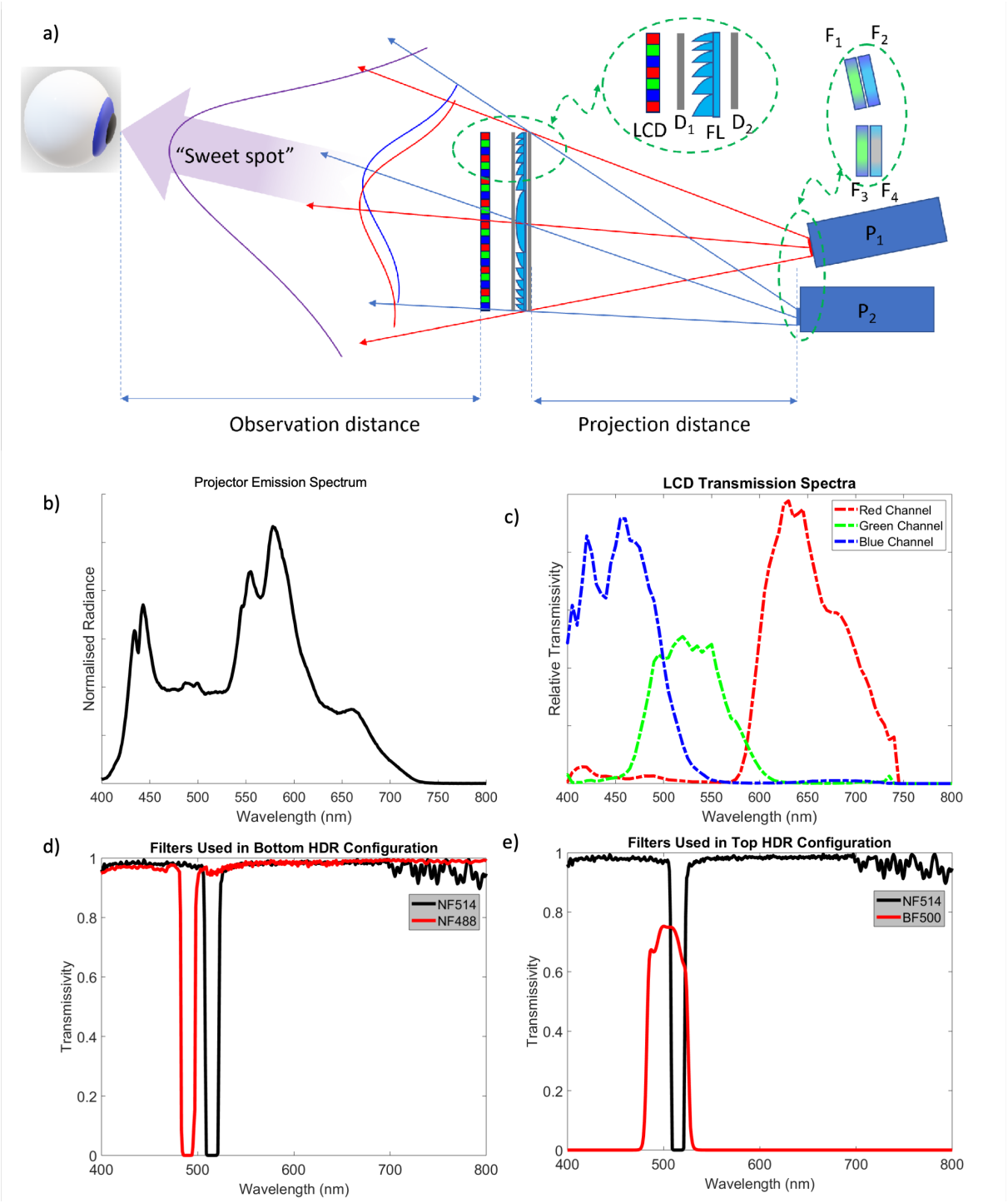


Figure 5.1: a) Schematic of the display setup. Observers should view the display from the "sweet spot", the optimal position for overlap of the backlight from each projector. b) The relative emission spectrum of the projector lamp after passing through the DLP chips used in the display (without filters). c) The transmissivity of the LCD panel used in the display, measured before construction of the set-up. The spectra of the DLPs and LCD transmission were measured using a JETI Specbos 1211. d) The transmission of the filters used in the bottom HDR configuration, where the bottom projector (P_2) illuminates the LCD. e) The transmission of the filters used in the top HDR configuration, where the top projector (P_1) illuminates the LCD. The transmission of the filters reported here is specification data provided by the manufacturer, ThorLabs (Thorlabs, 2020c, 2020b, 2020a).

Table 5.1: Components of the display system.

Component	**Brand**	**Properties**	**Modification**	**Acronym**
LCD Panel	LP097QX1 (No brand or manufacturer information available)	2048x1536 IPS 9.7inch	Reflective layer removed	LCD
Projectors	Acer x128H (Acer UK Ltd, Greater London, UK)	1024x768	Colour wheel removed	P_1, P_2
Fresnel Lens	Comar Optics (Comar Optics Ltd, Cambridge, UK)	Focal length 305mm	NA	FL
Diffusers	Rosco (Rosco, London, UK)	3028 1/4 Tough White Diffusion 3040 Powder Frost	NA NA	D_1 D_2
Filters	Thorlabs (Thorlabs LTD., Ely, UK)	NF514 BF500 NF514 NF488	NA NA NA NA	F_1 F_2 F_3 F_4

by recycling the light generated in the backlight. These layers, which are depicted in Fig. 5.2(a), were removed to reach the bare LCD panel. In the MPHDR configuration, the LCD panel is placed such that the front side of the LCD panel (shown as 1 in Fig. 5.2(a)) is directed towards the projector. This is because the front side of the panel contains an anti-reflective treatment, which helps to reduce the reflections of the DLP projection, which is acting as a backlight. The light throughput was significantly higher when placing the LCD panel in this orientation.

Modification of the DLP projectors

To maximize the brightness of each DLP projector, the colour wheel was removed from the light path. Control circuits prevent the projector from operating when the colour wheel is disconnected, so the colour wheel was retained inside the projector, just placed outside the beam path. After the colour wheel was removed, the exposed empty space was covered with a custom-made 3D printed part to prevent stray light inside the projector. Then, the emission spectra were modified by adding custom filters in front of the lens to improve the available gamut and melanopsin contrast of the resultant MPHDR. Custom designed 3D printed mounts were used to hold the filters in place at the exit aperture of the projector lens, as shown in the inset of Fig. 5.2(b). The key difficulty was in the design of the

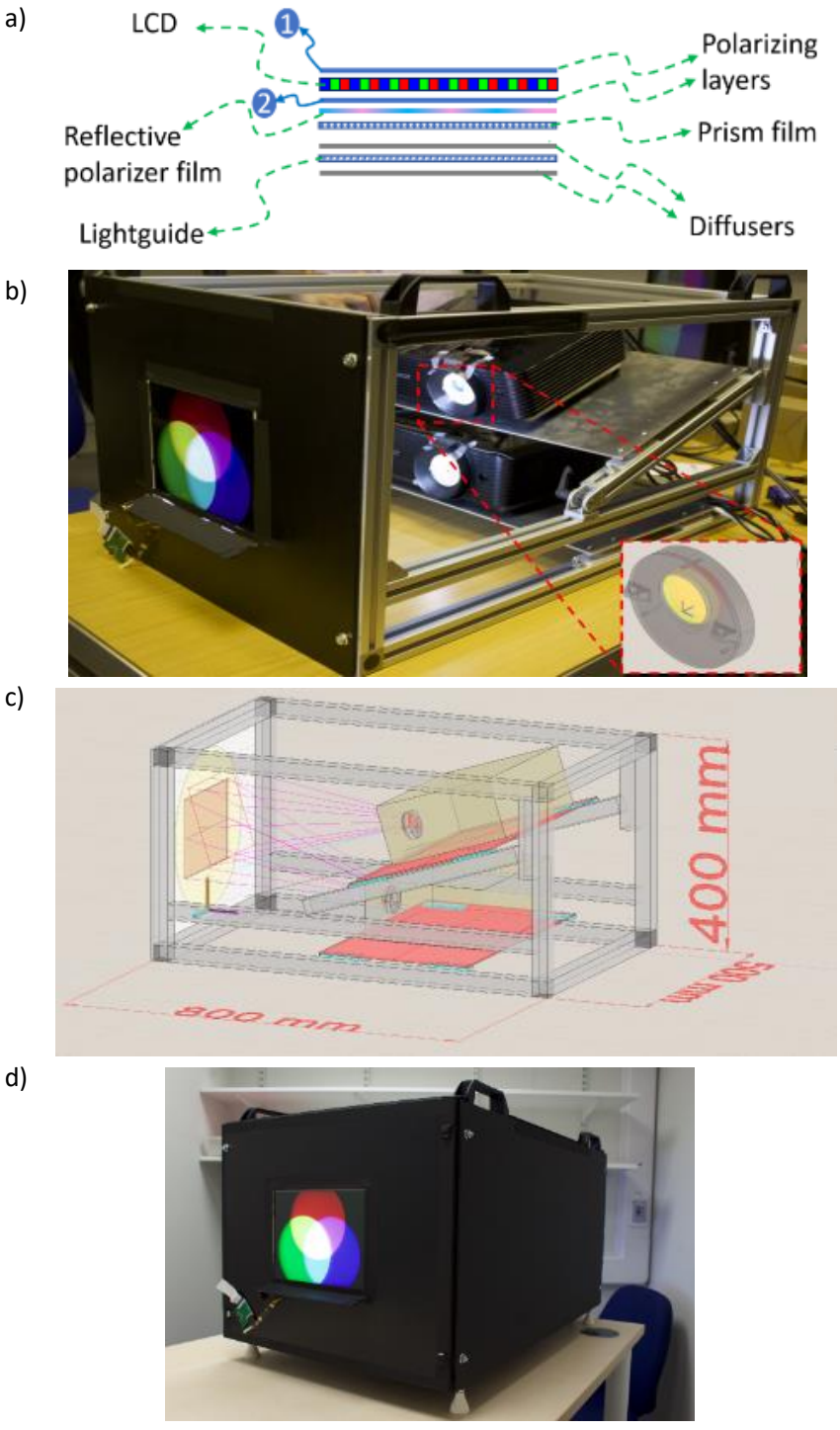


Figure 5.2: a) Layers of a generic LCD panel. b) Photograph of the MPHDR without the final casing. The CAD design of one of the 3D printed filter mounts is shown as an inset. c) A representative CAD drawing of the setup. d) Photograph of the MPHDR with final casing.

3D-printed parts so that they were not exposed to the direct beam path coming from the lamp. This was because the temperatures produced by the beam are significantly above the thermal limits of most materials used for 3D-printing.

Choosing the appropriate filters

The display was designed with independent control of melanopsin in mind, as the first batch of experiments to be run on the system will be using the method of silent substitution to modulate melanopsin stimulation. While appropriate for isolating any of the five human photoreceptor classes, the filters in our set-up, Fig. 5.1(d) and (e), were chosen to maximise melanopsin contrast and colour gamut. If the goal was instead to isolate S cones via silent substitution, one may wish to select filters such that S cone contrast and colour gamut were maximised.

By measuring the DLP emission spectra and the transmissivity of the LCD panel (Fig. 5.1(b) and (c)), one can predict the effective primaries in the display, as each effective primary is a simple multiplication of the relevant DLP emission spectrum with the relevant LCD transmission spectrum. Using the display algorithm described in Sec. 2.2.3, it is then possible to predict the spectral output of the display and calculate the available contrast of each photoreceptor and the available gamut. Without filtering the light from the DLPs so that the two DLPs have a different spectra, the display would simply be a typical three primary HDR. By filtering the light, one is able to achieve the multi-primary control described here.

To choose the appropriate filters for our set-up, an "n choose k" search was performed over all possible combinations of appropriate commercially available filters. Due to the physical positioning of the filters in front of the aperture of the DLPs, the filters had to be at least 25 mm in diameter so as not to clip a significant amount of the light emitted from the DLP. A non-exhaustive list of commercially available filters that fit this description was compiled. I then modelled the emission spectrum for each DLP with any combination of two, one, or none of the commercially available filters filtering the DLP emission spectrum. A combination search was performed on this space, to select the two filter combinations that were appropriate for the two DLPs to provide spectra with maximum melanopsin contrast and maximum available gamut.

The optical system design

Most projectors have an off-axis image formation, hence the light rays leave the objective lens above the optical axis. This has implications for the design of the filter mount and the design of the light

frustum. It was assumed the light rays are leaving a rectangular area on the lens surface the light frustum between the lens and the image on the screen was designed accordingly.

In order to direct the light towards the observer, and improve display uniformity and efficiency, a diffuser screen (Rosco, London, UK) and a Fresnel lens (f = 305mm, Comar Optics, Cambridge, UK) was placed in front of the modified LCD panel. However, when the diffuser was placed before the Fresnel lens (i.e. when we realise the image from the projector on the diffuser first) the transmitted light reflected between the LCD panel and the lens surface, since both surfaces are reflective. Therefore, a second diffuser was placed after the Fresnel lens. This also improved the visual quality by reducing the visibility of rings on the Fresnel lens, thereby making the illumination more uniform. The diffusers were stretched and fixed such that the gaps between the diffusers and the Fresnel lens were minimal, especially on the side with the lens grooves. The focal length of the lens was chosen so as to form converging light rays towards the observer.

The final display configuration can be thought of as two HDR configurations stacked on top of each other. I refer to the HDR configuration driven by the top projector (shown as P_1 in Fig. 5.1(a)) as the top HDR configuration and to the HDR configuration driven by the bottom projector (shown as P_2 in Fig. 5.1(a)) as the bottom HDR configuration. For the top HDR configuration, in which the top projector illuminates the LCD, the projected image was aligned according to the LCD position. Three options for creating overlapping images from the two HDR configurations were considered:

1. **Beam-splitter option:** The second projector is rotated by 90°, projecting images from an axis folded 90° relative to the main projector optical axis on the lateral plane. Although this would allow optimal alignment and calibration, this configuration suffers from light loss due to the beam-splitter.

2. **Flipped top projector option:** Rotating the top projector 180° around its own optical projection axis creating a mirror symmetry of the bottom projector with respect to the centre of the LCD panel. Although this design makes it easier for mechanical construction of the system, it suffers from significant divergence of light from the two projectors. In this arrangement, someone viewing the display from the center of the LCD, would receive significantly lower light from either the bottom or the top projector as the main beam paths would be above or below the central axis respectively.

3. **Slanted top projector option:** Ultimately, the top projector was mounted at an angle above the bottom one, aiming the beam towards the center of the LCD as shown in Fig. 5.2(c). This

configuration was the most convenient since the physical alignments were easier and the light loss was minimal. The beam paths were close to one another such that when the display is viewed from the center of the LCD, it is possible to receive significant amounts of light from both projectors as in Fig. 5.1(a). The top projector was positioned such that the image on the diffuser overlaps with the image projected from the bottom projector. The tilt angle of the top projector was measured as approximately 25°. Therefore, the overlapping region of the images, i.e. "sweet spot" in Fig. 5.1(a), is defined within this angle. The size and the focusing of the image were adjusted accordingly. Two drawbacks of this configuration were the keystoning correction requirement and the reduced light levels due to the off-axis positioning of the top projector with respect to the bottom projector.

The entire system was enclosed to prevent stray light reaching the viewing position (Fig. 5.2(d)). In order to maintain stable running conditions, two inlets and two exhaust fans were installed for the enclosure, as the projectors produce a large amount of heat.

5.2.2 Calibration

The MPHDR is calibrated in multiple stages. The calibration procedure is repeated for the two HDR configurations: the top HDR configuration and the bottom HDR configuration. In the first step, the geometric transformation from DLP pixel coordinates to the LCD pixel coordinates is determined, to align DLP and LCD pixels. This is achieved by displaying a grid of points first on the LCD, then on the DLP, capturing both with a camera (Canon 550D, Canon Inc., Tokyo, Japan), and finding both homographic transformation and mesh-based warping to obtain high quality alignment of the two displays. Next, a grid of points on the DLP is captured to measure and model the amount of blurring introduced on the DLP by the diffuser and due to its lower resolution. Then, an image of a uniform white field is taken and used to compensate for the spatial non-uniformity of the display. Because the slight vignetting caused by the filters on the DLP projectors prevented detection of the corners of the calibration grid, the filters were removed during geometric calibration. The keystoning of both projectors was first reduced using projector adjustments, and then the final alignment was achieved using the geometric calibration procedure.

It is also important to note the difference in the resolution reproduced by the LCD panel and the DLP projectors. The spatial resolution of the DLP projectors was half of the resolution of the LCD panel (refer to Table 5.1). In addition to that, the projected DLP images were blurred by the diffuser. The resulting blur was well approximated by a radially symmetric Gaussian function with $\sigma = 18.35$

[LCD pixels] for the bottom projector and $\sigma = 9.78$ for the top projector. The blur of the DLPs was compensated for by sharpening the image on the LCD panel.

Once the camera-based calibration is complete, the second step of the calibration procedure, the linearization, follows. In linearization, the display response curves (analogous to a gamma curve for CRTs) for both DLPs and the LCD panel are measured in turn using a JETI Specbos 1211 (JETI Technische Instrumente GmbH, Jena, Germany). To measure the display response curve of the LCD panel, the bottom HDR configuration is switched on to its maximum output, while the top HDR configuration is switched off. Display response curves are measured for each of the individual LCD channels, red, green, and blue, in turn. The display response curve is then measured with all three of the LCD channels simultaneously as a linearity check. The display response curve for each DLP is measured in turn, by setting the LCD panel to white (equal settings for red, green, and blue) at an appropriate transmissivity to measure the DLP display response curve, while switching off the other DLP.

The third and final step of calibration, spectroradiometric calibration follows. Spectroradiometric calibration was performed using a JETI Specbos 1211 (JETI Technische Instrumente GmbH, Jena, Germany) or Cambridge Research Systems SpectroCAL MKII Spectroradiometer (Cambridge Research Systems, Kent, UK). The spectroradiometric calibration measures the spectrum of each effective primary in turn, with the display set to an appropriate intensity to allow reliable measurements of the spectrum of each effective primary (see Sec. 2.3). The spectroradiometric calibrations were additionally verified using a SpectraScan PR-670 spectroradiometer (Photo Research, Syracuse, NY).

For the calibration results shown here, for all calibration steps, the calibration device (i.e. the camera for the camera-based calibration or the spectroradiometer for the linearization and spectroradiometric calibration), was placed at the "sweet spot" for the MPHDR display. This means that the calibration device was aligned at the point of maximal overlap of the primaries from the top HDR configuration and the primaries from the bottom HDR configuration. The calibration device was at a height defined by the sweet spot and a distance of 45 cm from the LCD panel. The device was oriented so its optical axis was perpendicular to the LCD panel.

5.2.3 Display algorithm

To drive the display, it is important to consider the control the user would want over the stimulus. To design cone, rod, and melanopsin isolating stimuli the experimenter would want to know, at each spatial location, the full spectrum the observer sees when viewing the display.

The spectral output of the MPHDR, $s(\lambda)$, can be thought of as the spectral emission from each DLP, $E_1(\lambda), E_2(\lambda)$, passed through the LCD with a certain transmission spectrum for each of the three channels in the LCD, $T_R(\lambda), T_G(\lambda), T_B(\lambda)$, as in Eq. 5.2.

For each spatial location, the experimenter has control of five independent parameters of the MPHDR display: the intensity of each of the DLPs, k_{E_1}, and k_{E_2}; and the intensity of each channel of the LCD, k_{T_R}, k_{T_G}, and k_{T_B}, forming the parameter vector κ as in Eq. 5.1, and depicted schematically in Fig. 5.3:

$$\kappa = [k_{E_1}, k_{E_2}, k_{T_R}, k_{T_G}, k_{T_B}]. \tag{5.1}$$

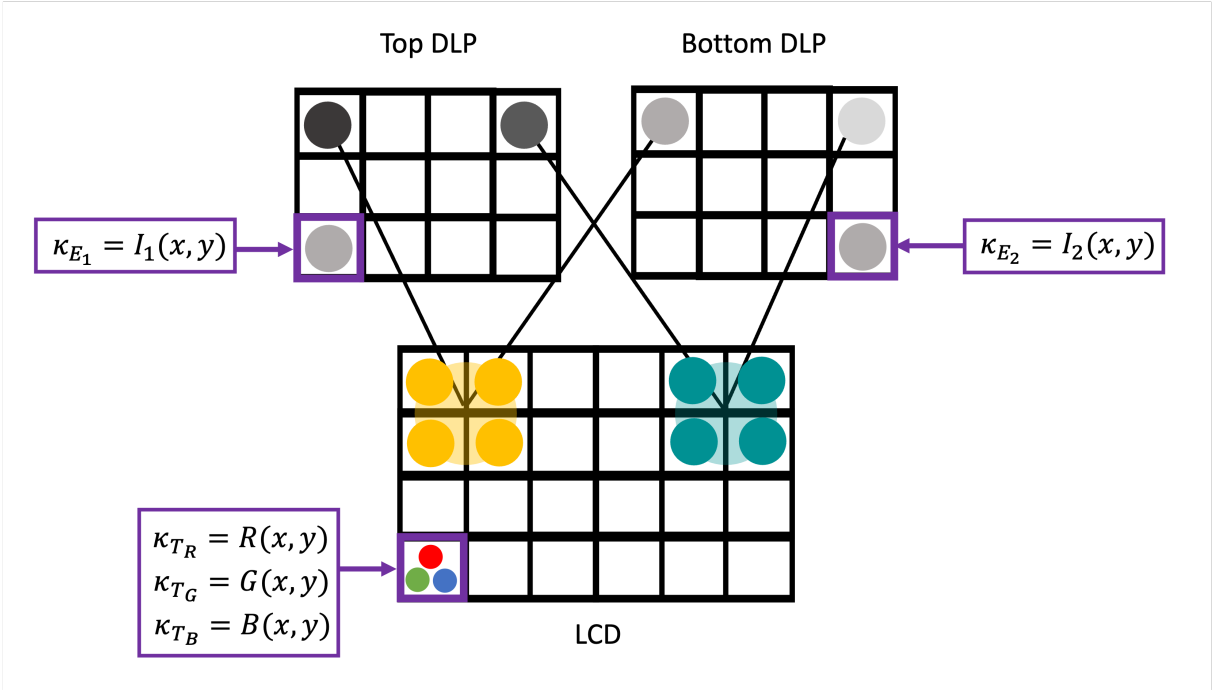


Figure 5.3: Schematic of the control the experimenter has over the MPHDR display. The experimenter can set the five independent parameters $\kappa = [k_{E_1}, k_{E_2}, k_{T_R}, k_{T_G}, k_{T_B}]$ where k_{E_1}, and k_{E_2} are the intensity of each of the DLPs; and k_{T_R}, k_{T_G}, and k_{T_B} are the intensity of each channel of the LCD. Thus, one essentially can change the achromatic intensity of each DLP (with spatial extent), and can control the intensity of each, R, G, B colour channel of the LCD (again with spatial extent). As the DLP has a lower resolution than the LCD, the DLP image is blurred, as described in Section 1.2.2, which results in an effective downsampling of the LCD resolution to achieve spectral-spatial control as described here. Note, while the DLPs are depicted to be achromatic here, in reality each DLP is filtered through a pair of filters which will impact their effective spectral output (but one can only vary the intensity, not the spectral composition, of that output through the display algorithm).

The value of all of these parameters can vary from effectively zero to 1. These parameters can never be 0 as this would result in a display with infinite contrast, which is not possible with this type of display configuration. Here, I use 0.0005 as the minimum driving value as it is found in practice that when each of these parameters is set to 0.0005 the output of the LCD transmission and projector

emission is equivalent to when they are driven with an 8-bit driving value of 0 i.e. this is their "black level" output.

The spectral output at each spatial location of the MPHDR is thus:

$$s(\lambda) = \sum_i \sum_j k_{E_i} k_{T_j} E_i(\lambda) T_j(\lambda) \tag{5.2}$$

where

$$i \in \{1,2\}$$

$$j \in \{R,G,B\}.$$

If one could measure the spectral transmission of each of the channels of the LCD and the emission spectra of the projectors, then one can generate the output spectra. However, due to physical limitations of the display, it is tricky to get a reliable measure of these spectra, $E_i(\lambda)$ and $T_j(\lambda)$. What can be easily and reliably measured are the effective primaries one gets as the output of the display, when the emission spectra of the projectors have passed through the spectral transmission of the LCD panel. One can consider six effective primaries, P_n, from the display, as shown below in Eq. 5.3.

$$P_n(\lambda) = E_i(\lambda) \cdot T_j(\lambda) \tag{5.3}$$

where

$$i \in \{1,2\}$$

$$j \in \{R,G,B\}$$

$$n \in \{1,2,3,4,5,6\}.$$

Therefore, written out in full form in Eq. 5.4, the output spectra from the display can be modelled as:

$$s(\lambda) = k_{E_1}\left(k_{T_R}P_1(\lambda) + k_{T_G}P_2(\lambda) + k_{T_B}P_3(\lambda)\right) + k_{E_2}\left(k_{T_R}P_4(\lambda) + k_{T_G}P_5(\lambda) + k_{T_B}P_6(\lambda)\right). \tag{5.4}$$

Where, for example, one can measure $P_1(\lambda)$ by setting k_{T_R}, $k_{E_1} = 1$ and k_{T_G}, k_{T_B}, k_{E_2} to ~ 0. In

practice, the measurement of $P_1(\lambda)$ will include some small contribution from the black levels of the remaining LCD channels and DLP (hence, why in practice, k_n is set to 0.0005). Analogous settings can be chosen to measure $P_2(\lambda)$, $P_3(\lambda)$, $P_4(\lambda)$, $P_5(\lambda)$, and $P_6(\lambda)$, resulting in a space where the user has six effective primaries (as shown in Fig. 5.4(a)), that they can control through five independent parameters to generate the spectral output from the MPHDR display (Fig. 5.4(b)).

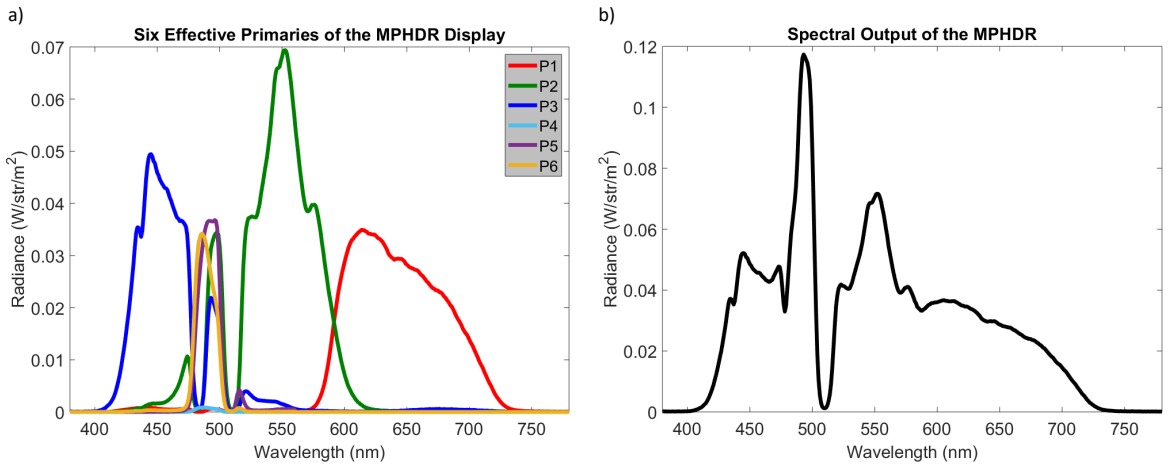


Figure 5.4: a) The six effective primaries available from the MPHDR display. Note, the spectra of primaries, $P_4(\lambda)$, $P_5(\lambda)$, and $P_6(\lambda)$, are measured from 380 nm to 560 nm given the transmission properties of the bandpass cut-off filter placed in the top HDR configuration. b) The spectral output of the MPHDR display when all six effective primaries are set to maximum.

The code to run the MPHDR display uses custom built drivers written in OpenGL (Khronos Group, Beaverton, OR, USA) and MATLAB (MathWorks, Natick, MA, USA). The experimenter can program an experiment using Psychtoolbox (Brainard, 1997), or other appropriate libraries, and simply render individually to each of the three components of the MPHDR. The experimenter will specify a pattern with colours defined as $[k_{T_R}, k_{T_G}, k_{T_B}]$ to send to the LCD screen, a pattern with colours defined as $[k_{E_1}, k_{E_1}, k_{E_1}]$ to send to DLP_1, and a pattern with colours defined as $[k_{E_2}, k_{E_2}, k_{E_2}]$ to send to DLP_2. The parameters $k_{T_R}, k_{T_G}, k_{T_B}, k_{E_1}, k_{E_2}$ are relative linear colour values.

5.2.4 Designing photoreceptor isolating stimuli

To generate pairs of photoreceptor-isolating stimuli, one needs to adjust the parameters of the MPHDR display system. Given the six effective primaries of the MPHDR, one cannot find a unique solution for the intensities of the individual primaries to design photoreceptor-selective spectral pairs. This therefore becomes an optimisation problem. Since the six primaries are controlled by five parameters one seeks two parameter vectors κ_1 and κ_2. These refer to five specific settings for the parameters described in the section above. These two parameter vectors are defined as in Eq. 5.5:

$$\kappa_1 = [k_{E_1;1}, k_{E_2;1}, k_{T_R;1}, k_{T_G;1}, k_{T_B;1}]$$

and

$$\kappa_2 = [k_{E_1;2}, k_{E_2;2}, k_{T_R;2}, k_{T_G;2}, k_{T_B;2}] \tag{5.5}$$

producing a photoreceptor-selective, or "isolating" spectral pair $s_1(\lambda)$ and $s_2(\lambda)$.

The defining characteristic of this spectral pair is that when alternating between them, the activation of only one photoreceptor class is changing (the "isolated" class), while the activation of the other photoreceptor classes (the "silent" classes) stays the same. The spectral sensitivities of the human photoreceptors are well-established and denoted here as $R_L(\lambda)$, $R_M(\lambda)$, $R_S(\lambda)$, $R_R(\lambda)$ and $R_I(\lambda)$, for the L, M and S cones, the rods, and melanopsin, respectively.

The activation of a photoreceptor $R(\lambda)$ to a given spectrum $s(\lambda)$ is given as in Eq. 5.6:

$$A = \int_{\lambda=380}^{780} R(\lambda)s(\lambda)\,d\lambda. \tag{5.6}$$

One can then calculate the Weber contrast for a given photoreceptor class as the percent difference in activation experienced under $s_1(\lambda)$ and $s_2(\lambda)$. One of these spectra serves as the background spectrum, $s_1(\lambda)$, giving activation A_1, and the other as the modulation spectrum, giving activation A_2. The Weber contrast between background and modulation, i.e. the percent difference in activation experienced under $s_1(\lambda)$ and $s_2(\lambda)$, is defined as in Eq. 5.7:

$$C = \frac{A_2 - A_1}{A_1}. \tag{5.7}$$

The two spectra are selected so that the contrast of the isolated photoreceptors is $C_{\text{isolated}} > 0$ and the contrast of the silent photoreceptors is $C_{\text{silent}} = 0$.

