The retinal mosaic, eye movements and colour vision

Sarah Elizabeth Regan

A thesis submitted for the degree of

Doctor of Philosophy



Pembroke College

Department of Experimental Psychology

University of Oxford

Trinity 2020

Abstract

Human colour vision is trichromatic: it is underpinned by the comparison of signals from three different classes of cone photoreceptor, those maximally sensitive to long (L), middle (M) and short (S) wavelengths. It can therefore be represented in a three-dimensional colour space in which the three axes represent the signals or combinations of signals from the three cone classes. Physiological data showing how these cone signals are combined and compared has motivated the creation of several important colour spaces, notably MacLeod-Boynton colour space, in which one axis of the space represents midget ganglion cells' comparison of L and M signals (Lee, Kremers & Yeh, 1998), while the other represents the small bistratified cells' comparison of S signals with summed L and M signals (Dacey & Lee, 1994). This space, along with other two- and three-dimensional colour spaces, have shaped the way researchers have imagined and investigated colour vision.

However, while physiological measurements as far as the early cortex are consistent with linear combinations of cone inputs (Derrington, Krauskopf & Lennie, 1984; Solomon & Lennie, 2007; Sun, Smithson, Zaidi & Lee 2006), psychophysics has revealed a number of important effects that cannot be captured in any Euclidean colour space. Chapter 2 focuses on hue (the subjective quality of appearing, for example, blue, bluish-green or green) and saturation (the subjective counterpart of colorimetric purity, where a saturated colour can be desaturated to a pastel shade by adding white). The 'super-importance of hue' is a nonlinearity of colour vision in which a smaller physical stimulus change is required to detect a change in hue than to detect a change in saturation (Judd, 1968). Chapter 2 presents the first explicit test of Kuehni's 2003 claim that hue super-importance is the result of separate neural channels for hue and saturation, a claim which was ultimately not supported by the data.

Further irregularities arise from individual differences in the cone mosaic that forms the gateway to the visual system. The cones of the three classes are not equal in number, and between individuals the ratio of L and M cones varies widely. Such differences are exacerbated locally when cones of the same class 'clump' together, which generally occurs to the degree that would be expected in a truly randomly distributed retina (Hofer, Carroll, Neitz, Neitz & Williams, 2005; Roorda, Metha, Lennie &

Williams, 2001). This should cause differences in spatiochromatic resolution within and between observers, but to date this has only been demonstrated in the fovea in a single observer with an extremely imbalanced L:M cone ratio (Miyahara, Pokorny, Smith, Baron & Baron, 1998) and in the parafovea (Danilova, Chan & Mollon, 2013). Chapter 3 reports the first evidence for an effect of L:M ratio on foveal spatiochromatic thresholds within and between observers who have more moderate L:M ratios.

Until recently, information about the human cone mosaic could only come from histology, but the introduction of adaptive optics (AO) technology to retinal imaging has revolutionised our ability to investigate living, seeing human retinae at high resolution. This technology allows eye movements to be measured on an unprecedentedly fine scale referenced to the photoreceptor mosaic (Young, Hauperich, Morris, Saunter & Smithson, 2017), as described in Chapter 4. Chapter 5 describes how this capacity has been used to investigate a possible functional role for eye movements in chromatic vision. Due to the unequal numbers of cones of different classes and local clumping, fine chromatic detail that can be detected easily by one part of the retina may not be detected by another. Therefore, the hypothesis was tested that eye movements may help observers to overcome the irregularity of the retinal mosaic by translating the image across the retina and ensuring that the image is sampled by all three classes of cone. While visual inspection of certain trends was promising, no significant associations were found between particular amounts of eye movement and spatiochromatic performance, between eye movement and L:M ratio, or between eye movement and scene content.

AO has more generally been used to extract important statistics about areas of retina, such as cone density, cone spacing and anisotropy in cone packing. However, the high volume of data generated requires automated methods of extraction, and such methods tend not to be readily generalisable between the images obtained from different labs' equipment. Chapter 6 describes the key prerequisites to improve the accuracy and automation of the Oxford Perception Lab's cone identification algorithm, comparing its outputs to carefully controlled manual cone counts.

In summary, this thesis probes irregularities in the story of trichromatic colour vision: the possibility of separable neural channels that underlie specific nonlinearities of colour space, and the consequences and a possible mediator of the imbalanced numbers of L and M cones in the human retina. It additionally advances the development and assessment of processing methods for a technology, AO retinal imaging, that shows great promise for the study of colour vision in humans *in vivo*.

Declaration

The work contained in this thesis is entirely that of Sarah E Regan (SER), but for the exceptions described below.

Chapter 1: Several paragraphs of this chapter are adapted from a publication (Regan, Lee, MacLeod & Smithson, 2018) of which SER is the first author.

Chapter 2: This chapter is adapted from a publication (Regan et al., 2018) of which SER is the first author.

Chapter 3: 65% of the participants for this chapter were recruited by SER and 35% were recruited by another DPhil student, Allie C Hexley (ACH). 50% of the participants were run by SER; 30% by an undergraduate student, Rebecca Williams (RW), under SER's supervision; and 20% by ACH. SER built a prototype of the heterochromatic flicker photometry (HFP) apparatus and some control software. The final version of the HFP apparatus and control software, and the software for the conversion of HFP measurements to L:M ratio, were built and written by ACH.

Chapters 4 and 5: The adaptive optics scanning laser ophthalmoscope (AOSLO) and associated software were built by Young, Morris, Saunter and Smithson (2018). 50% of the montages were made by ACH, RW and Samuel Ponting.

Chapter 6: The retinal images were collected in collaboration with Mital Shah and Anna-Katharina Hauperich. The adaptive optics scanning laser ophthalmoscope (AOSLO) and associated software were built by Young et al. (2018). The automatic cone localisation algorithm was adapted from an algorithm written by Hannah E Smithson, modified from similar code by Mark D Fricker.

Acknowledgements

First I must thank my supervisor, Hannah Smithson: your time, advice and encouragement during these years have been generous and indispensable. Truly, you have gone above and beyond.

Thank you also to my second supervisor, Holly Bridge, and the Vision Group for your fresh perspectives, and to my Interdisciplinary Bioscience DTP tutor Esther Becker for your pastoral support.

Laura Young, thank you for everything from the AOSLO tutorials to the Task −1's. Thank you to the other Oxford Perception Lab members: Alix, Biswajit, Dan, Joshua, Manuel, Mira, Mital, Rebekah and Rob. I owe particular thanks to Takuma for the long talks, and to my work wives: Allie, my impassioned collaborator; and above all, Anna, I am so fortunate to have had you by my side from the very beginning of this bizarre journey.

I am beyond grateful to my dear friends of many years, Hannah and Elliot, for the enormous support you have given me; to Susanna and Laura for providing me with a safe and necessary haven during the COVID-19 storm; and to David, for all that and more.

My final thanks are to my brother, for being the same but different, and most especially to my parents, without whom this thesis would not have been possible – your unconditional and unlimited love and support is essential to me and to everything I do.

This work was supported by funding from the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/M011224/1], and the University of Oxford's Tinbergen closure and COVID-19 extension funds.

Contents

1 Introduction **19**
- 1.1 Colour vision . . . 19
  - 1.1.1 Cone photoreceptors . . . 19
  - 1.1.2 Cardinal mechanisms . . . 20
  - 1.1.3 Non-cardinal mechanisms . . . 21
  - 1.1.4 Interim summary . . . 23
- 1.2 Impact of relative numbers of long- and middle-wavelength sensitive cones . . . 23
  - 1.2.1 Long- to middle-wavelength cone ratio . . . 24
  - 1.2.2 Genetics of L:M cone ratio . . . 24
  - 1.2.3 Assessment of L:M cone ratios . . . 25
  - 1.2.4 Impact of L:M cone ratio . . . 26
  - 1.2.5 Interim summary . . . 28
- 1.3 Fixational eye movements . . . 29
  - 1.3.1 Function of fixational eye movements . . . 30
  - 1.3.2 Interim summary . . . 32
- 1.4 Adaptive optics . . . 33
  - 1.4.1 Cone reflectance . . . 34
  - 1.4.2 Cone typing . . . 35
  - 1.4.3 The Oxford AOSLO . . . 36
  - 1.4.4 Interim summary . . . 37
- 1.5 Thesis overview . . . 37