For example, for a melanopsin-isolating modulation: $C_I > 0$, and $C_L = C_M = C_S = C_R = 0$. In practice, it is useful to allow for some tolerance, and accept modulations which are $C_I > 0$, and $C_L < \eta$, $C_M < \eta$, $C_S < \eta$, $C_R < \eta$, where η is an acceptable 'splatter' contrast. Alternatively, another metric, such as Euclidean pooled cone contrast could be subject to a constraint: $\sqrt{C_L^2 + C_M^2 + C_S^2} = 0$ for the exact case or $\sqrt{C_L^2 + C_M^2 + C_S^2} < \eta$ for the 'tolerant' case. For modulations that are cone-silent, instead of optimization in the photoreceptor space, one can also optimize in a derived space, such as the XYZ coordinates calculated from the LMS cone activations, or some other colour space.

Here, the relationship between background and modulation is expressed as the difference between

the log transformed activations (Eq. 5.8):

$$\log A_2 - \log A_1. \tag{5.8}$$

To find photoreceptor-isolating modulations, one can optimize over κ_1 and κ_2 to maximise the quantity in Eq. 5.9:

$$\log(A_{2;\text{isolated}}) - \log(A_{1;\text{isolated}}). \tag{5.9}$$

The optimization can then be described as (Eq. 5.10):

$$\underset{\kappa_1,\kappa_2}{\text{argmax}} \log(A_{2;\text{isolated}}) - \log(A_{1;\text{isolated}}). \tag{5.10}$$

The optimization is subject to the constraint in Eq. 5.11:

$$(\log(A_{2;\text{silent}}) - \log(A_{1;\text{silent}}))^2 = 0, \tag{5.11}$$

or for the soft constraint version (Eq. 5.12):

$$(\log(A_{2;\text{silent}}) - \log(A_{1;\text{silent}}))^2 < \epsilon. \tag{5.12}$$

The above procedure does not constrain luminance or chromaticity of the background and/or modulation spectra but simply produces a spectral pair with the largest contrast on the isolated photoreceptors. In situations where it is desirable to fix the background activations at some desired level, both the optimisation target and the constraint can be written as the difference to the desired set of activations.

5.3 Results

Here I present measurements taken from the MPHDR display that demonstrate the key benefits of the display over existing display systems and the ability of the MPHDR display to generate photoreceptor isolating stimuli. Measurements were taken using either the Cambridge Research Systems SpectroCAL MKII Spectroradiometer, the JETI Specbos 1211 (JETI Technische Instrumente GmbH, Jena, Germany), or the SpectraScan PR-670.

5.3.1 High-dynamic range and high luminance output

Typical CRT monitors may reach a maximum luminance of approximately 100 cd/m^{2}while typical LCD monitors may reach a maximum luminance of approximately 300 cd/m^2. The MPHDR display can reach a maximum luminance of 3200 cd/m^2. The luminance contribution from each of the HDR configurations in our set-up is not identical, with the bottom HDR configuration contributing a maximum luminance of 3000 cd/m^{2}and the top HDR configuration contributing a maximum luminance output of 200 cd/m^2. Both projectors and the LCD are subject to display response correction (analogous to gamma correction for CRTs), as shown in Fig. 5.5(a)-(c). After display response correction the MPHDR behaves linearly (Fig. 5.5(d)).

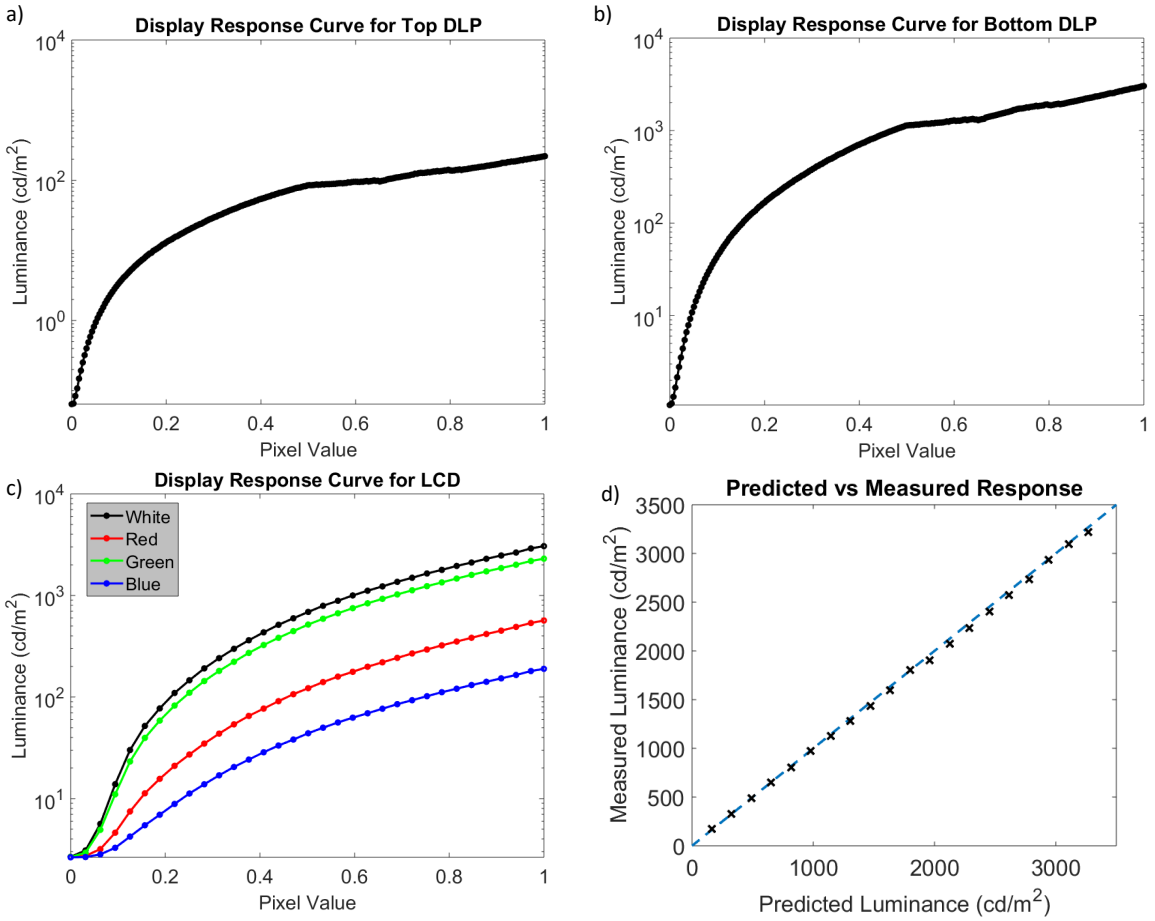


Figure 5.5: a) The display response curve for the top projector. b) The display response curve for the bottom projector. c) The display response curve for the LCD panel. d) Predicted and measured luminances (black cross symbols) match after display response correction.

The full on/off contrast range and ANSI contrast range of each of the individual components (the LCD, and each DLP) and configurations (each HDR configuration, and the full MPHDR set-up) of the MPHDR display were calculated by measuring the maximum and minimum luminance, L_{max} and

L_{min}, of the LCD, T, and each DLP, E_1 and E_2, in turn. The full on/off contrast range is the dynamic range achievable from a uniform black screen to a uniform white screen, while the ANSI contrast range is the dynamic range achievable in practice when a 2x2 white and black checkerboard pattern is displayed. Both the full on/off and ANSI contrast range, D, for each component and configuration can be calculated as described in Eq. 5.13:

$$D = \frac{L_{T_{max}}}{L_{T_{min}}} \cdot \frac{(L_{E_{1_{max}}} + L_{E_{2_{max}}})}{(L_{E_{1_{min}}} + L_{E_{2_{min}}})}, \tag{5.13}$$

where the choice of the parameter vectors κ for the luminance measures were as follows:

$$L_{T_{max}} : \kappa = [0,0,1,1,1]$$

$$L_{E_{1_{max}}} : \kappa = [1,0,0,0,0]$$

$$L_{E_{2_{max}}} : \kappa = [0,1,0,0,0]$$

$$L_{T_{min}} : \kappa = [0,0,0.0005,0.0005,0.0005]$$

$$L_{E_{1_{min}}} : \kappa = [0.0005,0,0,0,0]$$

$$L_{E_{2_{min}}} : \kappa = [0,0.0005,0,0,0].$$

For the individual components one can simply calculate the dynamic range as maximum measured luminance divided by minimum measured luminance from the component. For the individual HDR configurations Eq. 5.13 can be applied, where the luminance of the DLP not driving that configuration is simply zero. Both the full on/off and ANSI contrast range of the individual components and configurations and the full MPHDR display are reported in Table 5.2. The MPHDR display has a full on/off contrast range of 3,240,000:1 and an ANSI contrast range of 341,000:1.

5.3.2 Expanded colour gamut

The maximum colour gamut available from the MPHDR display is wider than the colour gamut provided by traditional sRGB displays shown in the CIE xy chromaticity diagram in Fig. 5.6(a). The two HDR configurations also contribute differently to the gamut due to the filters chosen for the DLPs, depicted in Fig. 5.6(a) and (b). To further visualize the expansion of the gamut, Fig. 5.7 shows the full gamut the MPHDR is able to achieve in CIE xyY chromaticity space, in both log scale (Fig. 5.7(a)) and linear scale (Fig. 5.7(b)) for luminance. Visualising the gamut in CIE xyY chromaticity space is useful to

Table 5.2: Table comparing the full on/off (global) and ANSI (local) contrast range of the individual components and configurations of the MPHDR display compared to the full MPHDR display configuration. Note, the top HDR configuration has a higher global but lower local contrast relative to the bottom HDR configuration, as these measurements are somewhat dependent on viewing geometry, which is asymmetric for the two projectors due to their orientations. As the priority for vision science is to characterise the display from the viewing geometry of the observer, here the measurements are made from the "sweet spot" location, as described in the section on calibration.

Display	**Full on/off (global) contrast**	**ANSI (local) contrast**
LCD Panel	1140:1	725:1
Top DLP	3430:1	100:1
Bottom DLP	2810:1	594:1
Top HDR Configuration	3,930,000:1	72,400:1
Bottom HDR Configuration	3,220,000:1	431,000:1
MPHDR	3,240,000:1	341,000:1

see how the available gamut changes as a function of luminance.

Using the CIE xy and CIE xyY chromaticity diagrams is a helpful tool in understanding the expanded colour gamut available from the MPHDR display, but it does not capture the full expansion of the gamut, as the CIE xy diagram is limited to L, M, and S cone responses. To truly gauge the expansion of the gamut, one needs to consider the expansion in rod and melanopsin space as well as cone space. Bar charts, showing the relative responses of each photoreceptor type, when each primary in turn is set to maximum while all other primaries are switched off are shown in Fig. 5.6(c), to more completely visualise the expanded gamut of the MPHDR display. The primaries $P_4(\lambda)$, $P_5(\lambda)$, and $P_6(\lambda)$, in particular allow additional modulation of rod and melanopsin responses. An even more complete visualization of the MPHDR, and five-primary displays more broadly, is presented in Chapter 6.

5.3.3 High-contrast photoreceptor-isolating stimuli

Using the algorithm described in the methods section (Sec. 2.3 and Sec. 2.4), one is able to produce photoreceptor isolating stimuli with the display. To demonstrate this, I consider the example of designing melanopsin isolating stimuli, as the filters used to filter the spectra of the DLPs were chosen to optimize the available melanopsin contrast and gamut, meaning that in its current form the MPHDR display is best suited to melanopsin isolating experiments.

The MPHDR is able to generate metameric pairs of melanopsin isolating stimuli such that the cones are silenced while the relative response of melanopsin to the two spectra differs. The spectra of the desired metamers can be displayed reliably on the MPHDR. An example case demonstrating this is shown in Fig. 5.8(a). In this instance rod activation is ignored.

The MPHDR is able to achieve a maximum melanopsin contrast of 117% at a luminance of 75 cd/m^2. Available melanopsin contrast decreases exponentially as a function of luminance, as depicted in Fig. 5.8(b). Even at a high luminance level of 2000 cd/m^{2}the MPHDR is still able to achieve a melanopsin contrast of 23%. Whilst our metamers should be identical to the cones (i.e. cone contrasts of 0%). In practice there are small cone contrasts between the metamers due to measurement uncertainty in the display spectral output (particularly in the wavelength range where there is low luminance, i.e. the wavelengths blocked by the filters). At luminance levels between 75 cd/m^2- 100 cd/m^{2}our metamers (at maximum melanopsin contrast) produce 7% cone contrast; at luminance levels between 500 cd/m^2- 1500 cd/m^{2}the cone contrasts are < 1%; and at other available luminance levels the cone contrasts are < 3%. If the experimenter wished to completely silence the cones, or to generate metamers at lower

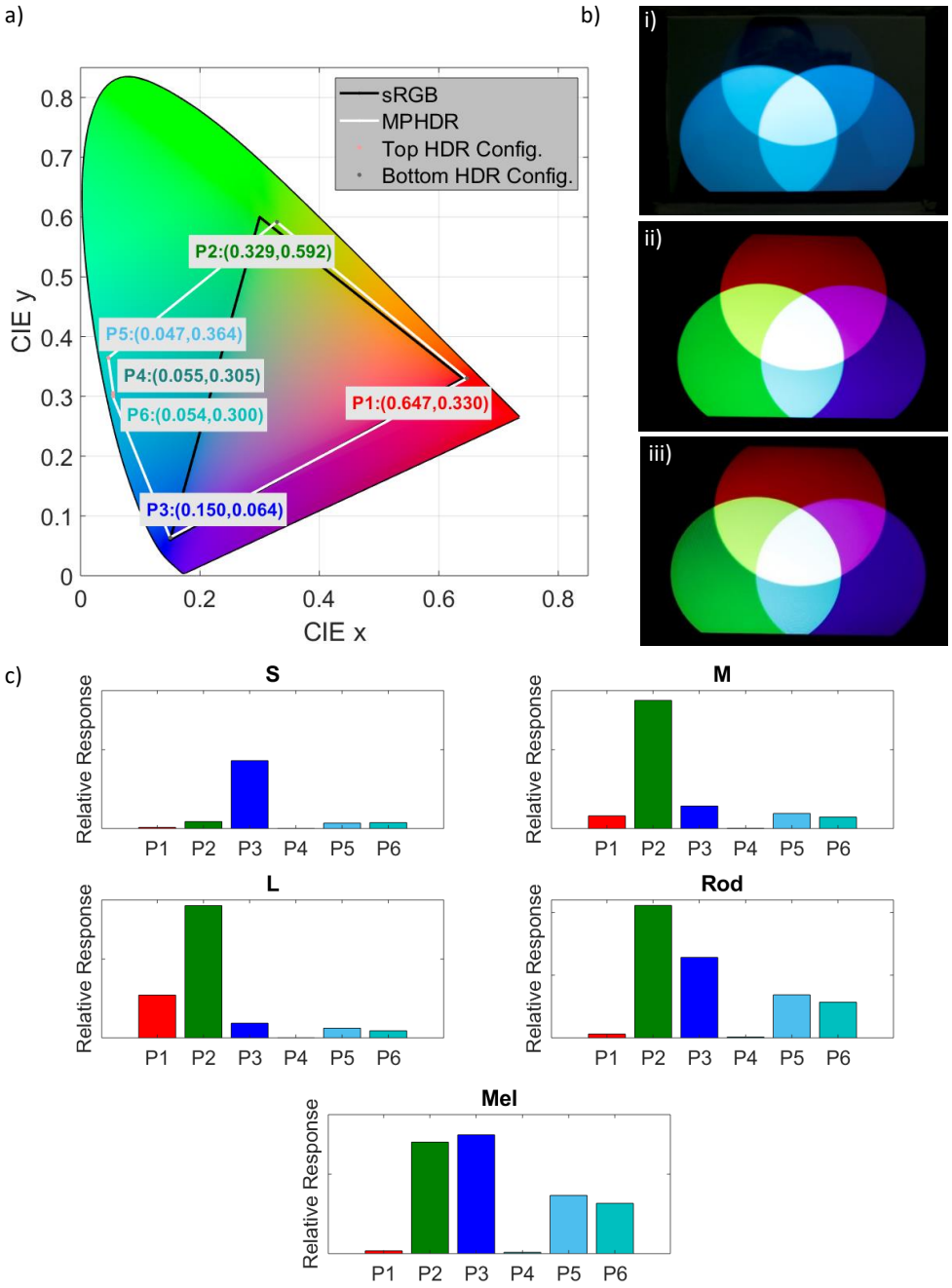


Figure 5.6: a) The colour gamut available from the MPHDR display, broken down into the two HDR configurations when all six effective primaries are available, expressed using the CIE 1931 xy chromaticity coordinates. The sRGB gamut is shown for comparison. b) Photographs of the display when i) only the top HDR configuration is on (for the circles top, bottom-left, and bottom right respectively: $\kappa = [1,0,1,0,0]$, $\kappa = [1,0,0,1,0]$, $\kappa = [1,0,0,0,1]$), ii) only the bottom HDR configuration is on ($\kappa = [0,1,1,0,0]$, $\kappa = [0,1,0,1,0]$, $\kappa = [0,1,0,0,1]$), iii) the MPHDR when both HDR configurations are contributing ($\kappa = [1,1,1,0,0]$, $\kappa = [1,1,0,1,0]$, $\kappa = [1,1,0,0,1]$). Note the fringes seen in b.iii) have been introduced by the camera's optics and are not visible on the display. c) Bar charts showing the relative response of each photoreceptor type to each primary, normalized for each photoreceptor class.

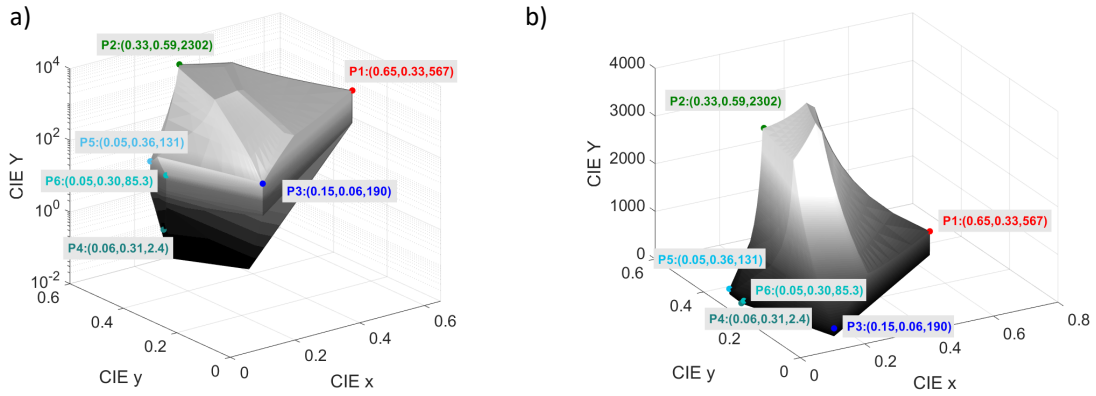


Figure 5.7: a) 3D colour gamut in CIE xyY chromaticity coordinates, plotted in log scale for the luminance axis. b) 3D colour gamut as in (a), shown again here in CIE xyY chromaticity coordinates, in linear scale for the luminance axis.

luminance levels for specific stimuli, the spectral output of the display could be fine-tuned. One way this might be achieved is during calibration measurements to simply set the spectral output to zero at any of the theoretically blocked wavelengths according to the filter response spectra.

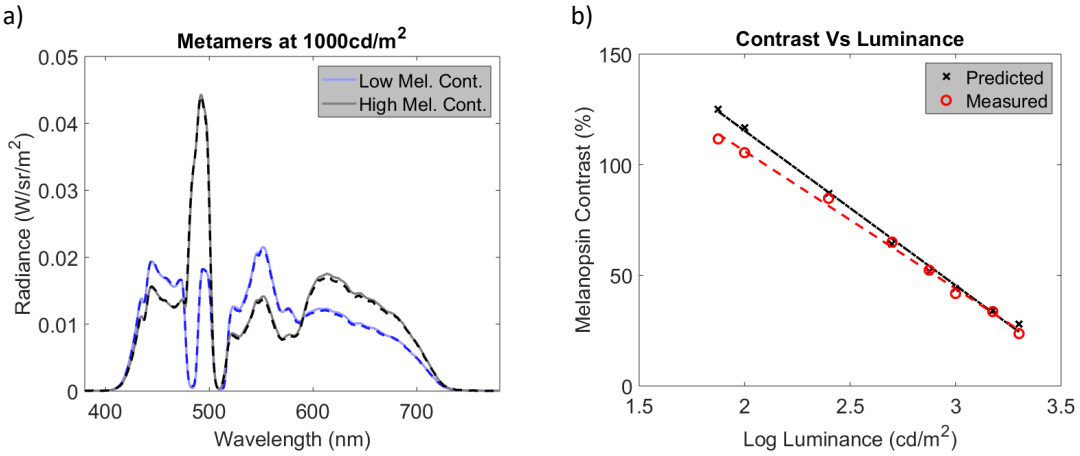


Figure 5.8: a) An example of a metameric pair of melanopsin isolating stimuli. The measured spectral output from the MPHDR display (shown as a dashed line) overlaps the desired spectral output from the MPHDR display (shown as a solid line) generated using the display algorithm. b) Available melanopsin contrast as a function of luminance, for both the measured contrast and predicted contrast.

5.4 Discussion

The goal of vision research is to understand how the human visual system responds to and integrates visual stimuli. To do so, vision researchers need a method to generate controlled and adjustable stimuli. Three common features of a stimulus that the experimenter may want to control are the spatio-temporal properties of the stimulus, the luminance and/or contrast range of the stimulus, and

the spectral content of the stimulus. Before the implementation of the MPHDR display, there was no display system that could give the experimenter complete control of all three of these features, meaning the experimenter always had to compromise.

For instance, to ask questions about the spatial resolution of the visual system, one needs to use a display that offers spatial control, such as a conventional CRT or LCD monitor (Carpenter & Robson, 1998). Such displays also offer intrinsic temporal control of the stimulus at the frame rate. However, a CRT or an LCD monitor would not be appropriate for an experiment aiming to probe the visual system across the full range of photopic light levels, and one may need to opt for a Maxwellian-view set-up for such an experiment (Westheimer, 1966). Multi-channel Maxwellian-view systems can also allow the experimenter control of more than three primaries, and could be used to achieve photoreceptor isolation, something that conventional three-primary displays such as CRTs cannot achieve (Cao et al., 2015; Pokorny et al., 2004). Such Maxwellian-view systems, however, are limited in their spatio-temporal control. Multi-primary spatially controllable stimuli are achievable through projector-based display systems, but such systems can only generate low-dynamic range stimuli (Packer et al., 2001; Ajito et al., 1999). Development of HDR display systems has improved the ability of vision researchers to conduct experiments at high luminance levels and high-contrast in a spatially controllable display, however existing systems are of this type have been limited to only three primaries (Seetzen et al., 2004).

The limitation of existing display systems means that the vision researcher has always had to make a choice: spatio-temporal control, high-dynamic range, multi-primary; pick two. The MPHDR display that I present in this chapter is the first system, to my knowledge, that offers the vision researcher control of all three desirable controls of a stimulus, as summarized in Table 5.3.

The high-dynamic range of the new display comes from replacement of the standard backlight from an LCD panel with a DLP projector, as described in earlier HDR systems (Seetzen et al., 2004). In addition to expanding the dynamic range, allowing the MPHDR to achieve a full on/off contrast ratio of $3,240,000:1$ and an ANSI contrast range of $341,000:1$, this setup also allows for extremely bright displays, with the MPHDR display reaching a peak luminance output of 3200 cd/m^2. The benefit of the high-dynamic range component of the MPHDR display is that it allows stimuli to be presented with a dynamic range and absolute light level that is closer to the range seen in real-world environments.

The multi-primary nature of the new display comes from adding in a second HDR configuration to the standard HDR set-up. The benefit of the multi-primary component of the MPHDR display is twofold. First, it allows for a wider colour gamut than traditional displays. Second, and crucially, it

Table 5.3: Table comparing features of the different systems available for vision experiments. The MPHDR display is the only system to our knowledge to incorporate all three key features: high-dynamic range, spatio-temporal control, and more than three primaries.

Display	High dynamic range	Spatio-temporal control	>3 primaries
CRT monitors	✗	✓	✗
LCD monitors	✗	✓	✗
HDR displays	✓	✓	✗
Multi-primary Maxwellian view systems	✓	✗	✓
Multi-primary projector based displays	✗	✓	✓
The multi-primary high dynamic range (MPHDR) display	✓	✓	✓

allows for independent control of the five classes of human photoreceptors on a spatially controllable high-dynamic range system, creating a new branch of possible experiments into the function of individual photoreceptor classes.

A key benefit of the design of the MPHDR display described in this chapter is that it is affordable and achievable. Any vision researcher who wishes to build such a display for their own experiments, should be able to follow the approach outlined above and construct their own MPHDR display using off-the-shelf components.

5.5 Future development

The model for driving the MPHDR takes as input the five parameters that the experimenter can control, namely the intensities of each DLP and the intensities of each channel of the LCD, and gives a spectral output, which can be considered as the addition of the six effective primaries of the MPHDR, as specified in Eq. 5.4. One may note the fact that the MPHDR does not have as many controllable independent parameters as it has effective primaries. This is due to the limiting factor of the LCD colour channels. Once one has set a value for say the red channel of the LCD, the range of settings for the "red" primary from both the top and bottom HDR configurations is constrained. The only further

control one then has of the value for these two "red" primaries is the intensity from each DLP.

To move towards an MPHDR with fully independent control of the six output primaries, the following alternative approaches could be considered:

1. **Temporal blurring:** One solution would be to modulate which DLP is illuminating the LCD at any given time. Currently the LCD has a frame rate of 60 Hz. If only one of the DLPs was turned on for one frame of the LCD to produce three independent output primaries, and then the other DLP turned on during the following frame, one could change the LCD settings between the two frames to produce an additional three independent output primaries. This would reduce the effective frame rate of the display to 30 Hz, as pairs of frames would be treated as one temporally blurred frame. To ensure synchronicity between the DLPs and LCD, and to ensure sharp transitions between the two DLP illuminations, one would likely opt for a shutter device to control the on/off switch of the DLPs. A 30 Hz refresh rate would be sufficient for some vision science experiments, but would be limiting for others.

2. **Achromatic LCD and colour DLPs:** Another alternative approach for fully independent six primary control would require a fundamental change to the hardware design presented in this chapter. In the set-up presented here the colour wheels of the DLPs have been removed, rendering the spectra from the projectors spatially uniform, with spatial control only of their intensity, while the LCD has chromatic transmission filters, namely the RGB channels. If a set-up instead kept the colour wheels of the DLPs intact and projected through an achromatic LCD, one could have independent control of the chromatic properties via the DLPs instead of the LCD. The concern for this option is that the colour wheels cause significant light loss in the DLP, so one would need an expensive high-end DLP to achieve the high-dynamic range and luminance output that is possible with the current approach.

3. **Custom designs:** A custom fully independent multi-primary HDR display could be developed by creating an LCD panel with four or more sub-pixels; by displaying multiple backlight colours in a single frame (Kauvar et al., 2015); or by adding colour channels to HDR light steering projectors (Damberg et al., 2016). Such displays, although feasible, would require custom manufacturing and substantial engineering effort. This is in contrast to the MPHDR design I describe here, which can be built from in-expensive off-the-shelf components and with limited engineering experience.

For the purpose of running photoreceptor isolating experiments on the display via silent substitu-

tion, our five-parameter control of the spectral output of the MPHDR is sufficient. It is nonetheless interesting to think about how one would develop a fully independent six primary MPHDR display.

Another area for future work is a consideration of the spatial-spectral calibration of the display. To find the "sweet spot" of the display the current procedure is to visually inspect the display to find the location from which the display appears brightest. Then, using a photometer, luminance is measured from that location [to the centre of the display], and from neighbouring locations [to the centre of the display] to find the viewing position that gives peak luminance. The location of the photometer for the maximum measurement is noted and this is where the observer's pupil is aligned. This is also the spot from which the spatial calibration is performed, and thus the display should be calibrated to this position. To ensure this position can be reliably returned to, the viewing axis can be marked by two spatially offset LEDs that can be placed in front of the display. The observer/calibration device should then be placed such that the two LEDs visually align. To help the observer maintain this location, one can use an eye piece that the observer must look through, and provide instructions to check periodically that the LEDs remain visually aligned. At present it has not been quantified how much variability in eye position happens in practice, and the consequences of this on display uniformity. To address this, in future work, I will use a hyperspectral camera to simultaneously measure the spectral-spatial uniformity of the display from the "sweet spot" and slight perturbations from it. From a visual check, there does not appear to be extreme spectral-spatial differences around the "sweet spot" position, but a proper quantification of this is left to future work.

It is also worth noting that the other issue that may occur in practice is the interaction between retinal inhomogenities and photoreceptor isolating stimuli presented on the display. One may wish to characterise the spectral output of the display as a function of retinal eccentricity, such that stimuli can be controlled independently for different retinal locations. Such spectral control should in theory be possible, and the implementation of this in practice is left to future work. Again, a hyperspectral camera will be useful to characterise the spatial-spectral output of the display for this purpose. Some retinal inhomogenities, such as quantified field size differences in spectral sensitivities (CIE, 2006), may be simpler to compensate, while others, such as retinal vessels, may be more challenging, and may cause difficulties in producing the desired photoreceptor isolation (Spitschan et al., 2015; Spitschan & Woelders, 2018).

5.6 Conclusions

In this chapter, I present, to my knowledge, the first implementation of a multi-primary high-dynamic range display system for use in vision experiments. The MPHDR display has five independent parameters that allow for the control of six effective primaries from the display. During the characterization of the MPHDR display, we achieved a maximum luminance output of 3200 cd/m^2, a full on/off contrast of 3,240,000:1, an ANSI contrast of 341,000:1, and an expanded colour gamut. I believe that this system will be useful for many different psychophysical experiments, particularly for modulation of individual photoreceptor classes. As an example case, I generated a range of metameric pairs of melanopsin isolating stimuli across different luminance levels, from an available melanopsin contrast of 117% at 75 cd/m^{2}to a melanopsin contrast of 23% at 2000 cd/m^2. I also envisage that new schemes in display architectures (such as temporal blurring and achromatic pixelated active masks with coloured backlighting) will enable full independent control of multiple primaries in a HDR display.

5.7 Acknowledgements

The work presented in this chapter was undertaken as a collaborative project, led by ACH, alongside Dr Manuel Spitschan, and Prof. Hannah Smithson, from the Department of Experimental Psychology at the University of Oxford, and Dr Ali Özgür Yöntem, and Dr Rafal Mantiuk, from the Department of Computer Science at the University of Cambridge. The text in this chapter is largely adapted from a publication (Hexley, Yöntem et al., 2020), for which ACH was awarded the 2020 JOSA A Emerging Researcher Best Paper Prize. The original text for this publication was written largely by ACH, with original text contributions also made from Dr Spitschan and Dr Yöntem. All authors made editorial contributions to the text in the publication. I thank all my co-authors on this project for their invaluable inputs to this work, and I would like to give a particular thanks to Dr Yöntem, as without his engineering expertise and practical help in building the display, the project would not have been possible.