2 Are hue and saturation carried in separate neural channels? **39**
- 2.1 Introduction . . . 39
  - 2.1.1 Colour mechanisms . . . 40
  - 2.1.2 Related work . . . 41
  - 2.1.3 The present study . . . 42
- 2.2 Methods . . . 43
  - 2.2.1 Ethics . . . 43
  - 2.2.2 Apparatus . . . 43
  - 2.2.3 Stimuli . . . 43
  - 2.2.4 Task . . . 47
  - 2.2.5 Procedures . . . 48
  - 2.2.6 Observers . . . 51
- 2.3 Results . . . 51
  - 2.3.1 Experiment 1: Phase-shifted non-cardinal thresholds . . . 51
  - 2.3.2 Experiment 2: Adaptation . . . 53
- 2.4 Discussion . . . 54
  - 2.4.1 The super-importance of hue . . . 55

2.4.2 Separate neural channels for hue and saturation? . . . 56
2.4.3 Detection by cardinal mechanisms . . . 56
2.4.4 Correlated neural noise . . . 57
2.4.5 Conclusions . . . 57
2.5 Appendix: Visualisation captions . . . 58

3 Does L:M ratio affect foveal spatiochromatic resolution? 59
3.1 Introduction . . . 59
3.1.1 Genetics of L:M ratio . . . 60
3.1.2 Assessment of L:M ratios using flicker photometry . . . 62
3.1.3 Spatial resolution . . . 65
3.1.4 L:M ratio affects spatiochromatic resolution . . . 67
3.1.5 The present study . . . 68
3.2 Methods . . . 69
3.2.1 Ethics . . . 69
3.2.2 Observers . . . 69
3.2.3 Acuity and colour vision deficiency assessment . . . 69
3.2.4 Heterochromatic flicker photometry . . . 70
3.2.5 Square C optotype thresholds . . . 71
3.3 Data preparation . . . 75
3.3.1 Colour vision deficiency tests . . . 75
3.3.2 HFP score and L:M ratio . . . 75
3.3.3 Square C optotype threshold difference . . . 75
3.3.4 Observer exclusion . . . 76
3.4 Results . . . 76
3.4.1 Heterochromatic flicker photometry and estimated L:M ratios . . . 76
3.4.2 Incremental vs decremental C optotypes . . . 77
3.4.3 Spatiochromatic thresholds . . . 77
3.4.4 Spatiochromatic threshold differences . . . 78
3.4.5 Correlations . . . 78
3.4.6 Gender and family history of red-green colour vision deficiency . . . 79
3.4.7 Ethnicity . . . 81
3.5 Discussion . . . 82
3.5.1 Does L:M ratio affect foveal spatiochromatic vision? . . . 82
3.5.2 Are thresholds for the square C optotype test primarily determined by opsin peak rather than L:M ratio? . . . 83
3.5.3 Foveal and parafoveal ratios . . . 84
3.5.4 Correlation strength . . . 86
3.5.5 Gender and family history of red-green colour vision deficiency . . . 88
3.5.6 Ethnicity . . . 89
3.5.7 Conclusion . . . 90

4 Capturing retinal images and eye movements using the Oxford adaptive optics scanning laser ophthalmoscope 91
4.1 Introduction . . . 91
4.2 Image acquisition . . . 92
4.3 Eye motion estimation . . . 94
4.4 Reference montage generation . . . 95
4.5 Axes of eye position space . . . 96
4.6 Eye position quality metric . . . 101
4.6.1 Creation of an eye position quality gold standard . . . 101

4.6.2 Analysis of the gold standard . . . 102
4.6.3 When does the quality metric fail to capture the effectiveness of eye position extraction? . . . 103

5 Eye movements with L- and M-cone isolating stimuli near the acuity limit 105
5.1 Introduction . . . 105
5.1.1 Sampling- and optics-limited regimes . . . 105
5.1.2 Resampling . . . 106
5.1.3 L and M cones and resampling . . . 109
5.1.4 Fixational eye movements and colour . . . 110
5.1.5 Quantifying eye movement . . . 112
5.1.6 The present study . . . 113
5.2 Methods of data collection . . . 114
5.2.1 Ethics . . . 114
5.2.2 Observers . . . 114
5.2.3 Apparatus . . . 115
5.2.4 Trial stimuli . . . 116
5.2.5 Data collection . . . 116
5.3 Methods of analysis . . . 116
5.3.1 L:M ratio metric . . . 116
5.3.2 Eye movement metric . . . 117
5.3.3 Trial selection . . . 118
5.3.4 Linear regression versus Spearman's rank correlation . . . 119
5.3.5 Hypothesis 1: Regardless of scene content, individuals with different L:M ratios move their eyes differently from one another . . . 119
5.3.6 Hypothesis 2: Eye movements are tuned to scene content . . . 119
5.3.7 Mean centre-to-centre cone spacing . . . 120
5.3.8 Hypothesis 3: Each participant has an amount of RMSE eye position associated with peak performance . . . 120
5.4 Results . . . 122
5.4.1 Eye position data quality . . . 122
5.4.2 Descriptives about RMSE spread . . . 122
5.4.3 HFP to L:M ratio . . . 122
5.4.4 Hypothesis 1: Regardless of scene content, individuals with different L:M ratios move their eyes differently from one another . . . 123
5.4.5 Hypothesis 2: Eye movements are tuned to scene content . . . 125
5.4.6 Hypothesis 3: Each participant has an amount of RMSE eye position associated with peak performance . . . 127
5.5 Discussion . . . 128
5.5.1 Hypothesis 1: Regardless of scene content, individuals with different L:M ratios move their eyes differently from one another . . . 129
5.5.2 Hypothesis 2: Eye movements are tuned to scene content . . . 130
5.5.3 Hypothesis 3: Each participant has an amount of RMSE eye position associated with peak performance . . . 131
5.5.4 Challenges and future directions . . . 132
5.5.5 Conclusion . . . 135
5.6 Appendix . . . 137