Chapter 6

Novel metrics for characterizing the reproduction of photoreceptor signals

Contents

6.1 Introduction 216

6.1.1 The human physiological response to light 216

6.1.2 The reproduction of light by displays 218

6.1.3 Towards more accurate physiologically relevant reproduction of light on displays 221

6.1.4 Quantifying a display's physiologically relevant reproduction of light 222

6.2 Methods: Introducing three novel metrics and an equal-luminance photoreceptor excitation diagram 223

6.2.1 Metric I: Photoreceptor Signal Reproduction Metric (PSRM) 225

6.2.2 Metric II: Photoreceptor Signal Distortion Metric (PSDM) 226

6.2.3 Metric III: Photoreceptor Correlation Distortion Metric (PCDM) 227

6.2.4 Equal-luminance photoreceptor excitation diagram 229

6.3 Results: Example applications of novel metrics 231

6.3.1 Photoreceptor Signal Reproduction Metric (PSRM) 232

6.3.2 Photoreceptor Signal Distortion Metric (PSDM) 234

6.3.3 Photoreceptor Correlation Distortion Metric (PCDM) 237

6.3.4 Equal-luminance photoreceptor excitation visualizations 237

6.4 Discussion 240

6.4.1 The applicability of the proposed metrics 240

6.4.2 Limitations of proposed metrics 241

6.4.3 Applying the metrics and visualization to multi-primary displays 244

6.5 Conclusion 245

6.6 Acknowledgements 245

6.1 Introduction

As explained in Chapter 1, the field of colorimetry, which quantifies the colour reproduction capabilities of displays, is well established. Yet, there are currently no metrics to characterise a display's ability (or inability) to reproduce the full set of five photoreceptor signals to light. As it becomes clearer that rod and ipRGC function may not be so readily ignored as traditionally believed, display manufacturers who continue to aspire towards creating hyper-realistic displays will need to begin to consider the rod and ipRGC reproduction capabilities of their systems. Metrics that characterize the full physiological response to light, beyond traditional colorimetry, are necessary to guide such development. Further, visualization methods are needed to illustrate the five-dimensional photoreceptor space. In this chapter, I describe three novel physiologically relevant metrics that I propose can be used to characterize a display's ability to reproduce all five photoreceptor signals. Additionally, I propose an equal-luminance photoreceptor excitation diagram, as an extension of the MacLeod-Boynton chromaticity diagram (MacLeod & Boynton, 1979), to allow for visualization of the full physiological sensor space.

6.1.1 The human physiological response to light

To understand the need to reproduce and quantify the reproduction of all five photoreceptor signals in order to characterize a display, one must look to the physiology behind the human light response. As described in detail in Chapter 1, the first step in the human response to light is the absorption of photons by photoreceptors in the retina. Photopic vision is an obvious example of a light-driven mechanism: photons are absorbed by the three different types of cones, the long-wavelength-sensitive (L), medium-wavelength-sensitive (M), and short-wavelength-sensitive (S) cones. Comparisons by post-receptoral neurons with centre-surround receptive fields support pre-processing of spatial variation. Comparisons between cone types with different spectral sensitivities support chromatic pre-processing.

The cones share the role of light sensation with the rods and the melanopsin-containing ipRGCs. The visual response to light, therefore, is more nuanced than the simple cone-driven model of photopic vision described above. Rods are well-known to contribute to visual function at low-light levels. While

rods have been traditionally thought to saturate in bright light conditions (Shapiro et al., 1996; Aguilar & Stiles, 1954) the exact light level at which rods begin to saturate, or the variability in this level between observers is not well-defined and can vary depending on field conditions, and recent evidence suggests a rod-response is still persistent even in light levels well into the photopic range (Sharpe & Stockman, 1999; Lall et al., 2010; Frederiksen et al., 2021). Even in the classical rod-saturation model, most office lighting levels would fall comfortably within the mesopic range, meaning rods are likely to contribute to much of our daily visual function (Stockman & Sharpe, 2006; Zele & Cao, 2015). There is also evidence that ipRGC responses may reach the visual cortex, and thus that melanopsin signals may also be contributing to some visual functions (Zaidi et al., 2007; Zele, Feigl et al., 2018; Spitschan et al., 2017; Cao et al., 2018; Brown et al., 2012; Zele, Adhikari et al., 2018).

Beyond vision, light is well-known to be one of the primary drivers of the entrainment of the internal circadian clock to the external light-dark cycle. Exposure to bright light in the evening and at night can suppress the production of melatonin, colloquially referred to as the "sleep hormone" (Blume et al., 2019; Brown, 2020; Duffy & Czeisler, 2009). Further, the spectral composition of light, not just the integrated radiance, is important for driving these circadian features, with the spectral sensitivity of these non-visual effects of light being consistent with the spectral sensitivity of melanopsin (Lucas et al., 2014; Brainard et al., 2001). Thus melanopsin excitation is often attributed as being the primary driver of these non-visual circadian effects of light. ipRGCs receive additional synaptic inputs from rods and cones (Dacey et al., 2005), and therefore rod and cone signals may also be contributing to the circadian effects of light.

ipRGCs are known to work alongside the cones to drive other non-visual responses to light. For example, pupil size is thought to be mediated by both melanopsin and cone signals (and perhaps, some rod signals) (Spitschan et al., 2017; Spitschan et al., 2019; Cao et al., 2015; Gamlin et al., 2007; Barrionuevo & Cao, 2016; Tsujimura et al., 2010; Tsujimura & Tokuda, 2011). L and M-cone signals combine synergistically with melanopsin signals to drive pupil size while S-cone and melanopsin signals influence pupil size in an opponent manner (Spitschan, 2019a; Spitschan et al., 2014). ipRGCs also play an important role in driving the light induced maintenance of alertness (Lok et al., 2018; Souman et al., 2018; Cajochen, 2007). Additionally, light exposure can impact mood, with light therapy being a treatment option for Seasonal Affective Disorder (SAD), and other mental health problems (Blume et al., 2019; Milosavljevic, 2019).

In sum, ipRGCs underlie key influences of light on physiology and behaviour. Rod and melanopsin signals are also potential contributors to visual function, even under conditions that have traditionally

been considered the sole domain of the cones (**Freeriksen3320**; Zaidi et al., 2007; Zele, Feigl et al., 2018; Spitschan et al., 2017; Cao et al., 2018; Brown et al., 2012; Zele, Adhikari et al., 2018; Lall et al., 2010). It is for these reasons that one must consider how to quantify the rod and melanopsin reproduction capability of displays, beyond traditional colorimetry.

6.1.2 The reproduction of light by displays

Most modern displays that make use of wide colour gamuts, high-resolution screens, and high-dynamic range, are capable of reproducing the spatial and chromatic characteristics of real-world scenes. Consider reproducing a single chromatic sample, such as the radiance spectrum shown in Fig. 6.1(b), on a display with primaries such as those in Fig. 6.1(c), as a simple uniform colour patch which occupies the entire field of view. To match the colour on the display to the colour from the real-world radiance spectrum one can simply calculate the relative intensity vector, $\vec{C} = [r, g, b]$, of the primaries of the display, $p_r(\lambda)$, $p_g(\lambda)$, and $p_b(\lambda)$, needed to generate a spectrum on the display (Fig. 6.1(d)) that reproduces the L, M, and S responses, $\vec{A} = [A_L, A_M, \text{and } A_S]$, from the real-world spectrum (where $\vec{A}$ is chosen to fall within the display gamut and intensity range), as in Eq. 6.1:

$$\vec{C} = B^{-1}\vec{A}$$

where

$$\vec{C} = \begin{bmatrix} r \\ g \\ b \end{bmatrix} \text{ and } \vec{A} = \begin{bmatrix} A_L \\ A_M \\ A_S \end{bmatrix} \tag{6.1}$$

and

$$B_{k,j} = \int_\lambda p_j(\lambda) \cdot\ T_k(\lambda)\, d\lambda$$

$$j \in [r, g, b]$$

$$k \in [L, M, S].$$

Note, the Stockman-Sharpe 10° cone fundamentals, $T_L(\lambda)$, $T_M(\lambda)$, $T_S(\lambda)$, as in the CIE S026 standard and depicted in Fig. 6.1(a), are used to calculate the L, M, and S responses (Stockman et al., 1999; Stockman & Sharpe, 2000; CIE, 2006, 2018). As the L, M, and S responses from both the display, $\vec{A}^*$, and the real-world spectrum, $\vec{A}$, will be equal, as shown in Fig. 6.1(e), the colour of the display and of the real-world spectrum is predicted to match. Conventional displays are designed to reproduce only the signals of the three cone classes using three primaries, equating the L, M, and S cone signals

from the real-world scene and the reproduced scene. As a consequence, the rod and melanopsin signals are uncontrolled in the reproduced scene. The reproduced rod and melanopsin signals, A_R^* and A_I^*, are determined by the spectral output, $P(\lambda)$, of the display, which is determined by the relative intensity vector $[r, g, b]$ set to reproduce the cone responses as described in Eq. 6.1. The distorted rod and melanopsin signals are described in Eq. 6.2, (note, the CIE S026 standard spectral sensitivity functions of rods, $T_R(\lambda)$, and melanopsin, $T_I(\lambda)$, are used here and depicted in Fig. 6.1(a) (Wyszecki & Stiles, 1982; Lucas et al., 2014; CIE, 2018)), and an example of how the signals are distorted when an example spectrum is displayed on the hypothetical three-primary display depicted in Fig. 6.1(c) is shown in Fig. 6.1(e):

$$\begin{gathered} P(\lambda) = r \cdot p_r(\lambda) + g \cdot p_g(\lambda) + b \cdot p_b(\lambda) \\ A_k^* = \int_\lambda P(\lambda) \cdot T_k(\lambda)\, d\lambda \\ \text{where} \\ k \in [R, I]. \end{gathered} \tag{6.2}$$

The inability of traditional displays to control rod and melanopsin excitation poses a problem as the world continues to move online and becomes heavily reliant on display technology in everyday life. As people may use computers, TVs, and phones well into the evening and night, the melanopsin signals driven by such evening light exposure, could lead to large disruptions in melatonin suppression (and subsequently sleep quality), circadian cycle, mood, and alertness in the general population (Arendt, 2006; Wood et al., 2006; Figueiro et al., 2011; Cajochen et al., 2011; Wood et al., 2013; Ostrin et al., 2017; Cain et al., 2020). While many newer displays are trying to combat this problem with a "night-shift" mode, which simply shifts the white balance of the display towards longer wavelengths, reducing the amount of short-wavelength light exposure of the evening, doing so alters the chromaticity and thus the visual appearance of the display. Current "night-shift" modes do not directly tackle the problem of controlling melanopsin signals, and instead try to suppress melanopsin activity by a general short-wavelength reduction, which in turn affects the relative cone signals. And if the goal is to minimise circadian disruption, it seems that such colour tuning does not have as strong an effect as dimming, meaning many current "night-shift" modes may not be effective at preventing melatonin suppression (Nagare et al., 2019).

In addition to minimizing inappropriate melanopsin activation from display viewing, one can also think of instances where the active control of the melanopic irradiance could be beneficial,

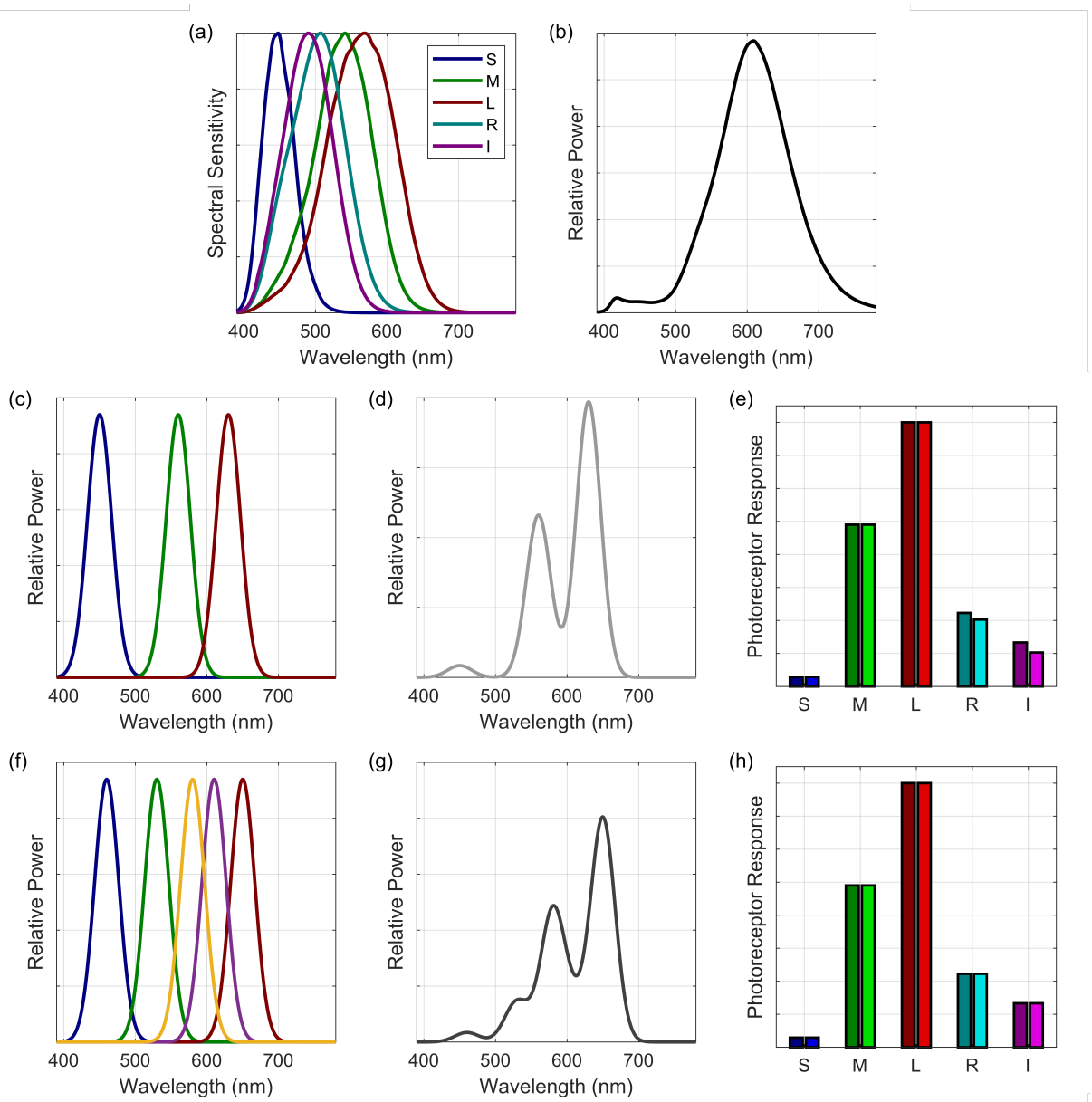


Figure 6.1: An example of how a display may reproduce and distort photoreceptor signals. (a) The spectral sensitivity functions, $T_k(\lambda)$, of the cones ($k \in [L, M, S]$), rods ($k = R$), and melanopsin ($k = I$) (as in the CIE S026 standard) (Stockman et al., 1999; Stockman & Sharpe, 2000; Wyszecki & Stiles, 1982; Lucas et al., 2014; CIE, 2018). (b) An example radiance spectrum from a real-world reflectance surface viewed under a real-world illuminant. (c) Example primaries, $p(\lambda)$, of a hypothetical three-primary display. (d) The spectral output, $P(\lambda)$, of the hypothetical three-primary display (c) that reproduces the same cone excitations as the real-world radiance spectrum (b). (e) The cone, rod, and melanopsin excitations from the real-world radiance spectrum, $\vec{A}$, ((b), the left-hand bar in each pair, in darker colours) and from the hypothetical three-primary display's spectral output, $\vec{A}^*$, ((d), the right-hand bar in each pair in lighter colours). (f) Example primaries, $p(\lambda)$, of a hypothetical five-primary display. (g) The spectral output, $P(\lambda)$, of the hypothetical five-primary display (f) that reproduces the same photoreceptor excitations as the real-world radiance spectrum (b). (h) The photoreceptor excitations from the real-world spectrum, $\vec{A}$, (b, left bar in each pair, darker colours) and the hypothetical five-primary display's spectral output, $\vec{A}^*$, (g, right bar in each pair, lighter colours). While there is a mismatch in rod and melanopsin signals for the three-primary displays, the five-primary displays can reproduce all photoreceptor signals.

such as tailored light exposure to mitigate circadian disruption in shift work, where individuals are particularly vulnerable to low mood and low alertness (Chellappa et al., 2020; Ganesan et al., 2019). Looking further to the future, topical announcements from technology giants about the so-called "metaverse" (augmented and virtual reality (AR and VR) and immersive technologies more generally) highlight the possibility that users may spend tens of hours of the week immersed in an artificial visual environment. Under these circumstances, characterising the full photoreceptor response to the available light becomes more pressing.

The role of ipRGCs in pupil size control could additionally have consequences for spatial or 3D reproduction. 3D displays have long been afflicted by the inability to reproduce the focus cues of a real image correctly, and therefore struggle to generate a 3D percept without associated visual discomfort (Ravikumar et al., 2011). Pupil size is physiologically linked to accommodation and vergence through the accommodation triad, and pupil size also affects depth of field. Whether or not the pupil size change driven by melanopsin would be of a sufficient magnitude to influence cue-conflict through these mechanisms remains unclear and an interesting area of future empirical testing.

6.1.3 Towards more accurate physiologically relevant reproduction of light on displays

Both hardware and software developments may help improve a display's ability to more accurately reproduce the five photoreceptor signals. Five-primary displays offer an exciting solution to control rod and melanopsin signals directly, without distorting the cone responses (Hexley, Yöntem et al., 2020; Allen et al., 2019; Cao et al., 2015), as described in Chapter 5. With five primaries one is able to directly control the signals of all five photoreceptors, as long as the signals are within gamut, as described in Eq. 6.3, and illustrated by an example in Fig. 6.1(f)-(h):

$$\vec{C} = B^{-1}\vec{A}$$

where

$$\vec{C} = \begin{bmatrix} r \\ g \\ b \\ c \\ m \end{bmatrix} \text{ and } \vec{A} = \begin{bmatrix} A_L \\ A_M \\ A_S \\ A_R \\ A_I \end{bmatrix} \tag{6.3}$$

and

$$B_{k,j} = \int_{\lambda} p_j(\lambda) \cdot \ T_k(\lambda)\, d\lambda$$

$$j \in [r, g, b, c, m]$$

$$k \in [L, M, S, R, I].$$

What might be done in software to provide a solution for most commercial displays that do not have the required number of primaries for an exact solution? For such three-primary displays, and indeed four-primary displays, it may be possible to design a more optimal chromaticity transformation which permits some distortion in the cone responses to introduce less distortion in the rod and melanopsin signals, though extensive research would be needed to demonstrate that any such algorithm produced a scene that had acceptable chromaticities and visual properties.

6.1.4 Quantifying a display's physiologically relevant reproduction of light

An immediate problem arises when one begins to seek solutions for controlling rod and melanopsin responses from image reproduction systems: what metrics should one use to quantify the distortion and reproduction of the rod and melanopsin signals? Colour gamuts based on the CIE 1931 XYZ functions, by definition, capture only chromaticity and photopic luminance, i.e. the cone signals, that a display can reproduce (CIE, 1932). The CIE 2015 (170-2) report introduces a move away from the CIE 1931 XYZ functions towards a physiologically relevant colour space, based on the CIE 2006 (170-1) cone fundamentals, though still only offers a quantification of the cone signals, not rods and melanopsin (CIE, 2006, 2015). There is therefore a need to look beyond traditional and physiologically relevant colour spaces that offer only a three-dimensional sub-space of the five-dimensional physiological space, in search of a better-suited, physiologically-relevant visualization and accompanying metrics to quantify display reproduction.

Three novel, physiologically relevant metrics are proposed in this chapter as a means of quantifying the reproduction and distortion of photoreceptor signals on displays. The first proposed metric, the photoreceptor signal reproduction metric (*PSRM*), quantifies the ability of a display to accurately reproduce a set of correlated quintuplets of photoreceptor signals. The second metric, the photoreceptor signal distortion metric (*PSDM*), quantifies any distortions introduced into any of the photoreceptor signals on the display compared to the desired signal. The final metric, the photoreceptor correlation distortion metric (*PCDM*), quantifies any distortions introduced to the correlations between the five photoreceptor signals from the desired signals to the displayed signals. In addition to the novel metrics, an equal-luminance photoreceptor excitation diagram, an extension of the MacLeod-Boynton chromaticity diagram (MacLeod & Boynton, 1979), is proposed to allow for a physiologically relevant visualization of the real-world and reproduced photoreceptor signals.

6.2 Methods: Introducing three novel metrics and an equal-luminance photoreceptor excitation diagram

One of the challenges for designing metrics to quantify reproduction of rod and melanopsin signals, is deciding what reference dataset to use. For chromaticity reproduction, the goal has often been to achieve a gamut that reaches the spectrum locus in order to display all physically possible visible colours, but in reality, displays are still a long way from achieving that. The sRGB, Adobe RGB, and Rec. 709 gamuts have become popular standard gamuts to act as a realistic reference against which to compare chromaticity reproduction in displays (International Electrotechnical Commission, 1999; Adobe Systems Incorporated, 2005; International Telecommunications Union, 2015). There is currently no equivalent gamut concept, or reference standard, that incorporates rod and melanopsin signals. Instead, one must define an appropriate reference dataset against which to compare rod and melanopsin signal reproduction. One option, which is presented in this chapter, is to consider the plausible spectra encountered in the real-world, or an approximation thereof, as the reference dataset. In the real-world, much of our visual input originates from objects that do not themselves emit light, but that are visible by virtue of the light they reflect. For matte surfaces this process is modelled by a surface being illuminated by a particular illuminant, and the multiplication of the surface reflectance and illuminant spectrum gives the resultant radiance spectrum that the observer sees. Illuminant spectra are available through specular reflection from glossy surfaces, or from direct viewing.

An ideal reference dataset might perhaps contain all possible real-world surfaces and illuminants,

weighted by the likelihood of an observer to interact with each surface under each illuminant. Such a dataset does not yet exist (Webler et al., 2019). However, there are datasets of real-world surfaces and real-world illuminants, from which a reference dataset of radiance spectra can be derived, that captures the variety of surfaces and illuminants that one may encounter in the real-world. Such a dataset is used here as a reference against which to quantify the reproduction and distortion of the reproduced photoreceptor quintuplet signals of a display.

The real-world reference dataset used here compromises 39,699 different radiance spectra, derived from 99 real-world reflectance samples under 401 different illuminants (Houser et al., 2013; David et al., 2015). The 99 reflectance samples used here are the samples used in the most recent derivation of the colour rendering index (David et al., 2015), and the 401 illuminants are the real illuminants under which those samples are rendered in this derivation (Houser et al., 2013; David et al., 2015). The set of 99 reflectance samples used here have been adopted by the Illuminating Engineering Society of North America (IES) in the IES TM-30-2015 standard, and have been proposed for adoption by the CIE. The term real-world spectra here refers to both natural and artificial illuminants and surfaces. The illuminant database contains the daylight series, black-body radiators, narrowband and broadband fluorescent lamps, LEDs, Tungsten Filament bulbs, and other miscellaneous illuminants. The reflectance database contains samples from nature, skin, textiles, paints, plastics, prints, and colour systems, and are chosen as the best representative set of samples that are uniformly distributed in colour space and wavelength space (David et al., 2015). Full details of the illuminant and reflectance datasets are detailed elsewhere (Houser et al., 2013; David et al., 2015), and the appropriateness of this example dataset is further evaluated in the Discussion of this chapter.

The radiant spectra, $\psi(\lambda)$, are calculated from all possible combinations of a reflectance spectrum, $\upsilon(\lambda)$, and an illuminant spectrum, $\iota(\lambda)$, as described in Eq. 6.4, for a single case:

$$\psi(\lambda) = \upsilon(\lambda) \cdot \iota(\lambda). \tag{6.4}$$

As the focus of this chapter is on the reproduction of the relative photoreceptor quintuplets, not the reproduction of the absolute signals, the illuminant spectra were normalized to have unit area under the curve, as in Eq. 6.5:

$$\int_{\lambda=400nm}^{\lambda=700nm} \iota(\lambda)\, d\lambda = 1. \tag{6.5}$$

6.2.1 Metric I: Photoreceptor Signal Reproduction Metric (PSRM)

The first proposed metric quantifies the proportion of the real-world photoreceptor signal quintuplets that can be accurately reproduced on the display which is useful for comparing the extent of the five-dimensional gamut across different displays. The metric is termed the photoreceptor signal reproduction metric, or *PSRM*. This can be thought of as a five-dimensional extension of the traditional colour gamut, i.e. the *PSRM* quantifies the proportion of all n real-world photoreceptor signal quintuplets, $\vec{A} = [A_S, A_M, A_L, A_R, A_I]$, that can be reproduced on the display, $\vec{A}^* = [A^*_S, A^*_M, A^*_L, A^*_R, A^*_I]$, defined as in Eq. 6.6:

$$PSRM = \frac{\sum_i^n \delta_{A_{k_i} A^*_{k_i}}}{n}$$

where

$$\delta_{A_{k_i} A^*_{k_i}} = \begin{cases} 0 & \text{if } A_{k_i} \neq A^*_{k_i} \text{ for any } k \in [S, M, L, R, I] \\ 1 & \text{if } A_{k_i} = A^*_{k_i} \text{ for all } k \in [S, M, L, R, I], \end{cases} \tag{6.6}$$

where the criteria for reproduction is that none of the primaries can be negative, and at least one of the primaries must be greater than zero (to avoid the case of counting a reproduction where it is necessary that the photoreceptor signals are scaled such that $\vec{A} = \vec{A}^* = 0$). Note, it is not explicitly required that the primaries be below one, as any cases where this is true may simply be scaled by a multiplicative factor so that reproduction can be achieved. Thus, the reproduced photoreceptor quintuplets are only considered plausible and thus counted in the above sum if all of the primaries of the display can be set to non-negative values and at least one can be set to non-zero values for the signal to be reproduced. All possible 39,699 real-world quintuplets are counted in the denominator, to quantify the proportion of real-world quintuplets that are accurately reproduced. The reproduced photoreceptor quintuplets, $[A^*_S, A^*_M, A^*_L, A^*_R, A^*_I]$, are only counted if they are equal to the real-world quintuplet that is being reproduced, $[A_S, A_M, A_L, A_R, A_I]$. The minimum possible *PSRM* in the case where the display does not accurately reproduce any of the real-world quintuplets is zero, while the maximum possible value in the case of perfect reproduction is 1 (or if expressed as a percentage, 100%).

In practice, a certain level of tolerance is necessary. Here, each photoreceptor signal for a particular spectrum, i, is considered to be reproduced on the display, i.e. $A^*_{k_i} = A_{k_i}$, if the reproduced signal, $A^*_{k_i}$, is within 1% of error of the real-world signal, A_{k_i} (the error metric is defined in Eq. 6.7), and if all of the primaries can be set to non-negative values, and one to a non-zero value to achieve such reproduction:

$$\left\| \frac{A^*_{k_i} - A_{k_i}}{A_{k_i}} \right\| < \epsilon$$
$$\text{where here } \epsilon = 0.01 \tag{6.7}$$
$$k \in [S, M, L, R, I].$$

In the *PSRM*, one does not explicitly consider the magnitude and direction of the distortion introduced by three-primary displays, though the effects of the distortion on the rod and melanopsin signals will simply reduce the binary count that specifies the *PSRM* of the display.

6.2.2 Metric II: Photoreceptor Signal Distortion Metric (PSDM)

The second metric, the photoreceptor signal distortion metric, or *PSDM*, explicitly quantifies the distortion introduced into the uncontrolled photoreceptor signals. This metric describes the degree of failure to reproduce each of the real-world photoreceptor signals, which is important to characterize a display's inability to reproduce the real-world signals and the impact the introduced distortions may have on the observer. For instance, a display that introduces a slight distortion in the melanopsin signals may be preferable to a display that introduces large distortions in the melanopsin signals when the *PSRM* of the two displays is equivalent. For a particular photoreceptor, k, and a particular spectrum, i, the $PSDM_{k_i}$ is defined as the contrast between the reproduced signal, A_{k_i}, and the real-world signal, $A^*_{k_i}$, as in Eq. 6.8:

$$PSDM_{k_i} = \frac{A^*_{k_i} - A_{k_i}}{A_{k_i}}$$
$$k \in [S, M, L, R, I]. \tag{6.8}$$

The $PSDM_{k_i}$ is calculated only for those spectra where the reproduced signal is reproducible within the gamut of the display, i.e. when all of the primaries are set to non-negative, and at least one is set to non-zero. This contrast can also be expressed as a percentage. One can see that the $PSDM_{k_i}$ defined in Eq. 6.8 is the error metric used for the *PSRM* in Eq. 6.7, which gives the tolerance for whether photoreceptor quintuplets are considered to be reproduced.

The $PSDM_{k_i}$ can be measured for all realisable spectra (i.e. all spectra that fall within gamut) individually, as in Eq. 6.8, but it is often useful to report the distortion pooled across photoreceptors and/or spectra as a summary metric. As such the average $PSDM_k$ per photoreceptor, k, is defined as the mean of the absolute values of the $PSDM_{k_i}$ for a particular individual photoreceptor, k, for all n realisable spectra, as in Eq. 6.9:

$$PSDM_k = \frac{\sum_i^n |PSDM_{k_i}|}{n}$$
$$k \in [S, M, L, R, I]. \tag{6.9}$$

Further, the overall *PSDM* for a display, pooled across photoreceptors and averaged over spectra, can be defined as the mean of the absolute value of the pooled $PSDM_k$ across all photoreceptors as in Eq. 6.10:

$$PSDM = \frac{\sum_k |PSDM_k|}{5}$$
$$k \in [S, M, L, R, I]. \tag{6.10}$$

The minimum possible (i.e. the most desirable) $PSDM_k$ for each photoreceptor, k, (and by extension the overall *PSDM*) is zero in the case that no photoreceptor signals are distorted by the display when reproduced. In theory there is no upper limit on the maximum *PSDM*, and the limit depends on the particular set of display primaries and real-world signals being reproduced. By design for current use-cases of colour reproduction on three-primary displays, and by definition in the analyses below, the colours within gamut of the display do not change the cone signals ($PSDM_L = PSDM_M = PSDM_S = 0$). It would however be possible to define a transformation matrix (B, in Eq. 6.3) that fixed the melanopsin signal instead of the M cone signal for example, thus meaning the melanopsin distortion would be zero by definition, and introducing distortions to the M cone signal, or devising a transformation that balances the distortion across all five photoreceptor classes.

6.2.3 Metric III: Photoreceptor Correlation Distortion Metric (PCDM)

It is not just the photoreceptor signals, but rather the correlations between the photoreceptor signals that provides information about the radiance spectrum that is falling on the retina. Spectral information from multiple photoreceptors depends on the correlations between them. For example, L and M cone signals are highly correlated and discrimination along the red-green colour opponent axis depends on the decorrelation between the L and M cone signals. In real-world spectra there are distinct correlations between the different photoreceptor excitations, as shown in Fig. 6.2. Fig. 6.2(a) shows the excitations for each photoreceptor pair from the 39,699 real-world spectra in the reference dataset, while Fig. 6.2(b) shows the Pearson correlation coefficient describing the correlations between each pair of photoreceptor signals. In particular the L and M cone excitations are highly correlated, as are the rod and melanopsin excitations.

If one or more of the photoreceptor signals are distorted when reproduced on a display, then the

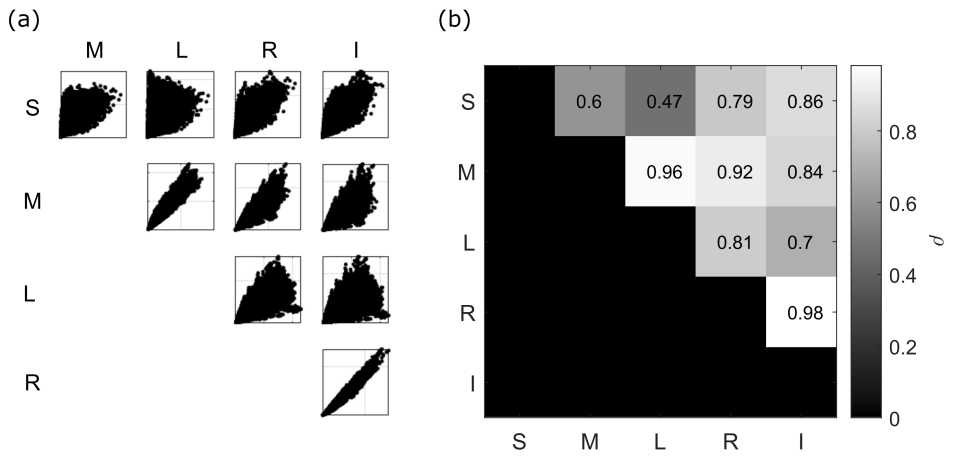


Figure 6.2: The correlations between the photoreceptor excitations from the 39,699 real-world spectra from our example reference dataset. (a) The photoreceptor excitations from the real-world spectra plotted pairwise. (b) The Pearson correlation coefficient matrix plotted as a heat map showing the correlation coefficient of each pairwise combination of photoreceptor signals.

correlations between that photoreceptor signal and the remaining photoreceptors may be distorted. Hence, the third and final metric, the photoreceptor correlation distortion metric, or *PCDM*, is defined. It is worth highlighting the distinction between the *PSDM* and the *PCDM* in terms of the information they provide about the nature of the distortions. While the $PSDM_k$ captures the degree of distortion in a particular photoreceptor signal (or pooled over photoreceptors in the overall *PSDM*), it does not directly describe the directionality of this distortion in relation to other photoreceptor signals, as the *PCDM* does. For example, it is possible for the *PSDM* to be non-zero while the *PCDM* is zero, if, for instance, all of the reproduced melanopsin signals are consistently less than the real-world melanopsin signals by a fixed factor. The *PCDM* considers pairwise correlations between photoreceptor signals for the set of real-world spectra, ρ_{k_1,k_2}, and the pairwise correlations for the reproduced spectra, $\rho^*_{k_1,k_2}$, (Eq. 6.11):

$$\begin{aligned} PCDM_{k_1,k_2} &= \rho^*_{k_1,k_2} - \rho_{k_1,k_2} \\ k_1, k_2 &\in [S, M, L, R, I], \end{aligned} \tag{6.11}$$

for all pairs of spectra where the reproduced spectra falls within gamut of the display i.e. when all of the primaries take non-negative values, and at least one takes a non-zero value. As with the *PSDM*, the minimum possible *PCDM* is zero in the case where all photoreceptor quintuplets can be reproduced without distortion (in the case where the *PSRM* is 1 (or 100%)). The theoretical maximum value for the *PCDM* would be 2 (or -2) corresponding to a complete reversal of the correlation (from a perfect positive correlation to a perfect negative correlation or vice versa) though in practice the *PCDM*

will never approach this maximum limit.

6.2.4 Equal-luminance photoreceptor excitation diagram

While the CIE 1931 xy chromaticity diagram provides a familiar and quick method to visualize chromaticities it provides no information of the distribution of rod and melanopsin signals. It is further worth noting that the CIE 1931 xy chromaticity space is no longer considered to be physiologically relevant, as it uses the CIE 1931 XYZ colour matching functions as its basis, which are not transformable to the physiologically-relevant 10 ° CIE 2006 standard cone fundamentals (as in the CIES026 standard) (Stockman et al., 1999; Stockman & Sharpe, 2000; CIE, 2006, 2018, 1932, 2015).

For a physiologically-relevant chromaticity space, one can turn to the MacLeod-Boynton chromaticity diagram (MacLeod & Boynton, 1979). While less common in the display industry, the MacLeod-Boynton chromaticity diagram is a widely-used colour space among vision scientists. The MacLeod-Boynton space provides a visualization of the cone excitations at equal luminance. The traditional MacLeod-Boynton chromaticity space plots $s_{\mathrm{MB}}(\lambda)$ against $l_{\mathrm{MB}}(\lambda)$, where $s_{\mathrm{MB}}(\lambda)$ and $l_{\mathrm{MB}}(\lambda)$ are defined as in Eq. 6.12:

$$
\begin{gathered}
s_{\mathrm{MB}}(\lambda) = \frac{a_S T_S(\lambda)}{V(\lambda)},\ l_{\mathrm{MB}}(\lambda) = \frac{a_L T_L(\lambda)}{V(\lambda)} \\
\text{where} \\
V(\lambda) = a_L T_L(\lambda) + a_M T_M(\lambda) \\
\text{and} \\
V(\lambda) \text{ and } s_{\mathrm{MB}}(\lambda) \text{ peak at unity.}
\end{gathered}
\tag{6.12}
$$

Throughout this thesis, the CIE 2006 10° spectral sensitivity functions (as in the CIE S026 standard) are used and normalised to unity peak (Stockman et al., 1999; Stockman & Sharpe, 2000; CIE, 2006, 2018). In terms of energy, the scaling constants, a_S, a_M, and a_L, take the following values (Stockman & Sharpe, 2000; MacLeod & Boynton, 1979):

$$a_L = 0.692839$$

$$a_M = 0.349676$$

$$a_S = 0.0554786.$$

The spectrum locus in the traditional MacLeod-Boynton diagram is shown in Fig. 6.3(a). The

MacLeod-Boynton space lends itself to a natural extension to include a melanopsin axis. Thus a new equal-luminance photoreceptor excitation diagram is proposed, which plots the melanopic efficacy of radiance i.e. the melanopsin excitation normalized by luminance (CIE, 2006, 2015). The newly proposed axis plots the variable $i_{\mathrm{MB}}(\lambda)$, which is defined in Eq. 6.13:

$$i_{\mathrm{MB}}(\lambda) = \frac{a_I T_I(\lambda)}{V(\lambda)}$$
$$\text{scaled so that} \tag{6.13}$$
$$i_{\mathrm{MB}}(\lambda) \text{ peaks at unity.}$$

The CIE S206 spectral sensitivity function for melanopsin is used here, and normalised to unity peak (Lucas et al., 2014; CIE, 2018). In terms of energy, the scaling constant, a_I, takes the following value in order for $i_{\mathrm{MB}}(\lambda)$ to peak at unity:

$$a_I = 0.131746.$$

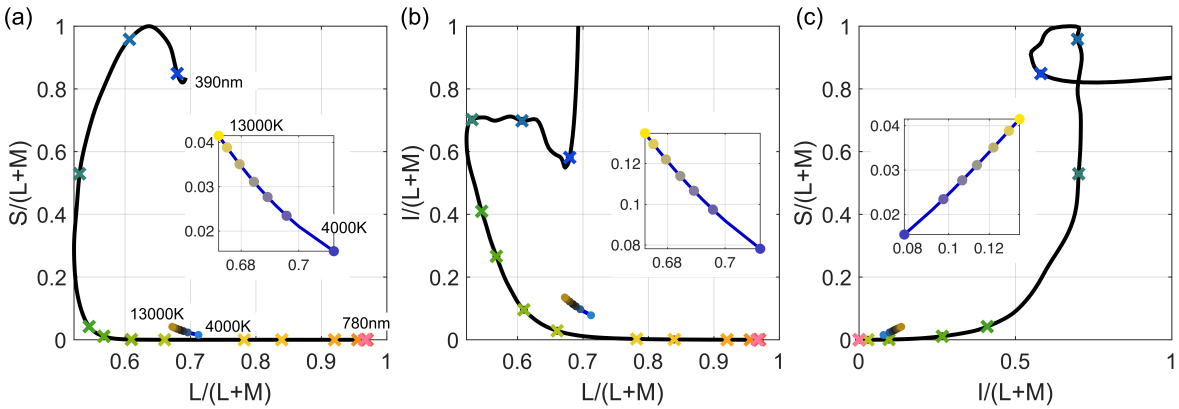


Figure 6.3: The proposed equal-luminance photoreceptor excitation space to capture melanopsin signals. (a) The spectrum locus plotted in the traditional MacLeod-Boynton chromaticity diagram, from 390 nm to 780 nm. Crosses indicate 25 nm steps and are coloured using a perceptually uniform rainbow colour-map to indicate the progression from short to long wavelengths along the spectrum locus (Kovesi, 2015). The daylight locus, from 4,000K to 13,000K is also shown, with a blue to yellow colour-map depicting the progression from cooler to warmer temperatures along the daylight locus. (b) The spectrum locus and daylight locus plotted in one projection of the proposed equal-luminance photoreceptor excitation diagram, plotting the melanopic efficacy of radiant flux (I/(L+M)) against the L/(L+M) excitation. (c) The other projection of the proposed equal-luminance photoreceptor excitation diagram, plotting the spectrum locus as S/(L+M) against I/(L+M). In each plot the daylight locus is shown magnified in the inset. Note, the concavities on the above plot are due to measurement uncertainty in the spectral sensitivity functions at short wavelengths.

The spectrum locus is depicted in the proposed equal-luminance photoreceptor excitation diagram in Fig. 6.3(b)-(c). Note, the concavities in this figure are due to measurement uncertainties in the spectral sensitivity functions at short wavelengths. The decision was made to use the full wavelength range over which the spectral sensitivities are measured i.e. 390 nm to 780 nm, to be consistent with

the way the traditional MacLeod-Boynton diagram normalisation is achieved (Stockman & Sharpe, 2000; MacLeod & Boynton, 1979). It is worth noting that the newly proposed equal-luminance photoreceptor excitation does not provide any visualization of the rod excitations. Previous work has derived visualizations of rod excitation on traditional chromaticity diagrams (Shapiro et al., 1996), so in this chapter I opted for an approach that allowed for visualization of the melanopsin signal. As the rod and melanopsin signals are very highly correlated (Fig. 6.2), and there is no known dominant post-receptoral pathway that compares rod and melanopsin signals, choosing to depict one of the rod or melanopsin signals allows for a three-dimensional visualization of the five-dimensional space that captures most of the relevant information necessary to understand the features of the space.

6.3 Results: Example applications of novel metrics

To demonstrate the effectiveness of the proposed metrics they are applied here to three different real three-primary displays, two research grade five primary-displays, and to two hypothetical five-primary displays, including the author's own system as described in Chapter 5.

Three real three-primary displays are evaluated: a CRT NEC Color Monitor MultiSync FP2141SB (NEC Display Solutions, Tokyo, Japan), a Dell U2715H LCD Monitor (Dell, Austin, TX, USA), and a CRS Display++ LCD Monitor (Cambridge Research Systems, Rochester, UK). The displays were measured with a CRS SpectroCal MKII Spectroradiometer (Cambridge Research Systems, Rochester, UK). The measured primaries of the displays are shown in Fig. 6.4(a)-(c). The two custom-built research displays that use five or more primaries, analysed here are a five-primary visual display unit (VDU) built at the University of Manchester (Spitschan et al., 2019; Allen et al., 2019), and the multi-primary high dynamic range (MPHDR) display built by the author at the University of Oxford (Hexley, Yöntem et al., 2020). The primaries for the Manchester five-primary VDU are shown in Fig. 6.4(d) and the primaries for the Oxford MPHDR are shown in Fig. 6.4(e). (Note, for the MPHDR which has six primaries, a solution for the *PSRM* (see below) is derived by permuting through all possible five primary combinations of the six primaries, in turn.) Two hypothetical five-primary displays are also evaluated: a hypothetical narrowband display, and a hypothetical broadband display (primaries depicted in Fig. 6.4(f)-(g)). The primaries in the hypothetical five-primary displays are either narrowband or broadband and their placement in the visible spectrum is arbitrary. The consequences of the primary placement are considered in the Discussion.

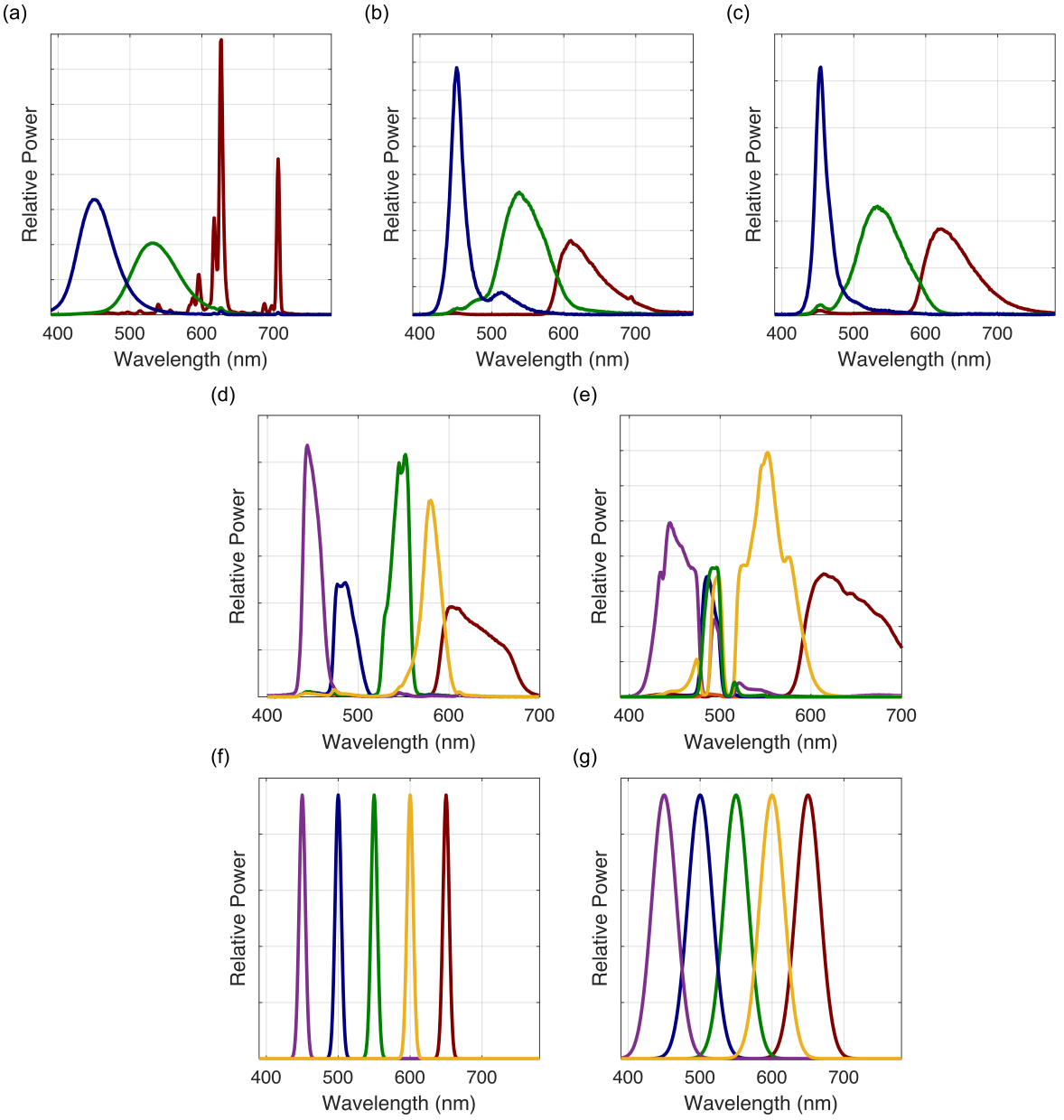


Figure 6.4: The display primaries used in example applications of the proposed metrics. (a) The primaries of the CRT NEC Color Monitor. (b) The primaries of the Dell U2715H LCD Monitor. (c) The primaries of the CRS Display++ LCD Monitor. (d) The primaries of the Manchester five-primary VDU. (e) The primaries of the Oxford MPHDR. (f) The primaries of the hypothetical narrowband five-primary display (all primaries with a full-width half-maximum of 10 nm and peaks at 450 nm, 500 nm, 550 nm, 600 nm, and 650 nm). (g) The primaries of the hypothetical broadband five-primary display (all primaries with a full-width half-maximum of 40 nm and peaks at 450 nm, 500 nm, 550 nm, 600 nm, and 650 nm).

6.3.1 Photoreceptor Signal Reproduction Metric (PSRM)

The majority of chromaticities of the real-world signals fall within the gamut of the chosen displays, as depicted in the CIE 1931 xy chromaticity diagram in Fig. 6.5(a). In fact, as shown in Fig. 6.5(b), at least 90% of the chromaticities can be reproduced by each display (CRT NEC Color Monitor: 90.8%; Dell U2715H LCD Monitor: 91.0%; CRS Display++ LCD Monitor: 94.1%; Manchester VDU: 97.7%;

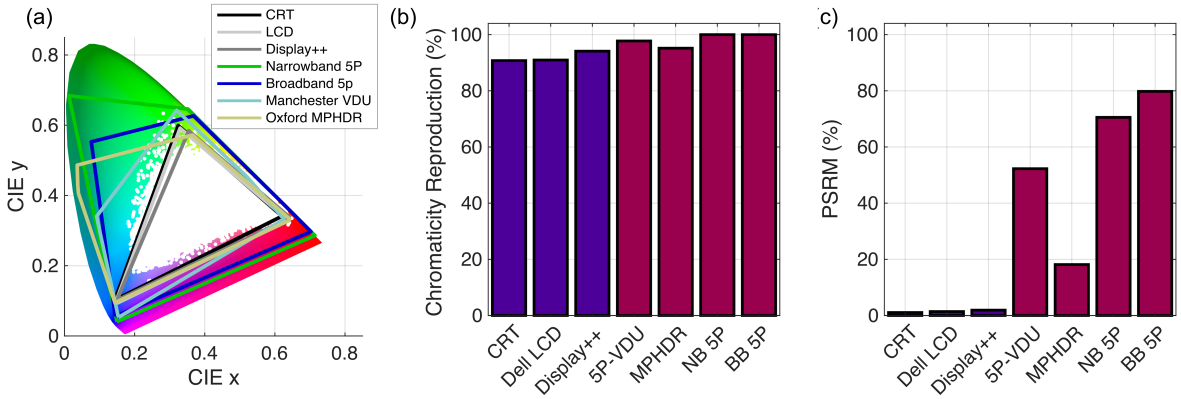


Figure 6.5: The chromaticity reproduction and *PSRM* of the real-world reference dataset spectra on the example displays. (a) Chromaticities of the 39,699 spectra plotted on the CIE 1931 xy chromaticity diagram, along with the colour gamuts of the example displays. (b) The chromaticity reproduction rate of the example displays. (c) The *PSRM* of the example displays. Despite the high chromaticity reproduction of all displays, the *PSRM* of the three-primary displays is very low, while it remains quite high for the five-primary displays. (In (b) and (c) the three-primary displays are shown in purple, and the five-primary displays in pink).

Oxford MPHDR: 95.2%; Narrowband Five-Primary Display: 100%; Broadband Five-Primary Display: 100%). Despite the excellent chromaticity reproduction, Fig. 6.5(c) shows the extremely low *PSRM* of the three-primary displays (CRT NEC Color Monitor: 1.1%; Dell U2715H LCD Monitor: 1.4%; CRS Display++ LCD Monitor: 1.9%). This is not surprising as any match between the reproduced rod and melanopsin signals to the real-world signals is by coincidence.

For both the hypothetical five-primary displays, the *PSRM* is quite high (Narrowband Five-Primary Display: 70.5%; Broadband Five-Primary Display: 79.8%). The *PSRM* does vary between the hypothetical five-primary displays which shows that whilst having five primaries is necessary to reproduce the five photoreceptor signals, it is not sufficient to use any arbitrary five primaries and expect to reproduce all desired real-world signals; similar to how one cannot simply pick any arbitrary three primaries and expect to reproduce a large colour gamut in three-primary displays. The placement of primaries along the wavelength axis is critical. The *PSRM* is lower for the research grade systems than for the hypothetical systems (*PSRM* of 52.3% for the Manchester five-primary VDU and a *PSRM* of 18.1% for the Oxford MPHDR). The reason for this comparatively low *PSRM* is that neither display was explicitly optimized for the purpose of reproduction of real-world signals. For example, the Oxford MPHDR display was designed to optimize the available melanopsin *contrast* and chromaticity gamut, as previously described in Chapter 5 (Hexley, Yöntem et al., 2020). This emphasises the point that the virtue of having five or more primaries alone is not enough to guarantee a high *PSRM*, and the choice of primaries for such systems must be optimized to achieve high reproduction. I therefore foresee the *PSRM* as a useful metric to guide the selection of primaries for future multi-primary display

technologies where the goal is real-world photoreceptor signal reproduction.

6.3.2 Photoreceptor Signal Distortion Metric (PSDM)

By definition, for all spectra that fall within gamut, the $PSDM_k$ for all photoreceptors ($k \in [L, M, S, R, I]$) for the two custom-built and two hypothetical five-primary displays is 0 (the ideal value of the $PSDM_k$ for perfect reproduction of all five photoreceptors signals). Fig. 6.6(a) demonstrates the rod and melanopsin distortions introduced in the Dell LCD (taken as an example representation of the three-primary displays) when the cone signals are matched for all realisable spectra, by plotting the reproduced signal, ${A_{ki}}^*$, against the real-world signal, A_{ki}, for each photoreceptor for each individual spectrum, i, for $k \in [R, I]$. If the reproduced signals were to match the real-world signals then all the data points would fall along the unity line, as they do for the cone responses in the three-primary displays and for all of the photoreceptor signals in the five-primary displays (not shown). The scatter shown here indicates the extent of the distortion. Fig. 6.6(b) shows the distribution of the rod and melanopsin $PSDM_{ki}$ for the reference dataset, $i = 1, 2, ...n$, for $k \in [R, I]$, for the Dell U2715H LCD Monitor. The equivalent plots for the CRT NEC Monitor and CRS Display++ LCD Monitor are shown in Fig. 6.7. For all three three-primary displays the $PSDM_{k_i}$ for both rods and melanopsin is widely spread (CRT NEC Color Monitor: $\sigma_{PSDM_{Ri}} = 16.0\%$, $\sigma_{PSDM_{Ii}} = 24.8\%$; Dell U2715H LCD Monitor: $\sigma_{PSDM_{Ri}} = 15.3\%$, $\sigma_{PSDM_{Ii}} = 23.6\%$; CRS Display++ LCD Monitor: $\sigma_{PSDM_{Ri}} = 13.6\%$, $\sigma_{PSDM_{Ii}} = 20.2\%$). For the CRT NEC Color Monitor and Dell U2715H the mean $PSDM_{k_i}$ for both rods and melanopsin is large (CRT NEC Color Monitor: $\mu_{PSDM_{Ri}} = 14.1\%$, $\mu_{PSDM_{Ii}} = 20.2\%$; Dell U2715H LCD Monitor: $\mu_{PSDM_{Ri}} = 10.4\%$, $\mu_{PSDM_{Ii}} = 15.2\%$). For the CRS Display++ LCD Monitor, the mean $PSDM_{ki}$ for both rods and melanopsin is near zero (CRS Display++ LCD Monitor: $\mu_{PSDM_{Ri}} = 0.9\%$, $\mu_{PSDM_{Ii}} = 1.1\%$) though it is worth noting that the spectra that can be reproduced with near zero-rod distortion introduce melanopsin distortion, and further that the mean of the absolute value of the $PSDM_{k_i}$ is non-zero (hence why the absolute value is taken in the definition of the pooled $PSDM_k$ in Eq. 6.9).

The $PSDM_k$ for each photoreceptor for each three-primary display, shows that the mean of the absolute value of the distortion introduced for rods and melanopsin across all spectra in the reference dataset is large (CRT NEC Color Monitor: $PSDM_R = 16.3\%$, $PSDM_I = 23.5\%$; Dell U2715H LCD Monitor: $PSDM_R = 13.8\%$, $PSDM_I = 20.1\%$; CRS Display++ LCD Monitor: $PSDM_R = 10.1\%$, $PSDM_I = 14.7\%$). The overall *PSDM* pooled across photoreceptors is also large (as expected) for the three-primary displays: for the CRT NEC Color Monitor the overall *PSDM* is 7.97%, for the Dell U2715H LCD Monitor the overall *PSDM* is 6.79%, and for the CRS Display++ LCD Monitor the overall *PSDM* is 4.96%.

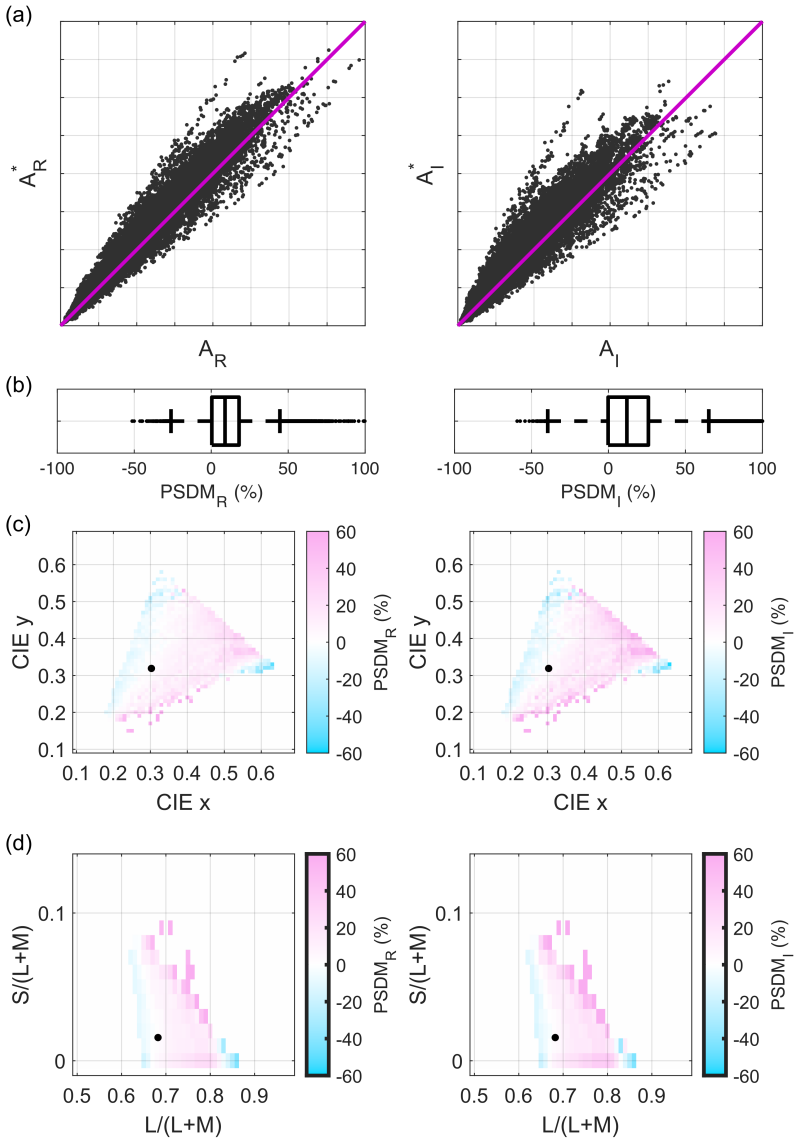


Figure 6.6: The *PSDM* applied to the Dell U2715H LCD Monitor as an illustrative example. (a) The rod ($k = R$) and melanopsin ($k = I$) photoreceptor signals reproduced by the display, A^*_k, are plotted against the rod and melanopsin signals from the corresponding real-world spectrum, A_k, for each spectrum in the reference dataset that falls in gamut. In the case the signals are undistorted, then they should fall along the line of unity (as is the case for the cone signals, which are not shown). (b) Box-plots showing the distribution of the rod and melanopsin $PSDM_{ki}$ (as defined in Eq. 6.8) over all individual spectra, i, that are within the chromaticity gamut for the display. (c) The mean rod and melanopsin distortion across the CIE 1931 xy chromaticity diagram. The distortions, $PSDM_{k_i}$, from all of the individual spectra whose chromaticities fall within a square on the evenly spaced 100×100 grid (with steps of 0.01 in both x and y) of the chromaticity diagram are taken together and the mean displayed in this plot as the $PSDM_{kxy}$. The black dot indicates the chromaticity of D65. (d) The mean rod and melanopsin distortion across the MacLeod-Boynton chromaticity diagram. Again, the set of spectra that fall within a particular bin (step size of 0.01 in both the L/(L+M) and S/(L+M) directions) in the 100×100 grid across the diagram, and the mean rod and melanopsin $PSDM_{ks,l}$ over the set is found. The black dot on the heat maps shows the chromaticity of D65.

The melanopsin and rod $PSDM_{k_i}$ (where $k \in [R, I]$) show a chromaticity dependence. This is not surprising as the photoreceptor signals are highly correlated with each other. To illustrate the

chromaticity dependence of the $PSDM_{k_i}$ the CIE 1931 xy chromaticity diagram is divided into a 100×100 grid (with each bin covering 0.01 in x and 0.01 in y), and the mean of all of the melanopsin and rod $PSDM_{k_i}$ values for the set of real-world spectra whose chromaticities fall within one grid square on the divided diagram is taken as the average $PSDM_{k_{xy}}$ and used to colour that grid square. This is illustrated for the Dell U2715H LCD Monitor in Fig. 6.6(c), and shown for the CRT NEC Monitor and CRS Display++ in Fig. 6.7. A similar principle is applied to the MacLeod-Boynton chromaticity diagram, whereby the chromaticity space is divided into a 100×100 grid and the mean melanopsin and rod $PSDM_{k_{s,l}}$ is calculated per square. An example is illustrated for the Dell U2715H LCD Monitor in Fig. 6.6(d).