6 Cone photoreceptor localisation in adaptive optics retinal images 139
6.1 Introduction . . . 139
6.1.1 Existing methods . . . 140
6.1.2 Method assessment . . . 141

6.1.3 The present study . . . 142
6.2 Methods . . . 143
6.2.1 Ethics . . . 143
6.2.2 Retinal image collection and selection . . . 143
6.2.3 Manual cone localisation . . . 144
6.2.4 Automatic cone localisation program . . . 151
6.3 Results . . . 153
6.3.1 Cone localisation gold standard . . . 153
6.3.2 Narrowing of algorithm parameters . . . 153
6.4 Discussion . . . 161
6.4.1 Manual localisation . . . 162
6.4.2 Automatic cone localisation . . . 162
6.4.3 Cone statistics . . . 165
6.4.4 Future directions . . . 165

7 General discussion **167**
7.1 Hue and saturation . . . 167
7.2 L:M ratio . . . 168
7.3 Eye movements and L:M ratio . . . 170
7.4 Cone localisation: The gold standard and an automated algorithm . . . 172
7.4.1 Gold standard for cone detection . . . 172
7.4.2 Automatic algorithm . . . 172
7.5 Conclusions . . . 173

List of Figures

2.1 Chromaticity of stimuli temporally varying in hue or saturation as defined in terms of the cardinal mechanisms . . . 44
2.2 Spatial characteristics of the adapting and test stimuli varying in hue and saturation 45
2.3 Detection thresholds for temporally varying stimuli, two in terms of hue and two in terms of saturation . . . 52
2.4 Detection thresholds for temporally varying stimuli, averaged across both saturation conditions and averaged across both hue conditions . . . 52
2.5 Post-adaptation detection thresholds for stimuli that temporally vary in terms of either hue or saturation, following adaptation to stimuli that vary either in terms of hue or saturation . . . 53
2.6 Post-adaptation detection thresholds for within-condition and between-condition adaptation . . . 54

3.1 Spatial configuration of the square C optotype . . . 72
3.2 Histograms of heterochromatic flicker photometry (HFP) scores and long- to middle-wavelength (L:M) cone ratio . . . 76
3.3 L−M log threshold differences for decremental versus incremental square C optotypes for size and contrast . . . 77
3.4 Examples of log thresholds for L cone-isolating and M-cone isolating square C optotypes for size and contrast . . . 78
3.5 Histograms for L−M log threshold differences for size and contrast thresholds for L cone-isolating and M-cone isolating square C optotypes . . . 79
3.6 Scatter plots showing the relationships between heterochromatic flicker photometry (HFP) score, long- to middle-wavelength (L:M) cone ratio estimates, and L−M log threshold differences . . . 80
3.7 Histograms for heterochromatic flicker photometry (HFP) score and L−M log contrast threshold differences between L cone-isolating and M-cone isolating square C optotypes for female and male participants . . . 81
3.8 Histograms for heterochromatic flicker photometry (HFP) score and L−M log contrast threshold differences between L cone-isolating and M-cone isolating square C optotypes for Asian/Asian British and White participants . . . 82
3.9 Illustrations of four possible relationships between linear middle- to long-wavelength (M:L) mean cone-to-cone spacing ratio and a given psychophysical measure . . . 85
3.10 Linear middle- to long-wavelength (M:L) threshold ratio for size and contrast plotted against estimated linear M:L mean foveal centre-to-centre cone spacing ratio . . . 86