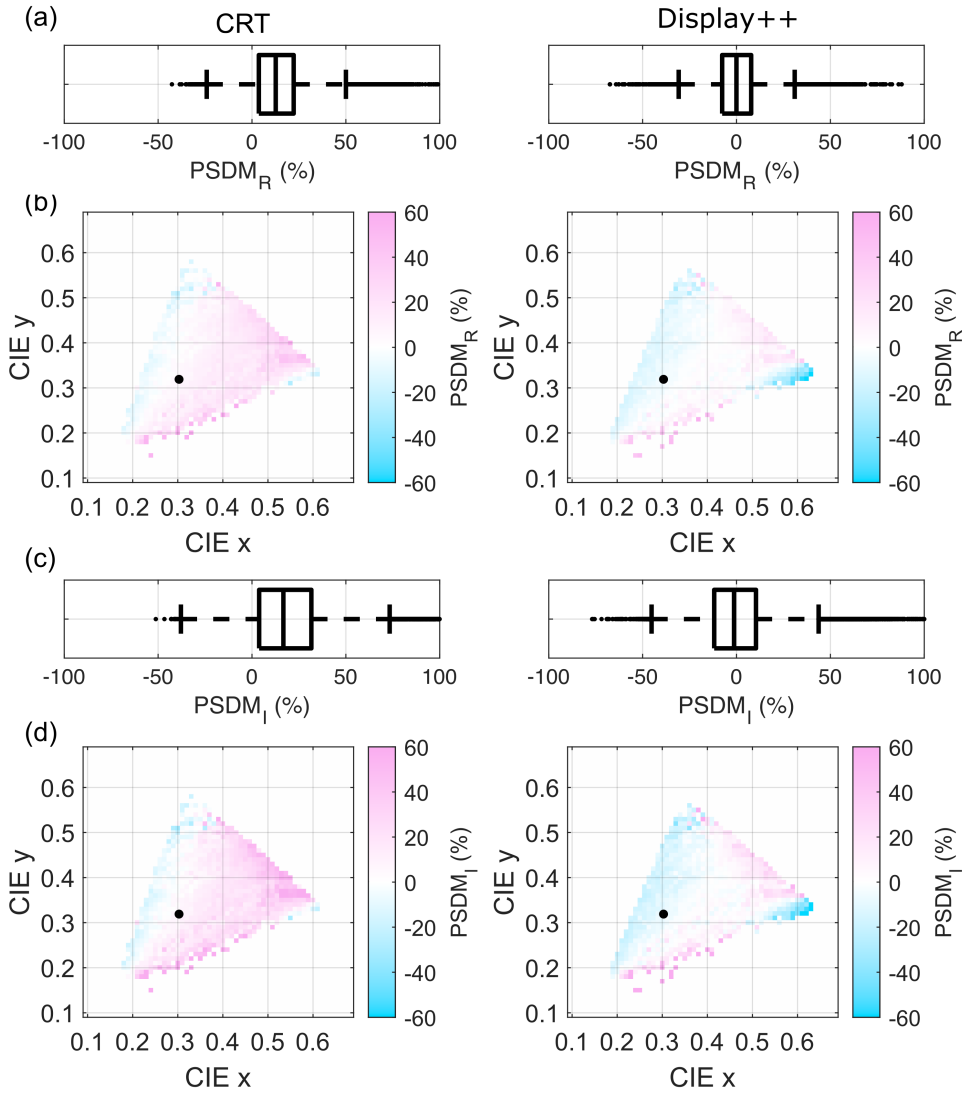


Figure 6.7: The *PSDM* of the CRT NEC Color Monitor and the CRS Display++ Monitor. (a) Box-plot showing the distribution of the rod $PSDM_{R_i}$ across all individual spectra, i that are realisable on the display. (b) The rod $PSDM_{R_{xy}}$ across the CIE 1931 xy chromaticity diagram. (c) Box-plot showing the distribution of the melanopsin $PSDM_{I_i}$. (d) The melanopsin $PSDM_{I_{xy}}$ across the CIE 1931 xy chromaticity diagram. The black dot on the heat maps shows the chromaticity of D65.

6.3.3 Photoreceptor Correlation Distortion Metric (PCDM)

By definition, the five-primary displays have a *PCDM* of zero for all photoreceptor pairs calculated for the spectra that are realisable on the display. Fig. 6.8 shows the *PCDM* for all photoreceptor pairs introduced by the three three-primary displays across the spectra within chromaticity gamut for each display. It is worth noting the large correlation distortions, particularly between the three cone excitations and the rod and melanopsin excitations, introduced by all the three-primary displays. The fact that the *PCDM* is non-zero indicates that it is not simply the case that all of the reproduced melanopsin excitations are less than the real-world melanopsin excitations by a constant factor, or in other words that the distorted photoreceptor signals are being mapped onto one of the undistorted photoreceptor signals. The *PCDM* for different photoreceptor pairs has different signs, showing that the correlations between different pairs of photoreceptors are sometimes increased and sometimes decreased. This means that not only will some of the correlations between real-world signals be lost during reproduction, but spurious correlations will be introduced by reproduction that do not exist between the photoreceptor excitations in the real-world.

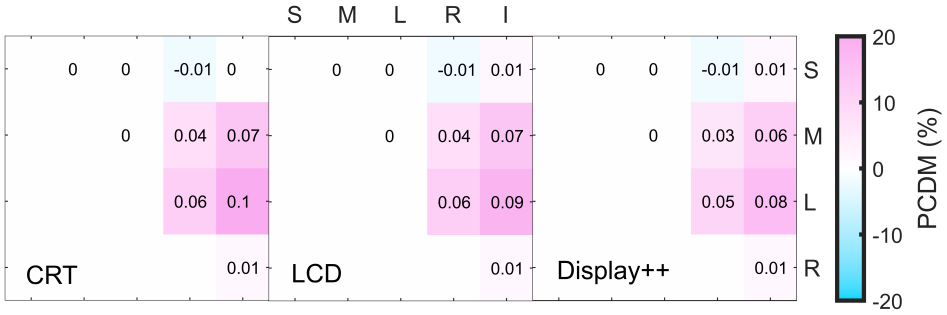


Figure 6.8: The *PCDM* of the three three-primary displays: the CRT NEC Color Monitor, the Dell U2715H LCD Monitor, and the CRS Display++ Monitor. By definition, the *PCDM* between the cone signals is 0, but large distortions are introduced in the correlation between the rod and melanopsin signals with each other photoreceptor.

6.3.4 Equal-luminance photoreceptor excitation visualizations

The 39,699 real-world spectra are plotted in the proposed equal-luminance photoreceptor excitation diagram in Fig. 6.9(a). As in the depiction of the *PSDM*, the space is divided into a 100 × 100 grid, and the frequency of the real-world spectra falling within a square on the grid is shown as the intensity of the heat map. Any spectra that are considered reproduced on a particular display (as in the *PSRM*), are depicted on the proposed equal-luminance photoreceptor excitation diagram space, as in Fig. 6.9(b) for the three-primary displays and in Fig. 6.10(b) for the five-primary displays. (Note the real-world

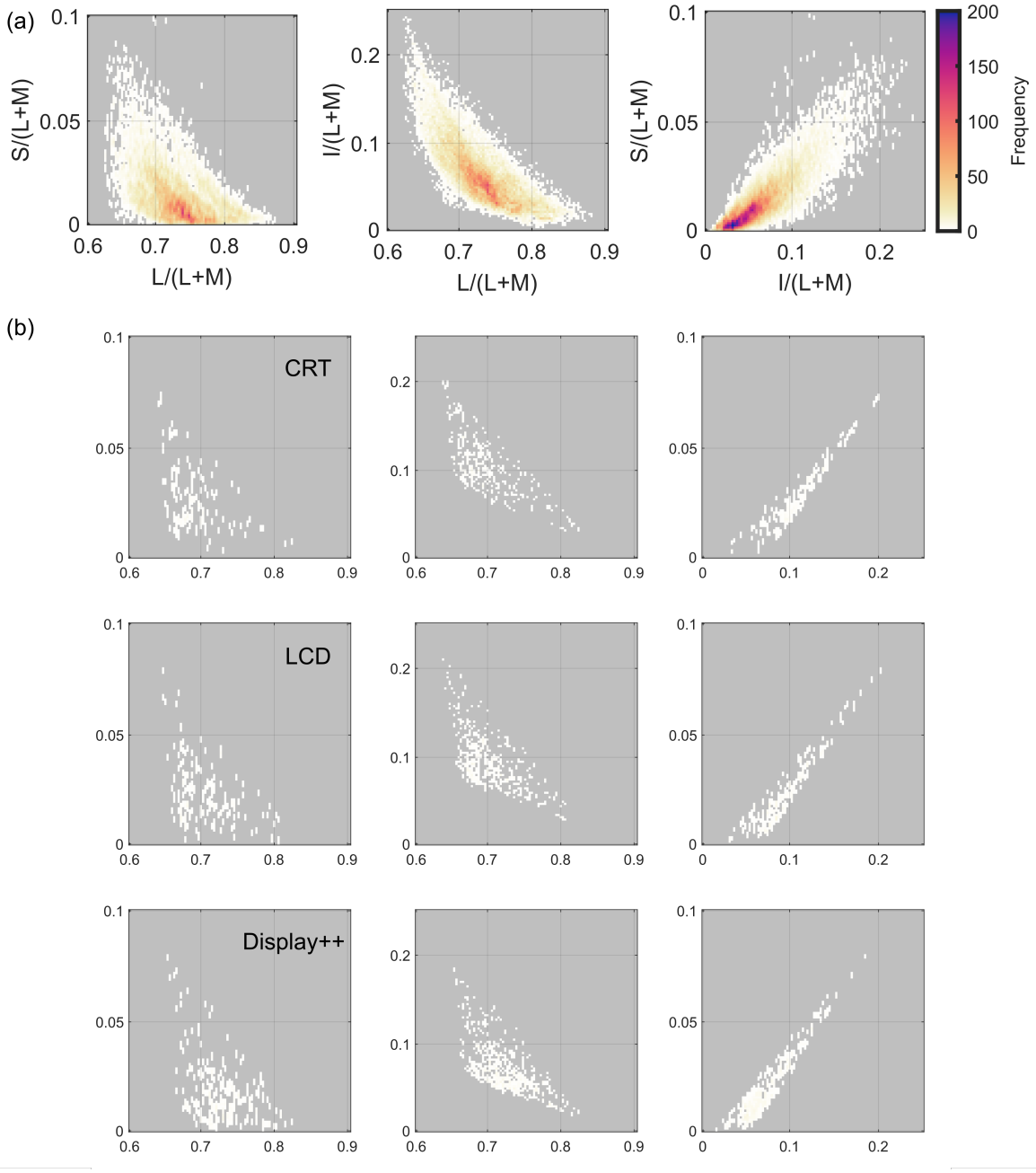


Figure 6.9: The photoreceptor signals from the real-world spectra and the reproduced display spectra depicted in the proposed visualization method of the extended MacLeod-Boynton space, along the traditional axis, and the two new projections including the extended melanopic axis. (a) The real-world spectra plotted as a heat map, depicting the number of real-world spectra that fall within a particular square when the axes shown are divided into a 100×100 grid. (b) The spectra that can be reproduced on the various three-primary monitors, depicted the same way (with the same axes and colorbar) as the real-world spectra in (a). Only the spectra that can be reproduced with a distortion within the allowed tolerance of the real-world photoreceptor signals for all photoreceptors i.e. $\frac{A_k^* - A_k}{A_k} < \epsilon$ for all $k \in [L, M, S, R, I]$, where here $\epsilon = 0.01$, are included in the plots in (b).

photoreceptor excitations are shown again in Fig. 6.10(a) for ease of comparison between the diagrams for the five-primary results figure).

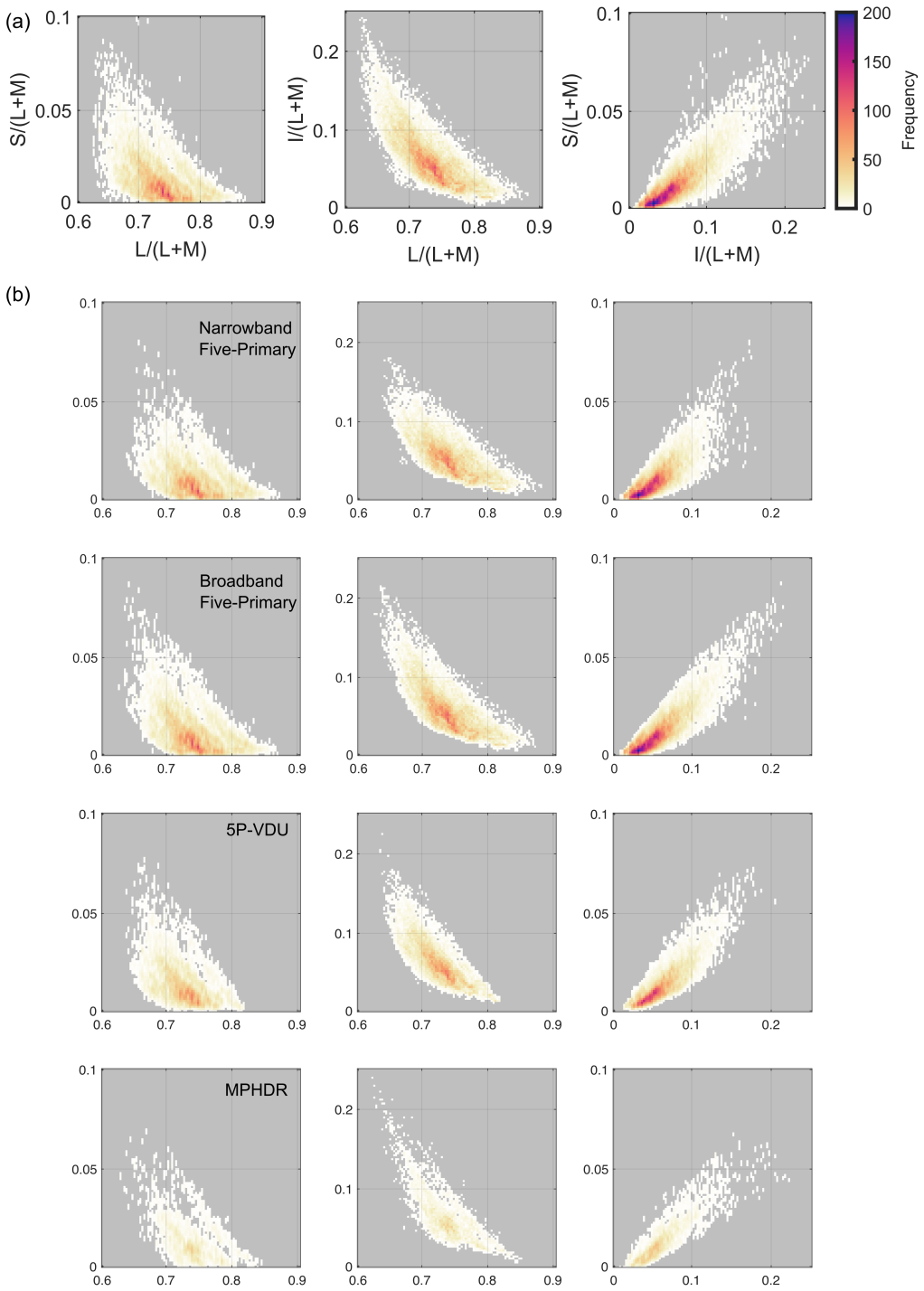


Figure 6.10: As in Fig. 6.9, the photoreceptor excitations elicited from the real-world spectra (a) and reproduced display spectra for the five-primary displays (b) shown in the proposed extended MacLeod-Boynton space.

In the case of the five-primary displays, the distribution of reproduced spectra looks very similar to the distribution of real-world spectra in the proposed equal-luminance photoreceptor excitation diagram space, though covering a slightly reduced area due to the non-perfect reproduction (i.e.

the *PSRM* is not 100%). For the three-primary displays however, the reproduced spectra on the proposed equal-luminance photoreceptor excitation diagrams cover a patchy and reduced area of the chromaticity diagram, indicating the inability of three-primary displays to reproduce all five real-world photoreceptor signals.

6.4 Discussion

Until now, metrics to characterize display reproduction have been focused solely on describing the reproduction of the visual signals of the image being reproduced. Colour gamuts have long been recognised as an important quantification of display reproduction, but they only describe chromaticity reproduction i.e. the reproduction of the cone photoreceptor signals. With mounting evidence suggesting that melanopsin strongly contributes to driving the non-visual (and perhaps even contributes to the visual) responses to light, and that rod saturation may not be as straightforward as once assumed, the display industry is beginning to consider the reproduction of rod and melanopsin responses on displays. To provide a metric-based framework for optimisation, there need to be metrics to describe a display's ability to reproduce, or in the case of displays that do not directly control the signals, the extent to which they distort, all five photoreceptor signals. Three such metrics were proposed in this chapter.

6.4.1 The applicability of the proposed metrics

The three novel metrics proposed in this chapter provide information about an image reproduction system's ability to capture all classes of photoreceptor signals available in the retina. The photoreceptor signal reproduction metric (*PSRM*) describes how many of the five photoreceptor signal quintuplets from a reference dataset can be reproduced on a display in a non-distorted manner, i.e. in the same ratios as in the real-world spectra. The *PSRM* is a single number that can be reported for any display type to describe the display's reproduction capability. The photoreceptor signal distortion metric (*PSDM*) describes what distortions are introduced when the relative photoreceptor quintuplet is distorted during the reproduction on a display using fewer than five primaries. There are various ways one could report the *PSDM*, as it can be quantified overall as both an average across photoreceptors and spectra, or per photoreceptor, k, and spectra, i ($PSDM_{k_i}$). I imagine the most useful application of the *PSDM* would be to report the overall *PSDM* as a single numerical metric for a given display, and to depict the $PSDM_{k_i}$ diagrammatically as shown in Fig. 6.6 and Fig. 6.7. Finally, the photoreceptor

correlation distortion metric (*PCDM*) quantifies how the correlations between the photoreceptor signals from the reference database are distorted by the display and should be depicted for the pairwise comparisons of the photoreceptors as in Fig. 6.8. Additionally, an equal-luminance photoreceptor excitation diagram is proposed as a visualization of signals that drive the human physiological response to light, which should provide a useful way to illustrate where distortions are introduced and where display reproduction fails.

As interest in "night-shift" modes, 3D displays, and real-world reproduction continues to grow in the display industry, the consideration of melanopsin signal reproduction by display manufacturers becomes increasingly important. It is hoped that the metrics proposed here highlight the fundamental limitation of display systems with fewer than five-primaries to control the non-visual physiological response to light, due to the lack of control and resulting distortions of the melanopsin signal. It is further hoped that the metrics and visualization proposed here offer a useful method to characterize an image reproduction system's ability to accurately reproduce the full set of photoreceptor signals from a real-world scene.

It is worth emphasising that the applicability of these metrics, and the "optimal" value for each metric depends on the particular use case of the display in question. For example, three-primary displays are not designed to reproduce the rod and melanopsin signals from real-world spectra, (and indeed neither were the two custom-built five primary displays) and thus it is an acceptable design choice that the *PSRM* is low. These metrics are not intended as a critique of such displays, but rather as a tool to inform the design of displays that move beyond colour reproduction and towards full photoreceptor signal reproduction as their dominant feature. The characterization of displays that are not designed for such purpose by these metrics is useful as a "benchmark" and to better understand the nature of the distortions introduced in the photoreceptor signals when left uncontrolled. The metrics proposed here are not being suggested as a replacement to current metrics, such as colour gamuts, but rather as an addition to current metrics.

6.4.2 Limitations of proposed metrics

There are limits to what the three metrics proposed here can quantify. One notable limitation is that the metrics are based on low-level physiological information, rather than on higher-level perceptual, circadian, and cognitive information. The metrics characterize the physiological signals available to the retina, rather than the physiological, psychological, and behavioural responses driven by these signals. One could imagine a mechanistic model combining the different photoreceptor signals for

predicting the non-visual impact of a display. However, more mechanistic modelling work is needed to quantify how the photoreceptor signals are combined to drive the non-visual consequences of light for such metrics to be developed. More work would be needed to quantify how a display that severely distorts melanopsin signals compares against a display that only minimally distorts melanopsin signals, and one that reproduces melanopsin signals exactly, in terms of, for example, the amount of melatonin suppression evoked by exposure to the display at a particular circadian phase. For now the metrics proposed here, which directly describe the physiologically relevant signals available to the retina, are an appropriate quantification of displays while the mechanisms driven by such signals remain under investigation.

A further limitation of the current metrics is that they do not consider absolute intensity. The proposed metrics do not take into consideration whether the relative photoreceptor signal quintuplet can be presented at the same intensity as they are found in the real-world spectra. Absolute intensity is not considered here as there is merit to quantifying the distortions introduced in the rod and melanopsin signals when the relative cone excitations can be reproduced, even when absolute intensity cannot be e.g. in the case of reproducing the daylight spectrum on a standard LCD monitor. If one wanted to incorporate absolute intensity into the metrics, then one could define the *PSRM* such that a photoreceptor quintuplet is only considered to be reproduced if it is reproduced at the same intensity as the real-world quintuplet. Alternatively one could report the proposed metrics alongside the intensity range of the display or quantify the absolute intensity range over which the three metrics proposed here apply.

A challenge of incorporating absolute intensity into the metrics proposed is knowing the absolute intensity range over which different real-world spectra are likely to exist. This is part of the more general requirement for a better characterization of the "spectral diet" most people experience (Webler et al., 2019). However, measuring the "spectral diet" in a way that captures intensity in a generalisable manner is tricky. Intensity depends on lighting geometry, as per the inverse square law $E = I/d^2$, which states that the irradiance of the illuminant in terms of intensity per unit area, E, varies in inverse proportion to the square of the distance, d, from the point source of intensity, I (Ryer et al., 1997). Additionally, there is a directional component to intensity in real-world spectra, such that it matters from which direction the observer is viewing the scene (Morimoto et al., 2019). These factors would need to be considered in any generalisable account of absolute intensity in the real-world reference set.

Here, the real-world reference dataset used was chosen as the reference because of its use in the

most recent derivation of the colour rendering index (David et al., 2015), and has previously been recognised as an appropriate standard reference set with its adoption in the IES TM-30-2015 standard. However, the current dataset does not include structural and interference colours, which, while less common, are particularly salient to our perceptual experience. The current real-world spectra dataset also does not weight the samples by likelihood of encountering a particular sample. More data are needed to understand the frequency of spectral encounters in the real-world (Webler et al., 2019; Cain et al., 2020). The metrics described in this paper do not rely on the choice of any particular reference dataset, just on a suitable reference dataset for the reproduction purpose. As more is understood about the spectral diet of humans, the reference dataset used to quantify reproduction here could be updated. It is also worth noting that a universal real-world reference set as used here may not be appropriate for all reproduction activity. For instance, if one wished to artificially decrease melanopsin signal in "night-modes" of displays then one may wish to define an altered real-world reference set to suit one's purpose.

A further challenge for display reproduction and quantification of display reproduction is the range of individual differences that exist in the photoreceptor spectral sensitivity functions, even amongst the normal vision population (as described in Chapter 2) (Asano et al., 2016). Differences in cone spectral sensitivity functions have long been of concern for display manufacturers, as extreme differences can result in metameric failure, i.e. a colour match on a display to a real-world stimulus for one individual that is not a colour match for another. As with the cone spectral sensitivities, one would expect differences in the rod and melanopsin spectral sensitivity functions driven by differences in lens, macular pigment, and optical density, though a characterization of the variance of rod and melanopsin functions remains to be derived, and at present only the standard functions are used. This means that even in displays where rod and melanopsin signals are reproduced without distortion for the standard observer, they may be distorted for different individual observers, though it is unclear how large these distortions would be in practice, and the impact that small distortions would have on the reproduction of non-visual features of light. It is worth noting that an advantage of my proposed metrics and visualizations is the ability to adapt an individual observer approach by simply using the individual spectral sensitivities as the basis for determining $\vec{A} = [A_L, A_M, A_S, A_R, A_I]$ and $\vec{A^*} = [A_L^*, A_M^*, A_S^*, A_R^*, A_I^*]$, and the MacLeod-Boynton axes rather than the standard observer spectral sensitivities.

6.4.3 Applying the metrics and visualization to multi-primary displays

It is first worth noting that the property of having sufficient primaries to control all five-photoreceptors alone is not enough to ensure a high *PSRM*, just as having an arbitrary three-primaries is not sufficient to give a large chromaticity gamut. This is demonstrated by the imperfect (<80%) *PSRM* for the hypothetical five-primary displays and the comparatively low *PSRM* for the research systems. In order to design five-primary displays capable of achieving a high *PSRM* one would need to carefully optimise selection of the primaries to maximise the *PSRM*. However, it is worth noting that it may not always be the user's goal to maximise the *PSRM* when designing five-primary systems and thus the applicability of the *PSRM* to evaluate the five-primary display depends on the use case of the display. For example, many five-primary systems built to date, such as the two custom-built systems described here, have been designed for specific research purposes, and use primaries that are optimized for that particular purpose, and thus may have a relatively low *PSRM* by consequence. The analysis of three-primary displays further emphasises this point as these displays are not designed to reproduce rod and melanopsin signals and so correspondingly the *PSRM* for just the cone signals (the chromaticity reproduction) is high but the *PSRM* over all five photoreceptors is low.

It is further worth reiterating that systems with five or more primaries are able to control all of the photoreceptor signals and thus by design, when driven using a control matrix that demands reproduction of all five photoreceptor signals, have a *PSDM* and *PCDM* of zero. The *PSDM* and *PCDM* are therefore only applicable for quantifying displays with fewer than five primaries, or a five-primary system that is not driven to control all five-photoreceptor signals. Correspondingly, the visualizations illustrated in Fig. 6.6 and 6.8 are only applicable when the *PSDM* and *PCDM* are non-zero. Therefore for the five-primary displays shown here (which are driven to control all photoreceptor signals) there is nothing to plot for those visualizations: they are simply used to illustrate the nature of the distortions introduced across chromaticity and between the photoreceptor signals when distortions are present. The proposed extended MacLeod-Boynton diagram is however applicable to any display (as shown in Fig. 6.9). I therefore imagine that this visualization (with the three projections in order to fully visualize the three-dimensional space) may be usefully plotted for each display evaluated, in the same way a chromaticity gamut would be useful.

I also foresee an application of the metrics to informing four-primary display design (Pokorny et al., 2004). As with three-primary displays, four-primary displays cannot reproduce all of the five photoreceptor signals, but can reproduce four out of five. Given the high correlation of the rod and melanopsin signals, there may be a reasonable trade-off between practical implementation and

reproduction that four-primary displays can offer i.e. controlling for the cone and melanopsin signals in a four-primary system may give a low *PSDM* and *PCDM* for the uncontrolled rod signals, and a high *PSRM* while requiring only four primaries, rather than five. I anticipate that four-primary display design could benefit from using the proposed metrics to optimise the primary choice and control matrix to maximise the *PSRM* and minimise the *PSDM* and *PCDM*.

6.5 Conclusion

In this chapter three novel, physiologically relevant display metrics to characterize the full photoreceptor response to light are proposed, along with a visualization of the photoreceptor response in the form of a new equal-luminance photoreceptor excitation diagram (an extension of the MacLeod-Boynton chromaticity diagram). These methods allow for the quantification of photoreceptor signals beyond the cones to evaluate the reproduction capabilities of different displays, and could be used to inform the optimal design of multi-primary displays for the reproduction of real-world photoreceptor signals. The three metrics also allow a characterization of the fundamental problem of photoreceptor signal distortions introduced by displays with fewer than five primaries, or five-primary displays that are not driven to control all five-photoreceptor signals. The metrics and visualizations proposed here are not envisaged to be the only metrics used in the future to quantify the full human response to light from displays, but rather, a starting point for evaluating photoreceptor signals currently ignored in common display metrics.

6.6 Acknowledgements

The text for this chapter has been largely adapted from a manuscript which has been released as a preprint on *BioRxiv* and is being prepared for submission to *Optics Express* (Hexley et al., 2021b). The preparation of this manuscript, including the analysis, and original draft writing was the original work of ACH. All authors of the preprint, Dr Takuma Morimoto, Dr Manuel Spitschan, Prof. Hannah Smithson, contributed to the conception of the work and editing of the manuscript. I thank all of my co-authors on this project for their contributions to the results presented in this chapter. The code for the analysis presented in this chapter (written in MATLAB R2021a (Mathworks, Natick, MA, USA)) is publicly available on GitHub (GitHub, Inc., San Francisco, CA, USA): https://github.com/AllieHexley/photoSim (Hexley et al., 2021a).

Chapter 7

General discussion

Contents

7.1 Chapter summaries 248

7.1.1 Chapter 2: Individual differences in cone photoreceptor signals 248

7.1.2 Chapter 3: Exploring the statistical regularities in photoreceptor signals from surfaces for illuminant estimation with machine-learning 248

7.1.3 Chapter 4: Online visual psychophysics: is it feasible to study photoreceptor signals with online experiments? 249

7.1.4 Chapter 5: Demonstrating a multi-primary high-dynamic range display for photoreceptor isolating experiments 250

7.1.5 Chapter 6: Novel metrics for characterizing the reproduction of photoreceptor signals 250

7.2 Future directions 251

7.2.1 The genetic underpinnings of individual differences in L:M ratio 251

7.2.2 Rod and ipRGC contributions to illuminant estimation 251

7.2.3 Utilising the MPHDR display to study the photoreceptor signals 252

7.2.4 Functional imaging of the human photoreceptor signals *in vivo* 253

7.2.5 Dynamics of the human photoreceptor signals 253

7.2.6 A complete model of the human photoreceptor signals 254

7.3 Thesis summary 254

This thesis presents a range of new methodologies, both in terms of custom-built equipment and experimental design, and in novel analytical approaches, to study, characterize, and quantify the human photoreceptor signals in novel ways. In this final chapter, I will briefly summarise the main

results presented in this thesis. I will discuss the limitations of the work presented and future directions from the current results. I will also briefly touch on what I foresee as the major directions for the future of the field, that will truly transform our current models of human photoreceptor signalling.

7.1 Chapter summaries

7.1.1 Chapter 2: Individual differences in cone photoreceptor signals

There are significant individual differences in the cone photoreceptor signals, yet we lack a quantification of the degree of this variability that can be explained by systematic differences between individuals. Chapter 2 investigated some of these systematic differences, namely the frequency of the L(S180A) polymorphism across ethnicity and L:M cone ratio across ethnicity, biological sex, and environmental light exposure. I presented an analysis of the large-scale gnomAD database, along with a review of previous literature, that supports the existence of a systematic ethnic bias in the frequency of the L(S180A) polymorphism, with the minor allele (L(A180)) being less common in African/African-American, East Asian, and South Asian ethnicities than in Caucasian or Latino/Admixed populations (Sharpe et al., 2005; McMahon et al., 2008; Hagen et al., 2019; Karczewski et al., 2020). I also reviewed past literature of differences in L:M ratio across ethnicity and biological sex, which suggest that L:M ratio is lower in African/African-Americans and East Asians than Caucasians, and that XX individuals have a higher L:M ratio than XY individuals (Hofer et al., 2005; Carroll et al., 2000; McMahon et al., 2008; Hagen et al., 2019; Kuchenbecker et al., 2014). In my own experimental results I did not find statistically significant differences in HFP score (as a proxy for L:M ratio) across biological sex or ethnicity, though due to the limited number of participants in my study (which was a consequence of the impact of the COVID-19 pandemic on in-person testing), this study was not sufficiently powered to uncover any such differences. I also investigated a relationship between season of birth and L:M ratio, to test the hypothesis that early environmental light exposure influences retinal development and preferential expression of a single cone type in low light conditions during development, and again did not find a statistically significant result (though again, did not have the statistical power to do so).

7.1.2 Chapter 3: Exploring the statistical regularities in photoreceptor signals from surfaces for illuminant estimation with machine-learning

Disambiguating the illuminant chromaticity from the chromaticity of surfaces in the scene is one of the biggest computational challenges the visual system faces. Mathematically speaking from a single

radiance sample it is impossible to disentangle the spectral contributions from the illuminant and the surface. However, there are gamut-based constraints that bound this problem: i.e. if the resultant radiance spectrum is very red, the illuminant can't have been strongly blue (Forsyth, 1990; Finlayson, 1996). Chapter 3 presented a machine-learning approach to quantify these statistical regularities, for a dataset of real-world surface and illuminant spectra. Specifically, I trained a random forest classifier to perform an illuminant classification task using just the cone photoreceptor signals generated by the radiance spectrum from which the classifier is trying to disentangle the illuminant. The classifier was able to learn regularities from the real-world surfaces and illuminants to perform above chance on the classification task, and when the classifier does misclassify the illuminant, it usually only does so by predicting the neighbouring class. By varying the illuminant classes, the surfaces on which the classifier is trained and tested, and the availability of the photoreceptor signals, I was able to probe what statistical properties are most informative for illuminant estimation. For example, when multiple surfaces are present in a scene, the optimal way to classify the illuminant, is to make a prediction from each surface and then take the average prediction.

7.1.3 Chapter 4: Online visual psychophysics: is it feasible to study photoreceptor signals with online experiments?

The COVID-19 pandemic began just over a year into my doctoral work, in March 2020, and caused mass disruption to in-person psychophysics testing worldwide. Many researchers turned to online testing to continue experimental work during the course of the pandemic (which was still ongoing and causing significant disruption to participant testing at the time of writing this thesis). This thesis is concerned with novel characterizations of the human photoreceptor signals, which require carefully controlled experimentation. Chapter 4 reviewed what is possible in terms of stimulus control for online visual psychophysics, and simulates what aspects of stimulus control are distorted when calibration is not precise. As such this chapter discussed the limits of what experimental work in vision science can be conducted remotely, and which experiments require the careful control that can only be offered in laboratories. In this chapter I concluded that there is great scope for conducting vision science experiments remotely thanks to various perceptual calibration techniques that are reviewed, but that, unfortunately, experiments probing human photoreceptor signals in the way this thesis strives to are not possible within the limits of remote testing.

7.1.4 Chapter 5: Demonstrating a multi-primary high-dynamic range display for photoreceptor isolating experiments

I presented in Chapter 5 the first implementation of a multi-primary high-dynamic range display system for studying the photoreceptor signals in isolation with spatio-temporal control across a large luminance range (Hexley, Yöntem et al., 2020). I described the control of the MPHDR display which utilises five independent parameter control of six effective primaries. Through measurements of the custom-built device, I demonstrated a maximum luminance output of 3200 cd/m^2, a full on/off contrast of 3,240,000:1, an ANSI contrast of 341,000:1, and an expanded colour gamut relative to sRGB displays. I also demonstrated the display's ability to generate melanopsin isolating stimuli over a broad range of luminance levels. I showed that the display is capable of generating metameric pairs of stimuli with a melanopsin contrast of 117% at 75 cd/m^{2}to a melanopsin contrast of 23% at 2000 cd/m^2. I also detailed future directions in which display architecture will be developed to enable full independent control of the six effective primaries in the MPHDR display.

7.1.5 Chapter 6: Novel metrics for characterizing the reproduction of photoreceptor signals

Rod and ipRGC signals have often been overlooked in quantitative models of vision, as it was historically assumed that these signals did not contribute to visual perception in bright light conditions. Yet, recent evidence suggests that these signals may indeed contribute to much of our daily visual experience (Zaidi et al., 2007; Zele, Feigl et al., 2018; Spitschan et al., 2017; Cao et al., 2018; Brown et al., 2012; Zele, Adhikari et al., 2018; Stockman & Sharpe, 2006; Zele & Cao, 2015). Additionally, it is clear that the human physiological response is not simply limited to vision, with ipRGC signals believed to be the main driver of circadian phase shifts and melatonin suppression, as well as influencing pupil size and mood (Spitschan, 2019b; Blume et al., 2019). Most displays have traditionally overlooked the need to control rod and melanopsin signals, and conventional colorimetry metrics do not quantify the reproduction of rod and melanopsin signals by displays. In Chapter 6 I presented three novel physiologically relevant metrics that can quantify the reproduction and distortion of the photoreceptor signals by visual displays. I additionally presented an extension of the well-known MacLeod-Boynton chromaticity diagram that includes a visualization of the melanopsin signals (Hexley et al., 2021b).

7.2 Future directions

The work presented in this thesis spurs many lines of future research, many of which I intend to pursue during postdoctoral work. I describe a few of these future directions in this section. Additionally, I use this section to discuss some of the additional future directions I foresee overhauling our models of human photoreceptor signals that do not follow directly from any of the experimental chapters presented in this thesis, namely the capability of adaptive optics retinal imaging technologies to characterise the retina in ways that are unachievable through any other technique. In fact much of my research in the past three years has focused on developing methodology and analysis paradigms to utilise data from an adaptive optics scanning laser ophthalmascope (AO-SLO) to image retinal function at high-resolution in the human eye *in vivo*. This work was not presented in the content of this thesis due to space constraints.

7.2.1 The genetic underpinnings of individual differences in L:M ratio

As described in detail in Chapter 2, one of the big mysteries in the differences in L:M ratio between individuals is the genetic underpinning of these differences (or environmental factors that influence L vs M expression during development). The two biggest candidates for genetic control of L:M ratio at present are 1) *RXRG*: the gene that codes for the RXRγ protein which is known to be involved cone differentiation (Lawrance-Owen et al., 2014; Onishi et al., 2010; Swaroop et al., 2010); and 2) the three-dimensional proximity of the LCR and *OPN1LW* and *OPN1MW* on the X chromosome array (McMahon et al., 2008). Future work is necessary to uncover which of these two candidates, if either, are responsible for the determination of L:M ratio in an individual, and whether (or perhaps more likely, how) these genetic underpinnings work hand-in-hand with environmental factors to influence the ontogenesis of the cone mosaic. I plan to address this in future work by recruiting more observers on my HFP experiment as described in Chapter 2, and genotyping *RXRG* and the upstream intron sequences in these observers, as well as collecting information on their early light exposure (by recording date and location of birth), to model the influence of these factors on L:M ratio control.

7.2.2 Rod and ipRGC contributions to illuminant estimation

A natural extension to the work presented in Chapter 3 would be to examine whether human observers are able to make use of the same statistical regularities the machine learning classifier is able to utilise to perform illuminant estimation. One particular experiment of pressing interest would be to

investigate whether a rod or melanopsin signal feeds into such higher-level colour processing. The results presented in Chapter 3 suggest that the availability of a rod or melanopsin signal would be advantageous for illuminant estimation. Thus a natural experiment to test whether such a rod or melanopsin signal is used to develop a percept of illuminant chromaticity would be to see if access to such information provided an advantage to the human observer in an illuminant estimation task. To test this, I plan to use the MPHDR display (Hexley, Yöntem et al., 2020), as described in Chapter 5, to vary melanopsin and rod signals while keeping the cone signals constant during an illuminant estimation task, to see whether we can observe the anticipated decrease in performance in the human observers as we see for the machine learning algorithm when the rod and ipRGC signals are not available.

7.2.3 Utilising the MPHDR display to study the photoreceptor signals

Chapter 5 describes the design, construction, and proof-of-principle of a multi-primary high-dynamic range display (MPHDR) display for photoreceptor isolating experiments. The obvious next step is to use the display to run a human psychophysics experiment. It had originally been planned that such an experiment would be completed as part of this DPhil, but due to the restrictions on in-person testing during the COVID-19 pandemic, this was unfortunately not possible. Thus such experiments will form the remit of future work, some of which I intend to pursue during postdoctoral work. As a first case, I propose the display be used to assess the melanopsin and luminance inputs to steady-state pupil size (Watson & Yellott, 2012). Specifically, I would test metameric pairs of stimuli at a given luminance such that one spectrum differs from the other in terms of melanopsin excitation, but not in terms of cone excitation, to see how stimuli with identical cone signals but differing melanopsin signals drive different steady-state pupil responses. This would thus provide a quantification of the interactions between the melanopsin and cone signals over a large operating range and enable one to determine the way in which photopic and melanopic excitations feed into pupil control across different mean light levels. This is just one of many such photoreceptor isolating experiments that are feasible on the MPHDR display, which I imagine will be an important tool for furthering our understanding of the human photoreceptor signals' contribution to many visual and non-visual light-mediated mechanisms.

7.2.4 Functional imaging of the human photoreceptor signals *in vivo*

One of the major transformational research directions of recent years has been the development of adaptive optics retinal imaging devices, such as the adaptive optics scanning laser ophthalmascope (AO-SLO), and adaptive optics optical coherence tomography (AO-OCT) (Roorda et al., 2002; Hermann et al., 2004; Marcos et al., 2017). Thus far these technologies have transformed our understanding of the structural arrangement of the cone photoreceptors in the human retina. However, these tools also provide the ability to capture *functional* activity at a cellular-scale. This functional measure is referred to as the optinoretinogram, and has been demonstrated in AO-OCT 3D slice recordings as a change in cone length (Pandiyan et al., 2020), and in AO-SLO images as an intensity change, where it is believed the cone acts as an interferometer and thus the length change leaves an intensity trace when viewed en-face (Cooper et al., 2017; Cooper et al., 2020). Work on characterizing the optinoretinogram is ongoing and new, with much work still to be done to characterize the nature of optinoretinogram under different stimulus conditions. The optinoretinogram is likely to prove a transformational tool as it is the first non-invasive measure of human photoreceptor activity at the cellular scale *in vivo*. It thus gives us the potential to study human cone function directly, and has major implications for potential use in answering many of the fundamental research questions raised in this thesis, and in diagnosis and treatment development for retinal disease.

7.2.5 Dynamics of the human photoreceptor signals

A major limitation of the way the photoreceptor signals are introduced and thought of in this thesis, and indeed in most accounts of the photoreceptor signals, is a lack of consideration of the spatio-temporal dynamics of retinal sampling. Here, we assumed the retinal image is a "snapshot" of a uniform external world at one point in time. But that is clearly an oversimplified model: the world is not uniform, the photoreceptor signals vary at different spatial locations across the retina due to both variation in the pre-receptoral filtering and photoreceptor distributions across the retina, and the fundamental nature of the retina as a discrete sampling array. Additionally, the photoreceptors are continually signalling and adapting (i.e. varying their sensitivity). Crucially, the external scene is continuously being translated over the retina due to fixational eye movements: eye movements that occur without conscious control, even when we are fixating. Without such fixational eye movements our perception of a stationary external scene would fade over time as the cones adapt. Fixational eye movement dynamics remain poorly understood, and thus future work is needed to develop models of this motion in order to feed this into models of the temporal dynamics of the photoreceptor signals.

Again, I envisage AO-SLO tools to be truly transformative here, as AO-SLO allows us to measure directly the movement of the retina during a fixation task, and thus characterise fixational eye movement dynamics at an unprecedentedly high-resolution that is matched to the cone mosaic sampling, and to the spatial scale of fixational eye movements.

7.2.6 A complete model of the human photoreceptor signals

Perhaps the best model currently available of the retinal level processes is the ISETBio framework (Cottaris et al., 2019). The ISETBio model considers the spatial extent of an external scene, how these are blurred through the eye's optics, and how these are translated into a pattern of cone excitations for a given cone mosaic. ISETBio also models the temporal persistence of the cone excitations by convolving the temporal pattern of cone excitations, which includes a model of fixational eye movements, with the cone photocurrent impulse response based on a model of phototransduction (Cottaris et al., 2019). The ISETBio model will be useful to model the cone photoreceptor signals in many scenarios. The ISETBio model has already been used to characterise ideal-observer pattern sensitivity and how individual differences between hypothetical observers would impact predicted performance on a pattern sensitivity task (Cottaris et al., 2019; Cottaris et al., 2020).

While ISETBio represents a significant step forward in the development of a model of human vision, ISETBio is still in its early days, and there is still much more we need to quantify about the human photoreceptor signals to improve the ISETBio model. As noted above, one such quantification is that of the dynamics of fixational eye movements. The cone mosaic model in the ISETBio is of a regular hexagonal mosaic, yet AO-SLO retinal imaging data have shown us that the cone mosaic is in fact only pseudo-hexagonal. Additionally, the ISETBio model at present does not incorporate the rod and ipRGC signals, and is solely focused on the cone signals. Future work in characterising the photoreceptor signals and their complex dynamics and interactions is thus still very much needed to inform the development of ISETBio into a more complete model of the human photoreceptor signals inputs to vision and non-visual light mediated neural mechanisms.

7.3 Thesis summary

The work presented in this thesis offers incremental steps forward in our understanding of the human photoreceptor signals and their influence on visual and non-visual light mediated neural mechanisms. Unfortunately, due to the impact of COVID-19 on restricting in-person human psychophysics testing

the psychophysical experimental work presented in this thesis is either 1) underpowered, and more data are needed to be able to draw conclusions from those experiments; or 2) missing where in an ideal world it would be present i.e. Chapter 5 would ideally have contained a psychophysical as well as a spectrophotometric demonstration of melanopsin isolation. The work in this thesis thus focused on utilising simulations and computational modelling tools and the exploitation of existing datasets in much of the analysis.

As a final summary, this thesis focused on five key experimental chapters. In the first, I investigated systematic drivers of individual differences in the cone photoreceptors. In the second, I presented a machine-learning approach to quantify the boundaries of illuminant perception using the photoreceptor signals from a real-world scene. In the third, I reported on the feasibility of using online psychophysics to study the photoreceptor signals. In the fourth, I presented a novel display technology which will enable previously unachievable tests of the photoreceptor signals in isolation. And in the fifth, I proposed new metrics and visualizations to quantify the reproduction of the photoreceptor signals on displays. This thesis thus offered novel methodology to study, and some novel characterizations of the human photoreceptor signals, with both breadth and depth in those characterizations. It is my belief and my hope that the work presented in this thesis thus provides useful contributions to the ways in which photoreceptor signals will be studied in the future, and to our understanding of the individual variability in, and the higher-level computations enabled by, the human photoreceptor signals.

Bibliography

Abramov, I., Gordon, J., Hendrickson, A., Hainline, L., Dobson, V. & LaBossiere, E. (1982). The retina of the newborn human infant. *Science*, *217*(4556), 265–267.

Adams, D. L. & Horton, J. C. (2003). The representation of retinal blood vessels in primate striate cortex. *Journal of Neuroscience*, *23*(14), 5984–5997.

Adobe Systems Incorporated. (2005). Adobe RGB (1998) Color Image Encoding.

Aguilar, M. & Stiles, W. (1954). Saturation of the Rod Mechanism of the Retina at High Levels of Stimulation. *Optica Acta: International Journal of Optics*, *1*(1), 1: 59–65.

Ajito, T., Obi, T., Yamaguchi, M. & Ohyama, N. (1999). Six-primary color projection display for expanded color gamut reproduction. *Multispectral Imaging*, *138*, 135–138.

Alexander, K. R. & Fishman, G. A. (1984). Rod-cone interaction in flicker perimetry. *British Journal of Ophthalmology*, *68*, 303–309.

Allen, A. E., Hazelhoff, E. M., Martial, F. P., Cajochen, C. & Lucas, R. J. (2018). Exploiting metamerism to regulate the impact of a visual display on alertness and melatonin suppression independent of visual appearance. *Sleep*, *41*(8), zsy100.

Allen, A. E., Martial, F. P. & Lucas, R. J. (2019). Form vision from melanopsin in humans. *Nature Communications*, *10*, 2274.

Álvaro, L., Linhares, J. M. M., Moreira, H., Lillo, J. & Nascimento, S. M. C. (2017). Robust colour constancy in red-green dichromats. *PLOS One*, *12*(6), 1–17.

Anstis, S. & Cavanagh, P. (1983). A minimum motion technique for judging equiluminance. *in Colour Vision: Psychophysics and Physiology, Mollon, J. D. and Sharpe, L. T., eds.*, 155–166.

Arendt, J. (2006). Melatonin and human rhythms. *Chronobiology International*, *23*(1-2), 21–37.

Asano, Y., Fairchild, M. & Blondé, L. (2016). Individual Colorimetric Observer Model. *PLOS One*, *11*, (2):e0145671.

Asenjo, A. B., Rim, J. & Oprian, D. D. (1994). Molecular determinants of human red/green color discrimination. *Neuron*, *12*(5), 5: 1131–1138.

Aston, S., Radonjić, A., Brainard, D. H. & Hurlbert, A. C. (2019). Illumination discrimination for chromatically biased illuminations: Implications for color constancy. *Journal of Vision*, *19*(3), 15–15.

Barrionuevo, P. A. & Cao, D. (2016). Luminance and chromatic signals interact differently with melanopsin activation to control the pupil light response. *Journal of Vision*, *16*(11), (11):29.

Baylor, D. A., Nunn, B. J. & Schnapf, J. L. (1987). Spectral sensitivity of cones of the monkey *Macaca fascicularis*. *Journal of Physiology*, *390*, 145–160.

Beck, J. (1959). Stimulus correlates for the judged illumination of a surface. *J Exp Psychol*, *58*, 267–274.

Beck, J. (1961). Judgments of surface illumination and lightness. *J Exp Psychol*, *61*, 368–375.

Belaid, N., van Overveld, I. M. C. J. & Martens, J.-B. (1997). Perceptual linearization as display standard: link between psychophysics and contrast discrimination models. In H. L. Kundel (Ed.), *Medical imaging 1997: Image perception* (pp. 130–142). SPIE.

Berns, R. S. (1996). Methods for characterizing CRT displays. *Displays*, *16*(4), 173–182.

Besuijen, J. (2007). Visual gamma measurement and methods to compare gamma models. *Journal of the Society for Information Display*, *15*(8).

Bieber, M. L., Kraft, J. M. & Werner, J. S. (1998). Effects of known variations in photopigments on L/M cone ratios estimated from luminous efficiency functions. *Vision Research*, *38*(13), 13: 1961–1966.

Birch, J. (2012). Worldwide prevalence of red-green color deficiency. *J. Opt. Soc. Am. A*, *29*(3), 313–320.

Bird, A. C. & Weale, R. A. (1974). On the retinal vasculature of the human fovea. *Experimental Eye Research*, *19*(5), 409–417.

Blume, C., Garbazza, C. & Spitschan, M. (2019). Effects of light on human circadian rhythms, sleep and mood. *Somnologie*, *23*, (3): 147–156.

Boehm, A. E., MacLeod, D. I. A. & Bosten, J. M. (2014). Compensation for red-green contrast loss in anomalous trichromats. *Journal of Vision*, *14*(13), 19–19.

Bone, R. A., Landrum, J. T., Guerra, L. H. & Ruiz, C. A. (2003). Lutein and zeaxanthin dietary supplements raise macular pigment density and serum concentrations of these carotenoids in humans. *J Nutr*, *133*(4), 992–998.

Bosten, J. M., Beer, R. D. & MacLeod, D. I. A. (2015). What is white? *Journal of Vision*, *15*(16), 5–5.

Bosten, J. (2019). The known unknowns of anomalous trichromacy. *Current Opinion in Behavioral Sciences*, *30*, 228–237.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Brainard, D. H. & Stockman, A. (2010). The Optical Society of America Handbook of Optics, 3rd edition, Volume III: Vision and Vision Optics. In M. Bass, C. DeCusatis, J. Enoch, V. Lakshminarayanan, G. Li, C. Macdonald, V. Mahajan & E. van Stryland (Eds.). New York: McGraw Hill.

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

Brainard, D. H. & Freeman, W. T. (1997). Bayesian color constancy. *J. Opt. Soc. Am. A, 14*(7), 1393–1411.

Brainard, D. H., Pelli, D. G. & Robson, T. (2002). Display characterization. *Signal Process, 80*, 2–067.

Brainard, D. H., Roorda, A., Yamauchi, Y., Calderone, J. B., Metha, A., Neitz, M., Neitz, J., Williams, D. R. & Jacobs, G. H. (2000). Functional consequences of the relative numbers of L and M cones. *J. Opt. Soc. Am. A, 17*(3), 607–614.

Brainard, G. C., Hanifin, J. P., Greeson, J. M., Byrne, B., Glickman, G., Gerner, E. & Rollag, M. D. (2001). Action spectrum for melatonin regulation in humans: Evidence for a novel circadian photoreceptor. *Journal of Neuroscience, 21,* (16): 6405–6412.

Braun, G. J. (2003). Visual display characterization using flicker photometry techniques. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human vision and electronic imaging VIII* (pp. 199–209). SPIE.

Brill, M. H. (1995). The relation between the color of the illuminant and the color of the illuminated object. *Color Research & Application, 20*(1), 70–76.

Brown, T. M. (2020). Melanopic illuminance defines the magnitude of human circadian light responses under a wide range of conditions. *Journal of Pineal Research, 69,* (1): e12655.

Brown, T. M., Tsujimura, S. I., Allen, A. E., Wynne, J., Bedford, R., Vickery, G., Vugler, A. & Lucas, R. J. (2012). Melanopsin-based brightness discrimination in mice and humans. *Current Biology, 22,* (12): 1134–1141.

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute, 310*(1), 1–26.

Burns, S. A. & Elsner, A. E. (1985). Color matching at high illuminances: The color-match-area effect and photopigment bleaching. *J. Opt. Soc. Am. A, 2*(5), 698–704.

Cain, S. W., McGlashan, E. M., Vidafar, P., Mustafovska, J., Curran, S. P. N., Wang, X., Mohamed, A., Kalavally, V. & Phillips, A. J. (2020). Evening home lighting adversely impacts the circadian system and sleep. *Scientific Reports, 10,* 19110.

Cajochen, C. (2007). Alerting effects of light. *Sleep Medicine Reviews, 11*(6), (6): 453–464.

Cajochen, C., Frey, S., Anders, D., Späti, J., Bues, M., Pross, A., Mager, R., Wirz-Justice, A. & Stefani, O. (2011). Evening exposure to a light-emitting diodes (LED)-backlit computer screen affects

circadian physiology and cognitive performance. *Journal of Applied Physiology, 110*, (5): 1432–1438.

Cajochen, C., Kräuchi, K. & Wirz-Justice, A. (2003). Role of melatonin in the regulation of human circadian rhythms and sleep. *Journal of Neuroendocrinology, 15*(4), 432–437.

Cao, D., Chang, A. & Gai, S. (2018). Evidence for an impact of melanopsin activation on unique white perception. *J. Opt. Soc. Am. A, 35*(4), (4): B287–B291.

Cao, D., Lee, B. B. & Sun, H. (2010). Combination of rod and cone inputs in parasol ganglion cells of the magnocellular pathway. *Journal of Vision, 10*(11), 4–4.

Cao, D., Nicandro, N. & Barrionuevo, P. A. (2015). A five-primary photostimulator suitable for studying intrinsically photosensitive retinal ganglion cell functions in humans. *Journal of Vision, 15*(1).

Carpenter, R. H. S. & Robson, J. G. (1998). *Vision research: A practical guide to laboratory methods*. Oxford University Press.

Carroll, J., Neitz, J. & Netiz, M. (2002). Estimates of L:M cone ratio from ERG flicker photometry and genetics. *Journal of Vision, 2(8)*, 531–542.

Carroll, J., McMahon, C., Neitz, M. & Neitz, J. (2000). Flicker-photometric electroretinogram estimates of L:M cone photoreceptor ratio in men with photopigment spectra derived from genetics. *J. Opt. Soc. Am. A, 17*(3), 3: 499–509.

Casagrande, V. A. & Xu, X. (2004). Parallel visual pathways: A comparative perspective. *Werner, J.S., Chalupa, L.S. (eds.) The New Visual Neurosciences* (pp. 494–506). MIT Press, Cambridge.

Casagrande, V., Yazar, F., Jones, K. & Ding, Y. (2007). The Morphology of the Koniocellular Axon Pathway in the Macaque Monkey. *Cerebral Cortex, 17*(10), 2334–2345.

Castańo, J. A. & Sperling, H. G. (1982). Sensitivity of the blue-sensitive cones across the central retina. *Vision Research, 22*(6), 661–673.

Charman, W. & Heron, G. (1988). Fluctuations in accommodation: A review. *Ophthalmic and Physiological Optics, 8*(2), 153–164.

Chellappa, S. L., Morris, C. J. & Scheer, F. A. (2020). Circadian misalignment increases mood vulnerability in simulated shift work. *Scientific Reports, 10*(1), 1–10.

Choi, H.-H., Kang, H.-S. & Yun, B.-J. (2020). Cnn-based illumination estimation with semantic information. *Applied Sciences, 10*(14).

CIE. (1932). Commission internationale de l'Eclairage proceedings, 1931. *Cambridge: Cambridge University Press*.

CIE. (2018). CIE System for Metrology of Optical Radiation for ipRGC-Influenced Responses to Light. CIE S 026/E:2018. *(Vienna: Central Bureau of the Commission Internationale de l'Éclairage)*.

CIE. (1976-a). Colorimetry. Part 4: CIE 1976 $L^*a^*b^*$ colour space. *(Vienna: Central Bureau of the Commission Internationale de l' Éclairage)*.

CIE. (1976-b). Colorimetry. Part 5: CIE 1976 $L^*u^*v^*$ colour space and $u'v'$ uniform chromaticity scale diagram. *(Vienna: Central Bureau of the Commission Internationale de l' Éclairage)*.

CIE. (2000). Colorimetry. Part 6: CIEDE2000 colour-difference formula. *(Vienna: Central Bureau of the Commission Internationale de l' Éclairage)*.

CIE. (2006). Fundamental chromaticity diagram with physiological axes. Part 1. Technical Report 170-1. CIE 170-1:2006. *(Vienna: Central Bureau of the Commission Internationale de l' Éclairage)*.

CIE. (2015). Fundamental chromaticity diagram with physiological axes. Part 2. Technical Report 170-2. CIE 170-2:2015. *(Vienna: Central Bureau of the Commission Internationale de l' Éclairage)*.

Ciulla, T. A., Curran-Celantano, J., Cooper, D. A., Hammond, B. R., Danis, R. P., Pratt, L. M., Riccardi, K. A. & Filloon, T. G. (2001). Macular pigment optical density in a midwestern sample. *Ophthalmology*, *108*(4), 730–737.

Collins, G. M., Bravo-Shugarman, M., Terasaki, P. I., Belzer, F. 0., Ashby, B. S., Dunphy, J. E., Frost, A. B., Ackerman, J., Tyree Finch, W., Manlove, A., Sinha, B. P., Atkinson, S. M., Pierce, Levy, J. M., Amer, M. V., Herrman, T. J., Turcotte, J. G., Anderson, M., Shear, C. T. G., ... Lines, E. J. (1972). Rods Cancel Cones in Flicker. *Nature*, *235*(2), 2: 173–174.

Colombo, E. & Derrington, A. M. (2001). Visual calibration of CRT monitors. *Displays*, *22*(3), 87–95.

Conner, J. & MacLeod, D. (1977). Rod photoreceptors detect rapid flicker. *Science*, *195*(4279), 698–699.

Cooper, R. F., Brainard, D. H. & Morgan, J. I. (2020). Optoretinography of individual human cone photoreceptors. *Optics Express*, *28*(26), 39326–39339.

Cooper, R. F., Tuten, W. S., Dubra, A., Brainard, D. H. & Morgan, J. I. (2017). Non-invasive assessment of human cone photoreceptor function. *Biomedical Optics Express*, *8*(11), 5098–5112.

Cornish, E. E., Xiao, M., Yang, Z., Provis, J. M. & Hendrickson, A. E. (2004). The role of opsin expression and apoptosis in determination of cone types in human retina. *Experimental Eye Research*, *78*(6), 1143–1154.

Cottaris, N. P., Jiang, H., Ding, X., Wandell, B. A. & Brainard, D. H. (2019). A computational-observer model of spatial contrast sensitivity: Effects of wave-front-based optics, cone-mosaic structure, and inference engine. *Journal of Vision*, *19*(4), 8–8.

Cottaris, N. P., Wandell, B. A., Rieke, F. & Brainard, D. H. (2020). A computational observer model of spatial contrast sensitivity: Effects of photocurrent encoding, fixational eye movements, and inference engine. *Journal of Vision*, *20*(7), 17–17.

Craven, B. & Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research*, *32*(7), 1359–1366.

Crawford. (1949). The scotopic visibility function. *Proceedings of the Physical Society*, *B62*, 321–334.

Crook, J. D., Davenport, C. M., Peterson, B. B., Packer, O. S., Detwiler, P. B. & Dacey, D. M. (2009). Parallel on and off cone bipolar inputs establish spatially coextensive receptive field structure of blue-yellow ganglion cells in primate retina. *Journal of Neuroscience*, *29*(26), 8372–8387.

Curcio, C. A., Allen, K. A., Sloan, K. R., Lerea, C. L., Hurley, J. B., Klock, I. B. & Milam, A. H. (1991). Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *J Comp Neurol*, *312*(4), 610–624.

Dacey, D. M. (1993). The mosaic of midget ganglion cells in the human retina. *Journal of Neuroscience*, *13*(12), 5334–5355.

Dacey, D. M. (1994). Physiology, morphology and spatial densities of identified ganglion cell types in primate retina. *Ciba Found Symp*, *184*, 12–28.

Dacey, D. M. (1996). Circuitry for color coding in the primate retina. *PNAS*, *93*(2), 582–588.

Dacey, D. M. & Lee, B. B. (1994). The 'blue-on' opponent pathway in primate retina originates from a distinct bistratified ganglion cell type. *Nature*, *367*(6465), 731–735.

Dacey, D. M., Liao, H.-W., Peterson, B. B., Robinson, F. R., Smith, V. C., Pokorny, J., Yau, K.-W. & Gamlin, P. D. (2005). Melanopsin-expressing ganglion cells in primate retina signal colour and irradiance and project to the LGN. *Nature*, *433*(7027), 749–754.

Dacey, D. M., Wool, L., Packer, O. & Wong, R. (2017). Confirmation of an S-OFF midget ganglion cell pathway using serial block-face scanning electron microscopy. *Journal of Vision*, *17*(7), 58–58.

Daly, S. J. & Lee, H. C. (U.S. Patent 5,754,222, May 1998). Visual characterization using display model.

Damberg, G., Gregson, J. & Heidrich, W. (2016). High Brightness HDR Projection Using Dynamic Freeform Lensing. *ACM Transactions on Graphics*, *35*(3), 1–11.

Danilova, M. V., Chan, C. H. & Mollon, J. D. (2013). Can spatial resolution reveal individual differences in the L:M cone ratio? *Vision Research*, *78*, 26–38.

Danilova, M. & Mollon, J. (2021). Bongard and Smirnov on the tetrachromacy of extra-foveal vision. *Vision Research*, 107952.

Dartnall, H. J. A., Bowmaker, J. K., Mollon, J. D. & Barlow, H. B. (1983). Human visual pigments: Microspectrophotometric results from the eyes of seven persons. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *220*(1218), 115–130.

David, A., Fini, P. T., Houser, K. W., Ohno, Y., Royer, M. P., Smet, K. A. G., Wei, M. & Whitehead, L. (2015). Development of the IES method for evaluating the color rendition of light sources. *Optics Express*, *23*(12), (12): 15888–15906.

Daw, N. W., Jensen, R. J. & Brunken, W. J. (1990). Rod pathways in mammalian retinae. *Trends in Neurosciences*, *13*(3), 110–115.

De Vries, H. (1948). The luminosity curve of the eye as determined by measurements with the flicker photometer. *Physica*, *14*, 319–348.

De Vries, H. (1949). The heredity of the relative numbers of red and green receptors in the human eye. *Genetica*, *24*(1), 199–212.