4.1 Effect of fixation relative to the AOSLO scan patch on the area of retina imaged . . . 94
4.2 Spacing of images for the retinal image montage . . . 96
4.3 Set of central montage images for participant AAM laid out with partial transparency in GNU Image Manipulation Program (GIMP) . . . 97
4.4 Intended versus actual location of a retinal image montage . . . 98

4.5 Adjustment of origin made to trial eye position output coordinates for a given observer based on relative location of central montage image . . . 100
4.6 Adjustment of origin made to trial eye position output coordinates for a given observer based on relative locations of multiple montage images . . . 101
4.7 Histogram of the distribution of the frame-averaged eye position extraction quality metric for the frames manually marked as accepted and rejected . . . 103

5.1 Effect of resampling on resolution . . . 107
5.2 Effect of resampling on resolution for a simulated array of L and M cones . . . 111
5.3 Histogram of the proportion of frames remaining for each participant following the exclusion of any frame with a mean quality score of less than 0.079 . . . 122
5.4 Box plots of the distributions of the amounts of eye movement (RMSE eye position, arcmin) included in the final analyses . . . 123
5.5 Box plots of the distributions of the amounts of eye movement (RMSE eye position, arcmin) for L- and M-isolating trials . . . 124
5.6 Linear regressions of eye movement versus L:M ratio metrics . . . 125
5.7 Assumptions of linear regressions of eye movement versus L:M ratio metrics . . . 126
5.8 Linear regressions of the differences in amounts of eye movement for L- and M-isolating trials versus L:M ratio metrics, and underlying eye movement values . . . 127
5.9 Assumptions of linear regressions of the differences in amounts of eye movement for L- and M-isolating trials versus L:M ratio metrics . . . 128
5.10 Assumptions of linear regressions of eye movement for L- and M-isolating trials versus HFP score . . . 128
5.11 Linear regressions of eye movement for L- and M-isolating trials versus mean centre-to-centre class-specific cone spacing . . . 129
5.12 Accuracy versus RMSE eye position, single logistic regressions . . . 137
5.13 Accuracy versus absolute difference between RMSE and median RMSE eye position, single logistic regressions . . . 138

6.1 Location of image capture and use for a right eye . . . 144
6.2 Images of human retina used for manual and automatic cone localisation . . . 145
6.3 Manual cone localisation display . . . 146
6.4 AOSLO cone image at an eccentricity of 1.5 deg overlaid with the cone locations identified by the manual raters . . . 148
6.5 An AOSLO cone image at an eccentricity of 1.5 deg showing the comparison of a test and reference set of cone coordinates . . . 149
6.6 Processing pipeline for the automatic cone localisation algorithm . . . 150
6.7 Heatmaps showing the mean $F1$ scores across eyes for the laplacian of gaussians convolution algorithm applied to images of human retina with final cone positions unmerged . . . 154
6.8 Effect of oversampling factor on algorithm performance . . . 155
6.9 Effect of filter standard deviation on performance for gaussian and laplacian of gaussians filtering algorithms with cone merging for superior retina . . . 156
6.10 Effect of merging implausibly close cones on mean performance ($F1$ score) . . . 157
6.11 Effect of different filters on mean performance ($F1$ score) . . . 159
6.12 Optimum laplacian of gaussians filter standard deviation for different eccentricities and directions from the fovea . . . 160
6.13 Optimum laplacian of gaussians filter width and histological estimates of cone-to-cone spacing at different eccentricities . . . 161
6.14 Optimum laplacian of gaussians filter widths, measured in the present study and predicted from histological measurements of cone density . . . 161

6.15 AOSLO cone image taken at an eccentricity of 1.5 deg at (a) its original size, and (b) upsampled by a factor of two. The images have been processed using the no-filter algorithm which has correctly detected all of the cones in (b) and most of the cones in (a; green dots) but has missed one cone in (a; red cross). 164

List of Tables

3.1 M:L spatiochromatic linear threshold ratios for C-shaped optotypes. 84
3.2 L:M cone ratio and M:L mean cone-to-cone spacing ratio estimates derived from flicker photometry. 85

Chapter 1

Introduction

Declaration: The work contained in this chapter is entirely that of the author (SER). Several paragraphs of this chapter are adapted from a publication (Regan et al., 2018) of which SER is the first author.