Deeb, S. S., Alvarez, A., Malkki, M. & Motulsky, A. G. (1995). Molecular patterns and sequence polymorphisms in the red and green visual pigment genes of japanese men. *Human Genetics*, *95*(5), 501–506.

Dees, E. W., Gilson, S. J., Neitz, M. & Baraas, R. C. (2015). The influence of L-opsin gene polymorphisms and neural ageing on spatio-chromatic contrast sensitivity in 20–71year olds. *Vision Research*, *116*, 13–24.

Delahunt, P. B., Webster, M. A., Ma, L. & Werner, J. S. (2004). Long-term renormalization of chromatic mechanisms following cataract surgery. *Visual Neuroscience*, *21*(3), 301–307.

Delange, H. (1952). Experiments on flicker and some calculations on an electrical analogue of the foveal systems. *Physica*, *18*(11), 935–950.

Delpero, W. T., O'Neill, H., Casson, E. & Hovis, J. (2005). Aviation-relevent epidemiology of color vision deficiency. *Aviation, space, and environmental medicine*, *76*(2), 127–133.

Derrington, A., Krauskopf, J. & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology*, *357*(1), 241–265.

Devlin, K., Chalmers, A. & Reinhard, E. (2006). Visual calibration and correction for ambient illumination. *ACM Transactions on Applied Perception*, *3*(4), 429–452.

Dickinson, S. J. (1999). Object representation and recognition. *What is Cognitive Science*, 172–207.

Do, M. T. H. & Yau, K.-W. (2010). Intrinsically photosensitive retinal ganglion cells. *Physiol Rev*, *90*(4), 1547–1581.

Duffy, J. F. & Czeisler, C. A. (2009). Effect of light on human circadian physiology. *Sleep Medicine Clinics*, *4*, (2): 165–177.

Ecker, J. L., Dumitrescu, O. N., Wong, K. Y., Alam, N. M., Chen, S.-K., LeGates, T., Renna, J. M., Prusky, G. T., Berson, D. M. & Hattar, S. (2010). Melanopsin-expressing retinal ganglion-cell photoreceptors: Cellular diversity and role in pattern vision. *Neuron*, *67*(1), 49–60.

Edgar, A. D. & Kasson, J. M. (U.S. Patent 5,298,993, Mar 1994). Display calibration.

Einabadi, F., Guillemaut, J.-Y. & Hilton, A. (2021). Deep neural models for illumination estimation and relighting: A survey. *Computer Graphics Forum*, *40*(6), 315–331.

Elliott, S. L. & Cao, D. (2013). Scotopic hue percepts in natural scenes. *Journal of Vision*, *13*(13), 15–15.

Elsner, A. E., Burns, S. A. & Webb, R. H. (1993). Mapping cone photopigment optical density. *J. Opt. Soc. Am. A*, *10*(1), 52–58.

Elze, T. & Tanner, T. G. (2012). Temporal properties of liquid crystal displays: Implications for vision science experiments. *PLOS One*, *7*(9), e44048.

Emery, K. J. & Webster, M. A. (2019). Individual differences and their implications for color perception. *Current Opinion in Behavioral Sciences*, *30*, 28–33.

Eskew Jr, R. T. (2009). Higher order color mechanisms: A critical review. *Vision Research*, *49*(22), 2686–2704.

Estevez, M. E., Fogerson, P. M., Ilardi, M. C., Borghuis, B. G., Chan, E., Weng, S., Auferkorte, O. N., Demb, J. B. & Berson, D. M. (2012). Form and function of the m4 cell, an intrinsically photosensitive retinal ganglion cell type contributing to geniculocortical vision. *Journal of Neuroscience*, *32*(39), 13608–13620.

Estévez, O. & Spekreijse, H. (1974). A spectral compensation method for determining flicker characteristics of the human colour mechanisms. *Vision Research*, *14*, 823–30.

Estévez, O. & Spekreijse, H. (1982). The "silent substitution" method in visual research. *Vision Research*, *22*, 681–91.

Estévez, O., Spekreijse, H., van Dalen, J. & Lunel, H. V. (1983). The OSCAR color vision test: Theory and evaluation (objective screening of color anomalies and reductions). *American Journal of Optometry and Physiological Optics*, *60*(11), 892–901.

F.lux software llc. (2011-2015). f.luxometer/rainbow

Fairchild, M. D. (2013). *Color appearance models*. John Wiley & Sons.

Field, G. D., Greschner, M., Gauthier, J. L., Rangel, C., Shlens, J., Sher, A., Marshak, D. W., Litke, A. M. & Chichilnisky, E. (2009). High-sensitivity rod photoreceptor input to the blue-yellow color opponent pathway in macaque retina. *Nature Neuroscience*, *12*(9), 1159–1164.

Figueiro, M. G., Wood, B., Plitnick, B. & Rea, M. S. (2011). The impact of light from computer monitors on melatonin levels in college students. *Neuroendocrinology Letters*, *32*(2), 158–163.

Finlayson, G. (1996). Color in perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(10), 1034–1038.

Finlayson, G. D., Hordley, S. D. & Hubel, P. M. (2001). Color by correlation: A simple, unifying framework for color constancy. *IEEE Trans. Pattern Anal. Mach. Intell.*, *23*, 1209–1221.

Flachot, A. & Gegenfurtner, K. R. (2018). Processing of chromatic information in a deep convolutional neural network. *J. Opt. Soc. Am. A*, *35*(4), B334–B346.

Flachot, A. C., Schuett, H. H., Fleming, R. W., Wichmann, F. & Gegenfurtner, K. R. (2019). Color constancy in deep neural networks. *Journal of Vision*, *19*(10), 298–298. https://doi.org/10.1167/19.10.298

Forsyth, D. A. (1990). A novel algorithm for color constancy. *International Journal of Computer Vision*, *5*(1), 5–35.

Foster, D. H. (2011). Color constancy. *Vision Research*, *51*(7), 674–700.

Foster, D. H. & Nascimento, S. M. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *257*(1349), 115–121.

Frederiksen, R., Morshedian, A., Tripathy, S. A., Xu, T., Travis, G. H., Fain, G. L. & Sampath, A. P. (2021). Rod photoreceptors avoid saturation in bright light by the movement of the g protein transducin. *Journal of Neuroscience*, *41*(15), 3320–3330.

Funt, B. & Finlayson, G. (1995). Color constant color indexing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*(5), 522–529.

Gamlin, P. D., McDougal, D. H., Pokorny, J., Smith, V. C., Yau, K. W. & Dacey, D. M. (2007). Human and macaque pupil responses driven by melanopsin-containing retinal ganglion cells. *Vision Research*, *47*, (7): 946–954.

Ganesan, S., Magee, M., Stone, J. E., Mulhall, M. D., Collins, A., Howard, M. E., Lockley, S. W., Rajaratnam, S. M. & Sletten, T. L. (2019). The impact of shift work on sleep, alertness and performance in healthcare workers. *Scientific Reports*, *9*(1), 1–13.

Ghodrati, M., Morris, A. P. & Price, N. S. C. (2015). The (un)suitability of modern liquid crystal displays (LCDs) for vision research. *Front Psychol*, *6*, 303.

Gijsenij, A., Gevers, T. & Van De Weijer, J. (2011). Computational color constancy: Survey and experiments. *IEEE transactions on image processing*, *20*(9), 2475–2489.

Gisbert, S., Feldkaemper, M., Wahl, S. & Schaeffel, F. (2020). Interactions of cone abundancies, opsin expression, and environmental lighting with emmetropization in chickens. *Experimental Eye Research*, *200*, 108205.

Gisbert, S. & Schaeffel, F. (2018). M to L cone ratios determine eye sizes and baseline refractions in chickens. *Experimental Eye Research*, *172*, 104–111.

Glasser, J. (1997). Principles of display measurement and calibration. *in: L.W.MacDonald, A.C. Lowe (Eds.), Display Systems: Design and Applications, JohnWiley Sons, Chichester*.

Gooley, J. J., Rajaratnam, S. M., Brainard, G. C., Kronauer, R. E., Czeisler, C. A. & Lockley, S. W. (2010). Spectral responses of the human circadian system depend on the irradiance and duration of exposure to light. *Science Translational Medicine*, *2*(31), 31ra33–31ra33.

Graham, D. M. & Wong, K. Y. (2016). *Melanopsin-expressing, intrinsically photosensitive retinal ganglion cells (ipRGCs). Webvision. Moran Eye Center*. Retrieved December 1, 2021, from https://webvision.med.utah.edu/book/part-ii-anatomy-and-physiology-of-the-retina/melanopsin-expressing-intrinsically-photosensitive-retinal-ganglion-cells/

Granzier, J. J. M., Brenner, E. & Smeets, J. B. J. (2009). Can illumination estimates provide the basis for color constancy? *Journal of Vision*, *9*(3), 18–18.

Granzier, J. J. M. & Valsecchi, M. (2014). Variations in daylight as a contextual cue for estimating season, time of day, and weather conditions. *Journal of Vision*, *14*(1), 22–22.

Grimes, W. N., Baudin, J., Azevedo, A. W. & Rieke, F. (2018). Range, routing and kinetics of rod signaling in primate retina. *Elife*, *7*, e38281.

Guild, J. (1931). The colorimetric properties of the spectrum. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *230*(681-693), 149–187.

Gunter, K., Neitz, M. & Neitz, J. (2006). L:M Cone Contribution to Heterochromatic Flicker Photometry. *Investigative Ophthalmology and Visual Science*, *47*, 3695.

Hack, I., Peichl, L. & Brandstätter, J. H. (1999). An alternative pathway for rod signals in the rodent retina: Rod photoreceptors, cone bipolar cells, and the localization of glutamate receptors. *PNAS*, *96*(24), 14130–14135.

Hagen, L. A., Arnegard, S., Kuchenbecker, J. A., Gilson, S. J., Neitz, M., Neitz, J. & Baraas, R. C. (2019). The association between L:M cone ratio, cone opsin genes and myopia susceptibility. *Vision Research*, *162*, 20–28.

Hammond, B. R., Wooten, B. R., Náñez, J. E. & Wenzel, A. J. (1999). Smoking and lens optical density. *Ophthalmic and Physiological Optics*, *19*(4), 300–305.

Hammond Jr, B. R., Curran-Celentano, J., Judd, S., Fuld, K., Krinsky, N. I., Wooten, B. R. & Snodderly, D. M. (1996). Sex differences in macular pigment optical density: Relation to plasma carotenoid concentrations and dietary patterns. *Vision Research*, *36*(13), 2001–2012.

Hannibal, J., Christiansen, A. T., Heegaard, S., Fahrenkrug, J. & Kiilgaard, J. F. (2017). Melanopsin expressing human retinal ganglion cells: Subtypes, distribution, and intraretinal connectivity. *J Comp Neurol*, *525*(8), 1934–1961.

Hart, W. (1987). The temporal responsiveness of vision. In R. Moses & W. Hart (Eds.), *Adler's physiology of the eye: Clinical application*. St. Louis: The C. V. Mosby Company.

Hayashi, S., Ueyama, H., Tanabe, S., Yamade, S. & Kani, K. (2001). Number and variations of the red and green visual pigment genes in japanese men with normal color vision. *Jpn J Ophthalmol*, *45*(1), 60–67.

He, J., Taveras-Cruz, Y. & Eskew, R. T. (2021). Modeling individual variations in equiluminance settings. *Journal of Vision*, *21*(7), 15–15.

Hecht, S., Shlaer, S. & Pirenne, M. H. (1942). Energy, quanta, and vision. *Journal of General Physiology*, *25*(6), 819–840.

Hecht, S., Shlaer, S., Smith, E. L., Haig, C. & Peskin, J. C. (1948). The visual functions of the complete colorblind. *Journal of General Physiology*, *31*(6), 459–472.

Helmholtz, H. (1866). *Handbuch der physiologishen optik 3rd edn*. [Transl. J. P. C. Southall 1962 *Helmholtz's treatise on physiological optics*, 2nd edn. New York: Dover.]

Hendrickson, A., Possin, D., Kwan, W., Huang, J. & Bourne, J. (2016). The temporal profile of retinal cell genesis in the marmoset monkey. *Journal of Comparative Neurology*, *524*(6), 1193–1207.

Hendrickson, A. & Zhang, C. (2019). Development of cone photoreceptors and their synapses in the human and monkey fovea. *Journal of Comparative Neurology*, *527*(1), 38–51.

Hendrickson, A. E. & Yuodelis, C. (1984). The morphological development of the human fovea. *Ophthalmology*, *91*(6), 603–612.

Hendry, S. H. C. & Reid, R. C. (2000). The koniocellular pathway in primate vision. *Annual Review of Neuroscience*, *23*(1), 127–153.

Hering, E. (1964). Outlines of a Theory of the Light Sense. *Harvard University Press*.

Hering, E. (1878). *Zur Lehre vom Lichtsinne: sechs Mittheilungen an die Kaiserl. Akademie der Wissenschaften in Wien*. C. Gerold's Sohn.

Hermann, B., Fernández, E., Unterhuber, A., Sattmann, H., Fercher, A., Drexler, W., Prieto, P. & Artal, P. (2004). Adaptive-optics ultrahigh-resolution optical coherence tomography. *Optics letters*, *29*(18), 2142–2144.

Hexley, A. C., Morimoto, T., Smithson, H. E. & Spitschan, M. (2021a). *Photosim: Physiologically-relevant simulations of real-world spectra and displays*. https://doi.org/10.5281/zenodo.5142677

Hexley, A. C., Dumbalska, T., Morimoto, T., Shiwen, L., Spitschan, M., Smithson, H. E. & Young, L. K. (2020). A review of perceptual calibration methods for online experiments with visual stimuli. *OSA Technical Group Special Webinar: Display Calibration for Internet and At-home Human Vision, Visual Perception and Color Research*.

Hexley, A. C., Morimoto, T., Smithson, H. E. & Spitschan, M. (2021b). Beyond colour gamuts: Novel metrics for the reproduction of photoreceptor signals. *bioRxiv*.

Hexley, A. C., Yöntem, A. Ö., Spitschan, M., Smithson, H. E. & Mantiuk, R. (2020). Demonstrating a multi-primary high dynamic range display system for vision experiments. *J. Opt. Soc. Am. A*, *37*(4), (4): A271–A284.

Hofer, H., Carroll, J., Neitz, J., Neitz, M. & Williams, D. R. (2005). Organization of the human trichromatic cone mosaic. *Journal of Neuroscience*, *25*(42), 9669–9679.

Hood, D. & Finkelstein, M. (1986). Sensitivity to light. In: K. Boff, L. Kaufman, J. Thomas (Eds.) *Handbook of Perception and Human Performance, 1*, (pp. 5-1-5–66) New York: Wiley.

Hood, S. M., Mollon, J. D., Purves, L. & Jordan, G. (2006). Color discrimination in carriers of color deficiency. *Vision Research*, *46*(18), 2894–2900.

Hordley, S. D. (2006). Scene illuminant estimation: Past, present, and future. *Color Research & Application*, *31*(4), 303–314.

Hornstein, E. P., Verweij, J., Li, P. H. & Schnapf, J. L. (2005). Gap-junctional coupling and absolute sensitivity of photoreceptors in macaque retina. *Journal of Neuroscience*, *25*(48), 11201–11209.

Houser, K. W., Wei, M., David, A., Kremers, M. R. & Shen, X. S. (2013). Review of measures for light-source color rendition and considerations for a two-measure system for characterizing color rendition. *Optics Express*, *21*, (8): 10393–10411.

Howells, O., Eperjesi, F. & Bartlett, H. (2013). Macular Pigment Optical Density in Young Adults of South Asian Origin. *Investigative Ophthalmology Visual Science*, *54*(4), 2711–2719.

Hu, M., Hu, Z., Xue, L., Yang, Z., Zheng, Z., He, D. & Zhang, X. (2011). Guinea pigs reared in a monochromatic environment exhibit changes in cone density and opsin expression. *Experimental Eye Research, 93*(6), 804–809.

Hu, Y. & Luo, M. R. (2019). Developing a visual method to characterize displays. *Color and Imaging Conference, 2019*(1), 75–79.

Huberman, A. D. & Niell, C. M. (2011). What can mice tell us about how vision works? *Trends Neurosci, 34*(9), 464–473.

Hurlbert, A. (2007). Colour constancy. *Current Biology, 17*(21), R906–7.

Iannaccone, A., Mura, M., Gallaher, K. T., Johnson, E. J., Todd, W. A., Kenyon, E., Harris, T. L., Harris, T., Satterfield, S., Johnson, K. C. & Kritchevsky, S. B. (2007). Macular Pigment Optical Density in the Elderly: Findings in a Large Biracial Midsouth Population Sample. *Investigative Ophthalmology Visual Science, 48*(4), 1458–1465.

International Electrotechnical Commission. (1999). IEC 61966-2-1:1999.

International Telecommunications Union. (2015). ITU-R BT.709-6: Parameter values for the HDTV standards for production and international programme exchange.

Isherwood, Z. J., Joyce, D. S., Parthasarathy, M. K. & Webster, M. A. (2020). Plasticity in perception: Insights from color vision deficiencies. *Fac Rev, 9*, 8.

Ives, H. (1912). The relation between the color of the illuminant and the color of the illuminated object (reprinted in: Color res. appl. 20, 70–75.) *Trans. Illuminant. Eng. Soc., 7*, 62–72.

Jacobs, G. H. (2008). Primate color vision: A comparative perspective. *Visual Neuroscience, 25*(5-6), 619–633.

Jägle, H., de Luca, E., Serey, L., Bach, M. & Sharpe, L. T. (2005). Visual acuity and X-linked color blindness. *Graefe's Archive for Clinical and Experimental Ophthalmology, 244*(4), 447.

Jägle, H., Heine, J. & Kurtenbach, A. (2006). L:M-cone ratio estimates of the outer and inner retina and its impact on sex differences in ERG amplitudes. *Doc Ophthalmol, 113*(2), 105–113.

Jameson, D. & Hurvich, L. M. (1959). Perceived color and its dependence on focal, surrounding, and preceding stimulus variables. *J. Opt. Soc. Am., 49*(9), 890–898.

Jordan, G. & Mollon, J. D. (1997). Sons and mothers: classification of colour-deficient and heterozygous subjects by counterphase modulation photometry. *Colour Vision Deficiencies XIII*, 385–392.

Jordan, G., Deeb, S. S., Bosten, J. M. & Mollon, J. (2010). The dimensionality of color vision in carriers of anomalous trichromacy. *Journal of vision, 10*(8), 12–12.

Judd, D. B., MacAdam, D. L., Wyszecki, G., Budde, H. W., Condit, H. R., Henderson, S. T. & Simonds, J. L. (1964). Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.*, *54*(8), 1031–1040.

Kalloniatis, M. & Luu, C. (2007a). *Temporal resolution. Webvision. Moran Eye Center*. Retrieved December 1, 2021, from https://webvision.med.utah.edu/book/part-viii-psychophysics-of-vision/temporal-resolution/

Kalloniatis, M. & Luu, C. (2007b). *Visual acuity. Webvision. Moran Eye Center*. Retrieved December 1, 2021, from https://webvision.med.utah.edu/book/part-viii-psychophysics-of-vision/visual-acuity/

Karatzas, D. & Wuerger, S. M. (2007). A hardware-independent colour calibration technique. *Annals of the British Machine Vision Association*, (3), 1–10.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., ... Consortium, G. A. D. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*, (7809):434–443.

Katz, D. (1936). The World of Colour, Translated by R.B. Macleod and C.W. Fox. *Philosophy*, *11*(43).

Kauvar, I., Yang, S. J., Shi, L., McDowall, I. & Wetzstein, G. (2015). Adaptive color display via perceptually-driven factored spectral projection. *ACM Transactions on Graphics*, *34*(6), 1–10.

Kim, U. S., Mahroo, O. A., Mollon, J. D. & Yu-Wai-Man, P. (2021). Retinal ganglion cells—diversity of cell types and clinical relevance. *Frontiers in Neurology*, *12*, 635.

Kling, A., Field, G. D., Brainard, D. H. & Chichilnisky, E. J. (2019). Probing computation in the primate visual system at single-cone resolution. *Annu Rev Neurosci*, *42*, 169–186.

Klooster, J., Vrensen, G., Müller, L. & Van der Want, J. (1995). Efferent projections of the olivary pretectal nucleus in the albino rat subserving the pupillary light reflex and related reflexes a light microscopic tracing study. *Brain Research*, *688*(1-2), 34–46.

Kolb, H. (2007). *Facts and figures concerning the human retina. Webvision. Moran Eye Center*. Retrieved December 1, 2021, from https://webvision.med.utah.edu/book/part-xiii-facts-and-figures-concerning-the-human-retina/

Kolb, H. (2012a). *Photoreceptors. Webvision. Moran Eye Center*. Retrieved December 1, 2021, from https://webvision.med.utah.edu/book/part-ii-anatomy-and-physiology-of-the-retina/photoreceptors/

Kolb, H. (2012b). *Simple anatomy of the retina. Webvision. Moran Eye Center*. Retrieved December 1, 2021, from https://webvision.med.utah.edu/book/part-i-foundations/simple-anatomy-of-the-retina/

Kovesi, P. (2015). Good colour maps: How to design them.

Kraft, J. M. & Brainard, D. H. (1999). Mechanisms of color constancy under nearly natural viewing. *PNAS of the United States of America*, *96*(1), 307–312.

Krauskopf, J., Williams, D. R. & Heeley, D. W. (1982). Cardinal directions of color space. *Vision Research*, *22*(9), 1123–1131.

Krauskopf, J., Williams, D., Mandler, M. & Brown, A. M. (1986). Higher order color mechanisms. *Vision Research*, *26*(1), 23–32.

Kremers, J., Scholl, H. P., Knau, H., Berendschot, T. T., Usui, T. & Sharpe, L. T. (2000). L/M cone ratios in human trichromats assessed by psychophysics, electroretinography, and retinal densitometry. *J. Opt. Soc. Am. A*, *17*(3), 517–526.

Krinov, E. L. (1947). Spectral reflectance properties of natural formations. *National Research Council of Canada*.

Kuchenbecker, J. A., Neitz, J. & Neitz, M. (2014). Ethnic variation in the ratio of long- to middle-wavelength sensitive cones. *Investigative Opthalmology and Visual Science*, *55*(13), 4539.

Kwak, Y. & MacDonald, L. (2000). Characterization of a desktop LCD projector. *Displays*, *21*(5), 179–194.

Lagroix, H. E. P., Yanko, M. R. & Spalek, T. M. (2012). LCDs are better: Psychophysical and photometric estimates of the temporal characteristics of crt and lcd monitors. *Atten Percept Psychophys*, *74*(5), 1033–1041.

Lall, G. S., Revell, V. L., Momiji, H., Enezi, J. A., Altimus, C. M., Güler, A. D., Aguilar, C., Cameron, M. A., Allender, S., Hankins, M. W. & Lucas, R. J. (2010). Distinct contributions of rod, cone, and melanopsin photoreceptors to encoding irradiance. *Neuron, 66*, 417–428.

Lamb, T. D. (2016). Why rods and cones? *Eye*, *30*(2), 179–185.

Laming, D. (1991). Contrast Sensitivity. Chapter 5. In S. LeVay, S. Nelson & J. Cronly-Dillon (Eds.), *Vision and visual dysfunction, Vol 5*. London: Macmillan Press.

Land, E. H. (1977). The retinex theory of color vision. *Scientific american*, *237*(6), 108–129.

Langenbucher, A. (2015). Emmetropia, definition. In U. Schmidt-Erfurth & T. Kohnen (Eds.), *Encyclopedia of ophthalmology* (pp. 1–1). Springer Berlin Heidelberg.

Lavin, Y., Silverstein, A. D. & Zhang, X. (1999). Visual experiment on the Web. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human vision and electronic imaging iv* (pp. 278–289). SPIE.

Lawrance-Owen, A. J., Bosten, J. M., Hogg, R. E., Bargary, G., Goodbourn, P. T. & Mollon, J. D. (2014). Counterphase modulation flicker photometry: phenotypic and genotypic associations. *J. Opt. Soc. Am. A, 31*(4), 4: A34–A37.

Lee, B. B., Smith, V. C., Pokorny, J. & Kremers, J. (1997). Rod inputs to macaque ganglion cells. *Vision Research, 37*(20), 2813–2828.

Lee, B., Martin, P. & Valberg, A. (1988). The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *The Journal of Physiology, 404*(1), 323–347.

Lee, B., Martin, P. & Valberg, A. (1989). Sensitivity of macaque retinal ganglion cells to chromatic and luminance flicker. *The Journal of Physiology, 414*(1), 223–243.

Lennie, P., Haake, P. W. & Williams, D. R. (1991). The design of chromatically opponent receptive fields. *Computational models of visual processing*, 71–82.

Lennie, P., Pokorny, J. & Smith, V. C. (1993). Luminance. *J. Opt. Soc. Am. A, 10*, 6: 1283–1293.

Lesmes, L. A. (2010). Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. *Journal of Vision, 10*(3), 1–21.

Lewis, A. & Zhaoping, L. (2006). Are cone sensitivities determined by natural color statistics? *Journal of Vision, 6*(3), 285–302.

Lewy, A. J., Wehr, T. A., Goodwin, F. K., Newsome, D. A. & Markey, S. (1980). Light suppresses melatonin secretion in humans. *Science, 210*(4475), 1267–1269.

Li, G., Yoshiura, K., Welander, U., Sällström, P. & McDavid, W. (2002). Visual linearization of the display of digital radiographs. *Dentomaxillofacial Radiology, 31*(2), 131–136.

Liao, H. W., Ren, X., Peterson, B. B., Marshak, D. W., Yau, K. W., Gamlin, P. D. & Dacey, D. M. (2016). Melanopsin-expressing ganglion cells on macaque and human retinas form two morphologically distinct populations. *Journal of Comparative Neurology, 524*, 2845–2872.

Liu, H., Prugnolle, F., Manica, A. & Balloux, F. (2006). A geographically explicit genetic model of worldwide human-settlement history. *The American Journal of Human Genetics, 79*(2), 230–237.

Lok, R., Smolders, K. C., Beersma, D. G. & de Kort, Y. A. (2018). Light, alertness, and alerting effects of white light: A literature overview. *Journal of Biological Rhythms, 33*, (6): 589–601.

Lucas, R. J., Peirson, S. N., Berson, D. M., Brown, T. M., Cooper, H. M., Czeisler, C. A., Figueiro, M. G., Gamlin, P. D., Lockley, S. W., O'Hagan, J. B. et al. (2014). Measuring and using light in the melanopsin age. *Trends in Neurosciences, 37*(1), 1–9.

Lutze, M. & Bresnick, G. H. (1991). Lenses of diabetic patients "yellow" at an accelerated rate similar to older normals. *Invest Ophthalmol Vis Sci*, *32*(1), 194–199.

Lutze, M., Cox, N. J., Smith, V. C. & Pokorny, J. (1990). Genetic studies of variation in rayleigh and photometric matches in normal trichromats. *Vision Research*, *30*(1), 149–162.

Lyon, M. (1961). Gene action in the X-chromosome of the mouse. *Nature*, *190*, 372–373.

MacLeod, D. I. A. & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *J. Opt. Soc. Am. A*, *69*(8), (8): 1183–1186.

Maloney, L. & Yang, J. (2003). The illumination estimation hypothesis and surface color perception. In R. Mausfeld & D. Heyer (Eds.), *Colour* (pp. 335–358). Oxford University Press.

Maloney, L. T. & Wandell, B. A. (1986). Color constancy: A method for recovering surface spectral reflectance. *J. Opt. Soc. Am. A*, *3*(1), 29–33.

Mantiuk, R. K. (2016). Perceptual display calibration. In: Displays: Fundamentals Applications. A K Peters/CRC Press, Rolf R. Hainich and Oliver Bimber (ed.)

Marcos, S., Werner, J. S., Burns, S. A., Merigan, W. H., Artal, P., Atchison, D. A., Hampson, K. M., Legras, R., Lundstrom, L., Yoon, G. et al. (2017). Vision science and adaptive optics, the state of the field. *Vision Research*, *132*, 3–33.

Martinovic, J. (2014). Magno-, parvo-, koniocellular pathways. In R. Luo (Ed.), *Encyclopedia of color science and technology* (pp. 1–5). Springer Berlin Heidelberg.

McCamy, C. S., Marcus, H. & Davidson, J. G. (1976). A color-rendition chart. *J. Appl. Photogr. Eng.*, *2*(3), 95–99.

McMahon, C., Carroll, J., Awua, S., Neitz, J. & Neitz, M. (2008). The L:M cone ratio in males of African descent with normal color vision. *Journal of Vision*, *8*(2), 5–5.

Merbs, S. L. & Nathans, J. (1992). Absorption spectra of the hybrid pigments responsible for anomalous color vision. *Science*, *258*(5081), 464–466.

Merbs, S. L., and Nathans, J. (1992). Absorption spectra of human cone pigments. *Nature*, *356*, 433–435.

Miller, M. E. & Yang, J. (2003). Visual display characterization using temporally modulated patterns. *Journal of the Society for Information Display*, *11*(1), 183.

Milosavljevic, N. (2019). How does light regulate mood and behavioral state? *Clocks & Sleep*, *1*(3), 319–331.

Mollon, J. D. (1999). Color vision: Opsins and options. *PNAS*, *96*(9), 4743–4745.

Mollon, J. (1997). '... aus dreyerley arten von membranen oder molekülen': George palmer's legacy. *Colour vision deficiencies XIII* (pp. 3–20). Springer.

Mollon, J. & Polden, P. (1978). On the time constants of tachistoscopes. *The Quarterly Journal of Experimental Psychology*, *30*(3), 555–568.

Mollon, J. (2006). Monge: the Verriest Lecture, Lyon, July 2005. *Visual Neuroscience*, *23*(3-4), 297–309.

Mollon, J. D. (2003). Origins of modern color science. In S. K. Shevell (Ed.), *The science of color. 2nd edn.* (pp. 1–39). Elsevier/Optical Society of America;Amsterdam/London.

Moore, R. Y. (2007). Suprachiasmatic nucleus in sleep–wake regulation. *Sleep Medicine*, *8*, 27–33.

Moreland, J. & Alexander, E. (1997). Effect of macular pigment on colour matching with field sizes in the 1° to 10° range. In C. R. Cavonius (Ed.), *Colour vision deficiencies xiii: Proceedings of the thirteenth symposium of the international research group on colour vision deficiencies, held in pau, france july 27–30, 1995* (pp. 363–368). Springer Netherlands.

Morimoto, T., Kishigami, S., Linhares, J. M., Nascimento, S. M. & Smithson, H. E. (2019). Hyperspectral environmental illumination maps: Characterizing directional spectral variation in natural environments. *Optics Express*, *27*(22), 32277–32293.

Morimoto, T., Kusuyama, T., Fukuda, K. & Uchikawa, K. (2021). Human color constancy based on the geometry of color distributions. *Journal of Vision*, *21*(3), 7–7.

Mulligan, J. B. (2009). Presentation of calibrated images over the web. *Human Vision and Electronic Imaging XIV*, *7240*, 724006.

Münch, M. & Kawasaki, A. (2013). Intrinsically photosensitive retinal ganglion cells: Classification, function and clinical implications. *Curr Opin Neurol*, *26*(1), 45–51.

Mure, L. S. (2021). Intrinsically photosensitive retinal ganglion cells of the human retina. *Frontiers in Neurology*, *12*.