1.1 Colour vision

1.1.1 Cone photoreceptors

When light enters the human eye, it passes through several layers of retinal cells before it reaches the photoreceptor layer where the cones, which are principally responsible for vision in the photopic range of light levels, are interleaved with the rods, responsible for scotopic vision. There are three classes of cone photoreceptor, maximally sensitive to either long-, middle- or short-wavelength (L, M and S) light. Microspectrophotometric measurements have shown that the probability of absorbing a photon of a given wavelength peaks for S cones at 419 nm, for M cones at 530 nm, and for L cones at 558 nm (Dartnall, Bowmaker & Mollon, 1983). However, prior to reaching the photoreceptors, light is attenuated at short wavelengths by the lens and macular pigment which shifts the effective peaks of the cone spectral sensitivities to longer wavelengths. The resulting cone sensitivity spectra, or 'cone fundamentals', have been estimated by Stockman, Sharpe and Fach (1999) and Stockman and Sharpe (2000) from colour matching data (Stiles & Burch, 1959) and from spectral sensitivity measurements made in non-trichromatic observers, for S cones by obtaining threshold measurements for a flickering test light over a range of wavelengths in S-cone monochromats, and for L and M cones by performing heterochromatic flicker photometry (HFP, described in § 1.2.3) in single-gene red-green dichromats.

Knowledge of the cone fundamentals allows us to define and present stimuli that will affect the cones in extremely precise ways. In displays such as cathode ray tube (CRT) or liquid crystal

display (LCD) monitors, the cone fundamentals can be combined with the emission spectra of the display primaries to determine the combination of primary intensities that will give rise to particular cone excitations. This permits, for example, the use of silent substitution to achieve cone isolation: variation of the stimulation of one cone while holding that of the others constant. The term silent substitution is used because the technique involves a substitution of one set of light intensities for another that would be unnoticed by all but one class of photoreceptor. These techniques allow us to study mechanisms: individual processes in the visual system that may or may not be linked with a particular class of cells. For example, combining cone fundamentals with display primaries also allows for carefully controlled adaptation, in which a field is presented in a sustained way until the responses of a mechanism or mechanisms are attenuated, either to demonstrate the existence of the mechanism by the ability to depress its responses or to expose those of another mechanism. Mechanism adaptation and cone isolation are deployed in Chapters 2, 3 and 5 with stimuli displayed on CRT and LCD monitors.

1.1.2 Cardinal mechanisms

Signals from the photoreceptors are passed via horizontal, bipolar and amacrine cells to the retinal ganglion cells (RGCs), the retina's projection neurons. Psychophysical experiments have been used to quantify the number and nature of post-receptoral mechanisms, additionally guided by physiological and anatomical knowledge of the underlying biological substrate. A psychophysical adaptation experiment in humans by Krauskopf, Williams and Heeley (1982) measured thresholds for light pulses following habituation to a wide range of fields produced by temporally modulating red, green and blue LEDs combined by a colour mixer. They found highly selective losses of chromatic or luminance sensitivity to specific types of modulation that suggested the existence of: (1) a mechanism opposing S signals with L+M; (2) a mechanism opposing L and M signals; and (3) a luminance mechanism. Other types of modulation produced broader sensitivity losses.

Physiological counterparts to these three psychophysical mechanisms were found by Derrington et al. (1984). They recorded from macaque neurons in the lateral geniculate nucleus (LGN), the stage in the image-forming visual pathway between RGCs and the primary visual cortex, while presenting stimuli motivated by specific combinations of cone signals on a CRT monitor. They defined a three-dimensional colour space (known as DKL space) in terms of the null planes passing through the white point of three key different cell types they found: one which did not respond to S−(L+M) or L+M+S modulation provided that L−M modulation was held constant; one which did not respond to variation in L−M or L+M+S signals where S−(L+M) was held constant; and one

which did not respond to L−M or S−(L+M) modulation where the L+M+S signal was held constant. This space extended MacLeod-Boynton space, whose two L/(L+M) and S/(L+M) axes represent the two chromatic cardinal mechanisms on an equiluminant plane (MacLeod & Boynton, 1979), with the addition of a luminance axis. The three classes of cell that were particularly strongly simulated by targets that modulated these three cardinal mechanisms are now believed to inherit their tuning from certain retinal ganglion cells, as discussed below.

There are at least 20 different types of primate retinal ganglion cells (Dacey, Peterson, Robinson & Gamlin, 2003), three of which are particularly interesting in terms of luminance and colour. Parasol cells draw inputs from L and M cells in a centre-surround receptive field structure in which the two fields generally have the same spectral sensitivity, meaning they are particularly affected by luminance contrast. They project to the magnocellular (M) layers of the LGN (De Monasterio & Gouras, 1975). The midget ganglion cells, which project to the parvocellular (P) LGN layers, have a centre-surround receptive field structure and draw inputs nonselectively from L and M cells (Martin et al. 2001, Field et al. 2010, Wool et al. 2018). This means that some midget ganglion cells can be affected by just luminance contrast, when the surround inputs happen to be drawn from only the same cone classes as the centre, and others may be affected by colour and luminance contrast, when the surround contains some inputs from a different class to the centre. The small bistratified ganglion cells, projecting to the koniocellular (K) LGN layers, oppose S cone signals with those of L and M cones, again providing a colour contrast signal, in a receptive field structure that has variously been reported to be spatially coextensive (Crook et al., 2009; Dacey & Lee, 1994) or centre-surround (Field et al., 2007).

1.1.3 Non-cardinal mechanisms

Since the establishment of the cardinal mechanisms (luminance, L vs M and S vs [L+M]), and their physiological counterparts, a number of non-cardinal mechanisms have been postulated. Krauskopf, Williams, Mandler and Brown (1986) reanalysed the data from their paper of 1982 and discovered that, in many cases, the maximum habituation of a test angle was achieved by adapting to a stimulus varying along that angle. The habituations were not as clear and consistent as those of the cardinal mechanisms, leading Eskew (2009) to describe these non-cardinal effects as 'very much second-order'. The psychophysical non-cardinal effects are echoed in neurophysiological measurements showing that the chromatically opponent cells found in V1 do not form populations sensitive to the two chromatic cardinal axes: instead, cells are maximally sensitive to modulation along intermediate directions, and no orientations predominate (Lennie, Krauskopf & Sclar, 1990).

The psychophysical non-cardinal effects and physiological recordings share other features (they both pass radially through the white point and are mostly linear), but lack the distinctive groupings into a small number of mechanisms shown by those mechanisms linked to processes earlier in the visual stream.

Other researchers have suggested a more limited number of non-cardinal mechanisms. Before the advent of the cardinal mechanism theory, Hering (1878) argued for two bipolar chromatic channels: a red-green and a yellow-blue opponent channel. These arose from the observations that the after-effect of adaptation to a red stimulus is green, and vice versa, while the after-effect of adaptation to blue is yellow, and vice versa. The hues (the subjective quality of appearing, for example, blue, bluish-green or green) at the poles of these channels were described as 'unique'. These four unique hues were phenomenologically unmixed, in that they appeared to contain no other colour. However, it is now agreed that the poles of the cardinal mechanisms do not correspond to the unique hues. This disparity has led some, for example De Valois and De Valois (1993), to postulate explicitly post-cardinal physiological processes, recombining the cardinal signals to yield six unipolar channels: intensity increments (L+M+S), intensity decrements ([-L]+[-M]+[-S]), red (L+[-M]+S), yellow (L+[-M]+[-S]), green ([-L]+M+[-S]) and blue ([-L]+M+S).

Taking a nonlinear version of the above cone combinations as his example, Eskew (2009) has pointed out that many of the psychophysical tasks that have been taken to support a large number of linear non-cardinal mechanisms, including habituation and noise masking experiments, could also be explained by a smaller number of nonlinear mechanisms. Other aspects of the visual system, such as cone adaptation state, demand a nonlinear explanation.

Hue and saturation

Another signature of the nonlinearity of the visual system relates to hue and saturation. As described in § 1.1.3, hue is the subjective quality of appearing, for example, blue, bluish-green or green. Saturation is the subjective counterpart of colorimetric purity, where a saturated colour can be desaturated to a pastel shade by adding white. In general, two colours may be different in hue if they are drawn from different angles around the white point in a colour space, and colours that are drawn from the same angle around but are different distances from the white point may appear different in saturation.