Nagare, R., Rea, M. S., Plitnick, B. & Figueiro, M. G. (2019). Effect of white light devoid of “cyan” spectrum radiation on nighttime melatonin suppression over a 1-h exposure duration. *Journal of Biological Rhythms*, *34*, (2): 195–204.

Nagel, W. (1924). Appendix: Adaptation, twilight vision and the duplicity theory. *Helmholtz's Treatise on Physiological Optics Translated from the Third German Edition by JPC Southall*, 313–343.

Nascimento, S. M., Amano, K. & Foster, D. H. (2016). Spatial distributions of local illumination color in natural scenes. *Vision Research*, *120*, 39–44.

Nathans, J., Merbs, S. L., Sung, C. H., Weitz, C. J. & Wang, Y. (1992). Molecular genetics of human visual pigments. *Annu Rev Genet*, *26*, 403–424.

Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B. & Hogness, D. S. (1986). Molecular genetics of inherited variation in human color vision. *Science*, *232*(4747), 203–210.

Nathans, J., Thomas, D. & Hogness, D. S. (1986). Molecular genetics of human color vision: The genes encoding blue, green, and red pigments. *Science*, *232*(4747), 193–202.

Nathans, J. (1992). Rhodopsin: Structure, function, and genetics. *Biochemistry*, *31*(21), 4923–4931.

Neitz, J. & Neitz, M. (2011). The genetics of normal and defective color vision. *Vision Research*, *51*(7), 633–651.

Neitz, M. & Neitz, J. (1998). Molecular genetics and the biological basis of color vision. In W. G. K. Backhaus, R. Kliegl & J. S. Werner (Eds.), *Color vision: Perspectives from different disciplines* (pp. 101–119). Berlin: Walter de Gruyter.

Neitz, M. & Neitz, J. (2000). Molecular genetics of color vision and color vision defects. *Archives of Ophthalmology*, *118*(5), 691–700.

Neumann, A., Artusi, A., Neumann, L., Zotti, G. & Purgathofer, W. (2006). Accurate display gamma functions based on human observation. *Color Research and Application*, *32*(4), 310–319.

Neumann, A., Artusi, A., Zotti, G., Neumann, L. & Purgathofer, W. (2003). An interactive perception-based model for characterization of display devices. In R. Eschbach & G. G. Marcu (Eds.), *Color Imaging IX: Processing, Hardcopy, and Applications* (pp. 232–241). SPIE.

Nolan, J., O'Donovan, O., Kavanagh, H., Stack, J., Harrison, M., Muldoon, A., Mellerio, J. & Beatty, S. (2004). Macular pigment and percentage of body fat. *Invest Ophthalmol Vis Sci*, *45*(11), 3940–3950.

Nowozin, C., Wahnschaffe, A., Rodenbeck, A., de Zeeuw, J., Hadel, S., Kozakov, R., Schopp, H., Munch, M. & Kunz, D. (2017). Applying melanopic lux to measure biological light effects on melatonin suppression and subjective sleepiness. *Current Alzheimer Research*, *14*(10), 1042–1052.

Nygaard, R. W. & Schuchard, R. A. (2001). SLO radiant power and brightness. *Journal of Rehabilitation Research and Development*, *38*(1), 123–8.

Nyquist, H. (1924). Certain factors affecting telegraph speed. *Bell System Technical Journal*, *3*(2), 324–346.

Onishi, A., Peng, G.-H., Chen, S. & Blackshaw, S. (2010). Pias3-dependent sumoylation controls mammalian cone photoreceptor differentiation. *Nature Neuroscience*, *13*(9), 1059–1065.

Ostrin, L. A., Abbott, K. S. & Queener, H. M. (2017). Attenuation of short wavelengths alters sleep and the ip rgc pupil response. *Ophthalmic and Physiological Optics*, *37*(4), 440–450.

Packer, O., Diller, L. C., Verweij, J., Lee, B. B., Pokorny, J., Williams, D. R., Dacey, D. M. & Brainard, D. H. (2001). Characterization and use of a digital light projector for Vision Research. *Vision Research*, *41*, (4) 427–439.

Pandiyan, V. P., Maloney-Bertelli, A., Kuchenbecker, J. A., Boyle, K. C., Ling, T., Chen, Z. C., Park, B. H., Roorda, A., Palanker, D. & Sabesan, R. (2020). The optoretinogram reveals the primary steps of phototransduction in the living human eye. *Science Advances*, *6*(37), eabc1124.

Parraga, A. C., Roca-Vila, J., Karatzas, D. & Wuerger, S. (2014). Limitations of visual gamma corrections in LCD displays. *Displays*, *35*(5), 227–239.

Patterson, S. S., Kuchenbecker, J. A., Anderson, J. R., Neitz, M. & Neitz, J. (2020). A color vision circuit for non-image-forming vision in the primate retina. *Current Biology*, *30*(7), 1269–1274.

Patterson, S. S., Mazzaferri, M. A., Bordt, A. S., Chang, J., Neitz, M. & Neitz, J. (2020). Another blue-on ganglion cell in the primate retina. *Current Biology*, *30*(23), R1409–R1410.

Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D. & Hurlbert, A. (2014). Chromatic illumination discrimination ability reveals that human colour constancy is optimised for blue daylight illuminations. *PLOS One*, *9*(2), 1–10.

Peli, E. (1992). Display nonlinearity in digital image processing for visual communications. *Optical Engineering*, *31*(11), 2374.

Percival, K. A., Jusuf, P. R., Martin, P. R. & Grünert, U. (2009). Synaptic inputs onto small bistratified (blue-on/yellow-off) ganglion cells in marmoset retina. *Journal of Comparative Neurology*, *517*(5), 655–669.

Pokorny, J., Smithson, H. & Quinlan, J. (2004). Photostimulator allowing independent control of rods and the three cone types. *Visual Neuroscience*, *21*, (3) 263–267.

Pokorny, J., Lutze, M., Cao, D. & Zele, A. J. (2006). The color of night: Surface color perception under dim illuminations. *Visual Neuroscience*, *23*(3-4), 525–530.

Pokorny, J., Lutze, M., Cao, D. & Zele, A. J. (2008). The color of night: Surface color categorization by color defective observers under dim illuminations. *Visual Neuroscience*, *25*(3), 475–480.

Pokorny, J. & Smith, V. C. (1976). Effect of field size on red–green color mixture equations. *J. Opt. Soc. Am.*, *66*(7), 705–708.

Polyak, S. L. (1941). *The Retina: The Anatomy and the Histology of the Retina in Man, Ape and Monkey, Including the Consideration of Visual Functions, the History of Physiological Optics and the Histological Laboratory Technique*. The University of Chicago Press.

Poynton, C. (2008). ColorChecker ('Macbeth') chart. *poynton.com*.

Prayag, A. S., Najjar, R. P. & Gronfier, C. (2019). Melatonin suppression is exquisitely sensitive to light and primarily driven by melanopsin in humans. *Journal of Pineal Research*, *66*(4), e12562.

Provencio, I., Rodriguez, I., Jiang, G., Hayes, W., Moreira, E. & Rollag, M. (2000). A novel human opsin in the inner retina. *Journal of Neuroscience*, *20*, (2) 600–605.

Purkinje, J. E. (1823). *Beobachtungen und versuche zur physiologie der sinne* (Vol. 1). In Commission der JG Calve'schen Buchhandlung.

Purkinje, J. (1819). Beobachtungen und versuche zur physiolgie der sinne: Beiträge zur kenntniss des schens in subjectiver hinsicht [observations and experiments on the physiology of the senses: Contributions to the knowledge of vision in its subjective aspect]. *Prague, Czechoslovakia: Calve*.

Race, Ethnicity, and Genetics Working Group et al. (2005). The use of racial, ethnic, and ancestral categories in human genetics research. *The American Journal of Human Genetics*, *77*(4), 519–532.

Radonjić, A., Ding, X., Krieger, A., Aston, S., Hurlbert, A. C. & Brainard, D. H. (2018). Illumination discrimination in the absence of a fixed surface-reflectance layout. *Journal of Vision*, *18*(5), 11–11.

Radonjić, 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), 2–2.

Ravikumar, S., Akeley, K. & Banks, M. S. (2011). Creating effective focus cues in multi-plane 3D displays. *Optics Express*, *19*(21), (21): 20940–20952.

Ribelayga, C., Cao, Y. & Mangel, S. C. (2008). The circadian clock in the retina controls rod-cone coupling. *Neuron*, *59*(5), 790–801.

Rohr, M. & Wagner, A. (2020). How monitor characteristics affect human perception in visual computer experiments: CRT vs. LCD monitors in millisecond precise timing research. *Scientific Reports*, *10*(1), 6962.

Roorda, A., Metha, A. B., Lennie, P. & Williams, D. R. (2001). Packing arrangement of the three cone classes in primate retina. *Vision Research*, *41*, 1291–1306.

Roorda, A., Romero-Borja, F., Donnelly III, W. J., Queener, H., Hebert, T. J. & Campbell, M. C. (2002). Adaptive optics scanning laser ophthalmoscopy. *Optics Express*, *10*(9), 405–412.

Rushton, W. A. H. (1972). Review lecture. Pigments and signals in colour vision. *The Journal of Physiology*, *220*(3), 1–31.

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

Ryer, A., Light, U. & Light, V. (1997). Light measurement handbook.

Sabesan, R., Hofer, H. & Roorda, A. (2015). Characterizing the human cone photoreceptor mosaic via dynamic photopigment densitometry. *PLOS One*, *10*(12), 12: 1–12.

Schiller, P. H. (2010). Parallel information processing channels created in the retina. *PNAS*, *107*(40), 17087–17094.

Schmidt, T. M., Alam, N. M., Chen, S., Kofuji, P., Li, W., Prusky, G. T. & Hattar, S. (2014). A role for melanopsin in alpha retinal ganglion cells and contrast detection. *Neuron*, *82*(4), 781–788.

Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L., Trentacoste, M., Ghosh, A. & Vorozcovs, A. (2004). High Dynamic range display systems. *ACM Trans. Graph.*, *23*(3), 3, 760–768.

Seo, J.-S., Yeom, T.-E. & Ko, J.-H. (2012). Experimental and Simulation Study of the Optical Performances of a Wide Grid Polarizer as a Luminance Enhancement Film for LCD Backlight Applications. *J. Opt. Soc. Korea*, *16*, 151–156.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423.

Shapiro, A. G., Pokorny, J. & Smith, V. C. (1996). Cone-rod receptor spaces with illustrations that use CRT phosphor and light-emitting-diode spectra. *J. Opt. Soc. Am. A*, *13*, (12): 2319–2328.

Sharma, G. & Bala, R. (2017). *Digital color imaging handbook*. CRC press.

Sharma, G., Wu, W. & Dalal, E. N. (2005). The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, *30*(1), 21–30.

Sharpe, L. T., Stockman, A., Jagla, W. & Jägle, H. (2005). A luminous efficiency function, V*(λ), for daylight adaptation. *Journal of Vision*, *5*(11), 3–3.

Sharpe, L. T. & Stockman, A. (1999). Rod pathways: The importance of seeing nothing. *Trends in Neurosciences*, *22*(11), (11): 497–504.

Sharpe, L. T., Stockman, A., Jägle, H., Knau, H., Klausen, G., Reitner, A. & Nathans, J. (1998). Red, Green, and Red-Green Hybrid Pigments in the Human Retina: Correlations between Deduced Protein Sequences and Psychophysically Measured Spectral Sensitivities. *Journal of Neuroscience*, *18*, 23: 10053–10069.

Sharpe, L. T., Stockman, A. & MacLeod, D. I. (1989). Rod flicker perception: Scotopic duality, phase lags and destructive interference. *Vision Research*, *29*, 1539–1559.

Simunovic, M. (2010). Colour vision deficiency. *Eye*, *24*(5), 747–755.

Smallwood, P. M., Wang, Y. & Nathans, J. (2002). Role of a locus control region in the mutually exclusive expression of human red and green cone pigment genes. *PNAS*, *99*(2), 1008–1011.

Smith, V. C. & Pokorny, J. (1975). Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm. *Vision Research*, *15*(2), 161–171.

Smith, V. C. & Pokorny, J. (2003). 3 - color matching and color discrimination. In S. K. Shevell (Ed.), *The science of color (second edition)* (Second Edition, pp. 103–148). Elsevier Science Ltd.

Smithson, H. E. (2005). Sensory, computational and cognitive components of human colour constancy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *360*(1458), 1329–1346.

Smithson, H. & Zaidi, Q. (2004). Colour constancy in context: Roles for local adaptation and levels of reference. *Journal of Vision*, *4*(9), 3–3.

Snodderly, D. M., Weinhaus, R. S. & Choi, J. (1992). Neural-vascular relationships in central retina of macaque monkeys (macaca fascicularis). *Journal of Neuroscience*, *12*(4), 1169–1193.

Solomon, S. G. (2021). Retinal ganglion cells and the magnocellular, parvocellular, and koniocellular subcortical visual pathways from the eye to the brain. In J. J. Barton & A. Leff (Eds.), *Neurology of Vision and Visual Disorders* (pp. 31–50). Elsevier.

Soucy, E., Wang, Y., Nirenberg, S., Nathans, J. & Meister, M. (1998). A novel signaling pathway from rod photoreceptors to ganglion cells in mammalian retina. *Neuron*, *21*(3), 481–493.

Souman, J. L., Borra, T., de Goijer, I., Schlangen, L. J., Vlaskamp, B. N. & Lucassen, M. P. (2018). Spectral tuning of white light allows for strong reduction in melatonin suppression without changing illumination level or color temperature. *Journal of Biological Rhythms*, *33*, (4): 420–431.

Spitschan, M. (2019a). Photoreceptor inputs to pupil control. *Journal of Vision*, *19*, (9):5, 1–5.

Spitschan, M., Gardasevic, M., Martial, F., Lucas, R. & Allen, A. (2019). Pupil responses to hidden photoreceptor-specific modulations in movies. *PLOS One*, *14*, (5):e0216307.

Spitschan, M. & Woelders, T. (2018). The Method of Silent Substitution for Examining Melanopsin Contributions to Pupil Control. *Front. Neurol.*, *9*, 941.

Spitschan, M. (2019b). Melanopsin contributions to non-visual and visual function. *Current Opinion in Behavioral Sciences*, *30*, 67–72.

Spitschan, M., Aguirre, G. K. & Brainard, D. H. (2015). Selective stimulation of penumbral cones reveals perception in the shadow of retinal blood vessels. *PLoS One*, *10*(4), e0124328.

Spitschan, M., Aguirre, G. K., Brainard, D. H. & Sweeney, A. M. (2016). Variation of outdoor illumination as a function of solar elevation and light pollution. *Scientific Reports*, *6*(1), 26756.

Spitschan, M., Bock, A. S., Ryan, J., Frazzetta, G., Brainard, D. H. & Aguirre, G. K. (2017). The human visual cortex response to melanopsin-directed stimulation is accompanied by a distinct perceptual experience. *PNAS*, *114*, (46): 12291–12296.

Spitschan, M., Jain, S., Brainard, D. H. & Aguirre, G. K. (2014). Opponent melanopsin and S-cone signals in the human pupillary light response. *PNAS*, *111*(43), (43): 15568–15572.

Sterling, P., Cohen, E., Freed, M. A. & Smith, R. G. (1987). Microcircuitry of the on-beta ganglion cell in daylight, twilight and star light. *Neurosci. Res. Suppl*, *6*, S269–S285.

Sterling, P., Freed, M. A. & Smith, R. G. (1988). Architecture of rod and cone circuits to the on-beta ganglion cell. *Journal of Neuroscience*, *8*(2), 623–642.

Stiles, W. S. & Burch, J. M. (1959). N.P.L. Colour-matching investigation: Final report (1958). *Optica Acta: International Journal of Optics*, *6*(1), 1: 1–26.

Stiles, W. & Burch, J. (1955). Interim report to the commission internationale de l'éclairage, zurich, 1955, on the national physical laboratory's investigation of colour-matching (1955). *Optica Acta: International Journal of Optics*, *2*(4), 168–181.

Stockman, A., Jagle, H., Pirzer, M. & Sharpe, L. T. (2008). The dependence of luminous efficiency on chromatic adaptation. *Journal of Vision*, *8*, 16:1, 1–26.

Stockman, A. & Sharpe, L. T. (2000). The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research*, *40*, (13): 1711–1737.

Stockman, A., Sharpe, L. T. & Fach, C. C. (1999). The spectral sensitivity of the human short-wavelength cones. *Vision Research*, *39*, (17): 2901–2927.

Stockman, A. (2019). Cone fundamentals and CIE standards. *Current Opinion in Behavioral Sciences*, *30*, 87–93.

Stockman, A. & Sharpe, L. T. (2006). Into the twilight zone: The complexities of mesopic vision and luminous efficiency. *Ophthalmic and Physiological Optics*, *26*(3), (3): 225–239.

Stricker, S. & Power, H. (1872). The eye: The retina. *Manual of human and comparative histology* (p. 666). London :The New Sydenham Society,

Sugiyama, T. & Kudo, Y. (2005). The display profiling method with eyes and colormatching accuracy in soft proofing. In R. Eschbach & G. G. Marcu (Eds.), *Color Imaging X: Processing, Hardcopy, and Applications* (pp. 385–392). SPIE.

Sugiyama, T. & Kudo, Y. (2006). The calibration accuracy of display white point by visual calibrator under various illuminations. In R. Eschbach & G. G. Marcu (Eds.), *Color Imaging XI: Processing, Hardcopy, and Applications* (pp. 190–199). SPIE.

Sun, H., Pokorny, J. & Smith, V. C. (2001a). Control of the modulation of human photoreceptors. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, *26*(S1), S69–S75.

Sun, H., Pokorny, J. & Smith, V. C. (2001b). Rod-cone interactions assessed in inferred magnocellular and parvocellular postreceptoral pathways. *Journal of Vision*, *1*(1), 5–5.

Sun, H., Smithson, H. E., Zaidi, Q. & Lee, B. B. (2006a). Do magnocellular and parvocellular ganglion cells avoid short-wavelength cone input? *Visual Neuroscience*, *23*(3-4), 441–446.

Sun, H., Smithson, H. E., Zaidi, Q. & Lee, B. B. (2006b). Specificity of cone inputs to macaque retinal ganglion cells. *Journal of Neurophysiology*, *95*(2), 837–849.

Sung, C.-H., Davenport, C. M., Hennessey, J. C., Maumenee, I. H., Jacobson, S. G., Heckenlively, J. R., Nowakowski, R., Fishman, G., Gouras, P. & Nathans, J. (1991). Rhodopsin mutations in autosomal dominant retinitis pigmentosa. *PNAS*, *88*(15), 6481–6485.

Swaroop, A., Kim, D. & Forrest, D. (2010). Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nature Reviews Neuroscience*, *11*(8), 563–576.

Szmajda, B. A., Grünert, U. & Martin, P. R. (2008). Retinal ganglion cell inputs to the koniocellular pathway. *Journal of Comparative Neurology*, *510*(3), 251–268.

Teplitz, R. L. (1965). Sex chromatin of cone cells of human retina. *Science*, *150*(3705), 1828–1829.

Thibos, L. N., Hong, X., Bradley, A. & Applegate, R. A. (2004). Accuracy and precision of objective refraction from wavefront aberrations. *Journal of Vision*, *4*(4), 9.

Thibos, L. N., Ye, M., Zhang, X. & Bradley, A. (1992). The chromatic eye: A new reduced-eye model of ocular chromatic aberration in humans. *Applied Optics*, *31*(19), 3594.

Thomas, J. B. & Gerhardt, J. (2012). Webcam based display calibration. *Final Program and Proceedings - IS and T/SID Color Imaging Conference*, 82–87.

Thomas, P. B. M. & Mollon, J. D. (2004). Modelling the rayleigh match. *Visual Neuroscience*, *21*, 477–482.

Thorlabs. (2020a). *Bandpass filter, central wavelength 500 nm, full-width half-maximum 40 nm transmission data*.

Thorlabs. (2020b). *Notch filter, central wavelength 488 nm transmission data*.

Thorlabs. (2020c). *Notch filter, central wavelength 514 nm transmission data*.

To, L., Woods, R. L., Goldstein, R. B. & Peli, E. (2013). Psychophysical contrast calibration. *Vision Research*, *90*, 15–24.

To, L., Woods, R. L. & Peli, E. (2009). 17.3: Visual Calibration of Displays for Accurate Contrast Reproduction.

Tominaga, S., Ebisui, S. & Wandell, B. A. (2001). Scene illuminant classification: Brighter is better. *J. Opt. Soc. Am. A*, *18*(1), 55–64.

Tregillus, K. E., Isherwood, Z. J., Vanston, J. E., Engel, S. A., MacLeod, D. I., Kuriki, I. & Webster, M. A. (2021). Color compensation in anomalous trichromats assessed with fmri. *Current Biology*, *31*(5), 936–942.

Tsujimura, S. & Tokuda, Y. (2011). Delayed response of human melanopsin retinal ganglion cells on the pupillary light reflex. *Ophthalmic & Physiological Optics*, *31*, (5): 469–479.

Tsujimura, S., Ukai, K., Ohama, D., Nuruki, A. & Yunokuchi, K. (2010). Contribution of human melanopsin retinal ganglion cells to steady-state pupil responses. *Proceedings of the Royal Society B Biological Sciences*, *277*, 2485–92.

Tyler, C. (1997). Colour bit-stealing to enhance the luminance resolution of digital displays on a single pixel basis. *Spatial Vision*, *10/4*, 369–377.

Uchikawa, K., Fukuda, K., Kitazawa, Y. & MacLeod, D. I. A. (2012). Estimating illuminant color based on luminance balance of surfaces. *J. Opt. Soc. Am. A*, *29*(2), A133–A143.

Vartanian, G. V., Li, B. Y., Chervenak, A. P., Walch, O. J., Pack, W., Ala-Laurila, P. & Wong, K. Y. (2015). Melatonin suppression by light in humans is more sensitive than previously reported. *Journal of Biological Rhythms*, *30*(4), 351–354.

Verweij, J., Dacey, D., Peterson, B. & Buck, S. (1999). Sensitivity and dynamics of rod signals in H1 horizontal cells of the macaque monkey retina. *Vision Research*, *39*(22), 3662–3672.

Vollrath, D., Nathans, J. & Davis, R. W. (1988). Tandem array of human visual pigment genes at Xq28. *Science*, *240*(4859), 1669–1672.

von Kries, J. (1905). Influence of adaptation on the effects produced by luminous stimuli. In M. (D. L. (Ed.), *Sources of color science* (pp. 120–126). Boston, MA: MIT Press.

Wagner, G. & Boynton, R. M. (1972). Comparison of Four Methods of Heterochromatic Photometry. *J. Opt. Soc. Am. A*, *62*(12), 12: 1508–1515.

Wald, G. (1945). Human vision and the spectrum. *Science*, *101*, 653–658.

Wang, Y., Macke, J. P., Merbs, S. L., Zack, D. J., Klaunberg, B., Bennett, J., Gearhart, J. & Nathans, J. (1992). A locus control region adjacent to the human red and green visual pigment genes. *Neuron*, *9*(3), 429–440.

Wang, Y., Smallwood, P. M., Cowan, M., Blesh, D., Lawler, A. & Nathans, J. (1999). Mutually exclusive expression of human red and green visual pigment-reporter transgenes occurs at high frequency in murine cone photoreceptors. *PNAS*, *96*(9), 5251–5256.

Watson, A. B. & Yellott, J. I. (2012). A unified formula for light-adapted pupil size. *Journal of Vision*, *12*(10), 12–12.

Webler, F. S., Spitschan, M., Foster, R. G., Andersen, M. & Peirson, S. N. (2019). What is the 'spectral diet' of humans? *Current Opinion in Behavioral Sciences*, *30*, 80–86.

Webster, M. A. & Macleod, D. I. A. (1988). Factors underlying individual differences in the color matches of normal observers. *J. Opt. Soc. Am. A*, *5*(10), 10: 1722–35.

Webster, M. A. (2015). Individual differences in color vision.

Webster, M. A., Miyahara, E., Malkoc, G. & Raker, V. E. (2000). Variations in normal color vision. II. Unique hues. *J. Opt. Soc. Am. A*, *17*(9), 1545–1555.

Webster, M. A., Mizokami, Y. & Webster, S. M. (2007). Seasonal variations in the color statistics of natural images. *18*, 3: 213–233.

Webster, M. (2014). Environmental influences on color vision. In R. Luo (Ed.), *Encyclopedia of color science and technology* (pp. 1–6). Springer Berlin Heidelberg.

Weiss, D., Witzel, C. & Gegenfurtner, K. (2017). Determinants of colour constancy and the blue bias. *i-Perception*, *8*(6), 2041669517739635.

Welbourne, L. E., Morland, A. B. & Wade, A. R. (2015). Human colour perception changes between seasons. *Current Biology*, *25*(15), R646–R647.

Went, L. & Pronk, N. (1985). The genetics of tritan disturbances. *Human Genetics*, *69*(3), 255–262.

Werner, J. S. & Schefrin, B. E. (1993). Loci of achromatic points throughout the life span. *J. Opt. Soc. Am. A*, *10*(7), 1509–1516.

Westheimer, G. (1966). The maxwellian view. *Vision Research*, *6*, (11-12) 669–682.

Wiesel, T. N. & Hubel, D. H. (1966). Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *J Neurophysiol*, *29*(6), 1115–1156.

Winderickx, J., Battisti, L., Hibiya, Y., Motulsky, A. G. & Deeb, S. S. (1993). Haplotype diversity in the human red and green opsin genes: Evidence for frequent sequence exchange in exon 3. *Hum Mol Genet*, *2*, (9):1413–1421.

Winderickx, J., Lindsey, D. T., Sanocki, E., Teller, D. Y., Motulsky, A. G. & Deeb, S. S. (1992). Polymorphism in red photopigment underlies variation in colour matching. *Nature*, *356*(6368), 431–433.

Witzel, C., Olkkonen, M. & Gegenfurtner, K. R. (2018). A Bayesian model of the memory colour effect. *i-Perception*, *9*(3), 2041669518771715.

Wood, A. W., Loughran, S. P. & Stough, C. (2006). Does evening exposure to mobile phone radiation affect subsequent melatonin production? *International Journal of Radiation Biology*, *82*(2), 69–76.

Wood, B., Rea, M. S., Plitnick, B. & Figueiro, M. G. (2013). Light level and duration of exposure determine the impact of self-luminous tablets on melatonin suppression. *Applied Ergonomics*, *44*(2), 237–240.

Woods, A. T., Velasco, C., Levitan, C. A., Wan, X. & Spence, C. (2015). Conducting perception research over the internet: A tutorial review. *PeerJ*, *3*, e1058.

Wool, L. E., Crook, J. D., Troy, J. B., Packer, O. S., Zaidi, Q. & Dacey, D. M. (2018). Nonselective wiring accounts for red-green opponency in midget ganglion cells of the primate retina. *Journal of Neuroscience*, *38*(6), 1520–1540.

Wool, L. E., Packer, O. S., Zaidi, Q. & Dacey, D. M. (2019). Connectomic identification and three-dimensional color tuning of s-off midget ganglion cells in the primate retina. *Journal of Neuroscience*, *39*(40), 7893–7909.

Wright, H. R. & Lack, L. C. (2001). Effect of light wavelength on suppression and phase delay of the melatonin rhythm. *Chronobiology International*, *18*(5), 801–808.

Wright, H. R., Lack, L. C. & Kennaway, D. J. (2004). Differential effects of light wavelength in phase advancing the melatonin rhythm. *Journal of Pineal Research*, *36*(2), 140–144.

Wright, W. (1930). A re-determination of the mixture curves of the spectrum. *Transactions of the Optical Society*, *31*(4), 201.

Wright, W. (1952). The characteristics of tritanopia. *J. Opt. Soc. Am.*, *42*(8), 509–521.

Wright, W. D. (1929). A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society*, *30*(4), 141.

Wyszecki, G. & Stiles, W. (1982). *Color science: Concepts and methods, quantitative data and formulae.* (2^{nd} edition), Wiley.

Xiao, K., Fu, C., Karatzas, D. & Wuerger, S. (2011). Visual gamma correction for LCD displays. *Displays*, *32*(1), 17–23.

Xiao, M. & Hendrickson, A. (2000). Spatial and temporal expression of short, long/medium, or both opsins in human fetal cones. *Journal of Comparative Neurology*, *425*(4), 545–559.

Yamada, E. S., Silveira, L. C. & Perry, V. H. (1996). Morphology, dendritic field size, somal size, density, and coverage of m and p retinal ganglion cells of dichromatic cebus monkeys. *Visual Neuroscience*, *13*(6), 1011–1029.

Yamauchi, Y., Katsu, K., Kuchenbecker, J., Neitz, M. & Neitz, J. (2013). L:M cone ratio of japanese derived with ERG flicker photometry method. *Investigative Opthalmology and Visual Science*, *54*(15), 6119.

Young, L. K. & Smithson, H. E. (2021). Emulated retinal image capture (ERICA) to test, train and validate processing of retinal images. *Scientific Reports*, *11*(1), 11225.

Young, T. (1802). II. The Bakerian Lecture. on the theory of light and colours. *Philosophical Ttransactions of the Royal Society of London*, (92), 12–48.

Yung, A., Cardoso-Leite, P., Dale, G., Bavelier, D. & Green, C. (2015). Methods to test visual attention online. *Journal of Visualized Experiments*, *96*.

Yuodelis, C. & Hendrickson, A. (1986). A qualitative and quantitative analysis of the human fovea during development. *Vision Research*, *26*(6), 847–855.

Zaidi, F., Hull, J., Peirson, S., Wulff, K., Aeschbach, D., Gooley, J., Brainard, G., Gregory-Evans, K., Rizzo, J., Czeisler, C., Foster, R., Moseley, M. & Lockley, S. (2007). Short-wavelength light sensitivity of circadian, pupillary, and visual awareness in humans lacking an outer retina. *Current Biology*, *17*, (24): 2122–2128.

Zaidi, Q. (1998). Identification of illuminant and object colors: Heuristic-based algorithms. *J. Opt. Soc. Am. A*, *15*(7), 1767–1776.

Zele, A., Feigl, B., Adhikari, P., Maynard, M. & Cao, D. (2018). Melanopsin photoreception contributes to human visual detection, temporal and colour processing. *Scientific Reports*, *8*, 3842.

Zele, A. J., Adhikari, P., Feigl, B. & Cao, D. (2018). Cone and melanopsin contributions to human brightness estimation. *J. Opt. Soc. Am. A*, *35*, B19–B25.

Zele, A. J. & Cao, D. (2015). Vision under mesopic and scotopic illumination. *Frontiers in Psychology*, *5: 1584*, (1594).

Zele, A. J., Kremers, J. & Feigl, B. (2012). Mesopic rod and s-cone interactions revealed by modulation thresholds. *J. Opt. Soc. Am. A*, *29*(2), A19–A26.

Zhang, F., Kurokawa, K., Lassoued, A., Crowell, J. A. & Miller, D. T. (2019). Cone photoreceptor classification in the living human eye from photostimulation-induced phase dynamics. *PNAS, 116*, 16: 7951–7956.

Zhou, N., Atchison, D. A., Zele, A. J., Brown, B. & Schmid, K. L. (2015). Cone ratios in myopia and emmetropia: A pilot study. *Optom Vis Sci*, *92*(1), e1–5.