Nickerson (1936) measured observers' abilities to discriminate Munsell chips along dimensions equivalent to hue, saturation and brightness and created a colour formula that described the change needed to generate a step of just-noticeable difference along any dimension from any

chromaticity. In a Euclidean space based on this formula, the number of just-noticeable steps around the perimeter of a circle centred on the completely desaturated white point ought to be $2\pi r$, where r is the number of just-noticeable steps along the radius. In fact, Judd calculated it to be slightly more than $4\pi r$ (Judd, 1968). This means that a smaller physical stimulus change is required to detect a change in hue than to detect a change in saturation, which he called 'the super-importance of hue'. This particular example of the nonlinearity of colour models has been taken to suggest the existence of hue channels: mechanisms encoding colour differences that are tangential to circles centred on the white point (Kuehni, 2003), separate from channels representing saturation. The MacLeod-Boynton and DKL colour spaces are therefore unlikely to be appropriate spaces within which to represent putative 'hue' or 'saturation' mechanisms or the stimuli that isolate them.

1.1.4 Interim summary

In summary, higher-order mechanisms have most commonly been conceived as mechanisms that are tuned to a narrow range of hue-angles and that respond to modulation through or from the white-point (i.e. to changes in saturation at a fixed hue angle). Some of these mechanisms have possible biological correlates. However, there have been fewer discussions of hue mechanisms, which do not pass through the white point, and no proposed biological correlate. Meanwhile, biologically-motivated colour models using the cardinal mechanisms fail to predict and represent postcardinal processes. Until we have a better understanding of colour-appearance spaces, our ability to visualize the tuning of these mechanisms is limited. The experiment presented in Chapter 4 therefore investigates the neural encoding of colour with specific reference to the nonlinearities implied by hue super-importance.

1.2 Impact of relative numbers of long- and middle-wavelength sensitive cones

The three classes of cones that underpin human colour vision are neither equal in number nor uniformly distributed across the retina. Furthermore, there are striking differences in the relative numbers of L and M cones (L:M cone ratio) between individuals (Carroll, McMahon, Neitz and Neitz 2000; Carroll, Neitz and Neitz 2002; Hofer et al. 2005). This should cause differences in spatial resolution for L-isolating and M-isolating stimuli both within and between observers, yet to date this has only been demonstrated in a very limited way (Danilova et al., 2013). It has also proved difficult to find any impact on subjective judgements of unique hues (Brainard et al., 2000; Miyahara

et al., 1998; Schmidt, Touch, Neitz & Neitz, 2015), and detection thresholds for chromatic targets are only elevated in those believed to have the most imbalanced of ratios (Hood, Mollon, Purves & Jordan, 2006). The experiment presented in Chapter 3 expands our knowledge of the impact of the relative numbers of L and M cones on foveal spatiochromatic resolution, while Chapter 5 explores a possible mechanism for the limited extent of the effect of the relative numbers of cones.

1.2.1 Long- to middle-wavelength cone ratio

On average, there are 2-2.5 times as many L cones as M cones (Carroll et al., 2000, revised upwards by Hofer et al., 2005), but the L:M cone ratio varies between individuals at least from 0.37:1 to 16.5:1 (Carroll et al., 2002; Hofer et al., 2005). S cones are scarce, numbering 5% of cones (Curcio et al., 1991), and are completely absent from the centre of the fovea (Castańo & Sperling, 1982). Furthermore, as would be expected in any truly randomly distributed retina, cone classes show 'clumping', in which a local area has fewer examples of one class than is indicated by the overall ratio (Roorda & Williams, 1999). An L cone in a truly random retina skewed strongly towards a high L:M cone ratio will have many more L neighbours than M neighbours; even minor clumping in a given area will exacerbate this existing tendency to have several L cones adjacent to one another.

1.2.2 Genetics of L:M cone ratio

A genetic basis for the large range in L:M cone ratios has yet to be found. The L and M photopigments are encoded by genes on the X chromosome, among which are a single copy of the L pigment gene, OPN1LW, and one or more copies of the M pigment gene, OPN1MW (Nathans, Piantanida, Eddy, Shows & Hogness, 1986; Nathans, Thomas, Hogness, Shows & Hogness, 1986). However, the ratio of expressed L and M pigment mRNA is not correlated with the number of copies of OPN1MW (Yamaguchi, Motulsky & Deeb, 1997), and while some other genetic factors have been linked to cone ratio (Deeb, Dorschner, Shafer, Kutyavin & Stamatoyannopolous, 2004; Gunther, Neitz & Neitz, 2008; Lawrance-Owen et al., 2014), their influence falls far short of explaining the extent of the variation.

Females, who have two X chromosomes, have an additional source of variation to males: random X inactivation takes place prior to pigment development, leading to half of a female's L and M cones being guided by the L and M gene array of one parent, and half by that of the other parent. Jordan and Mollon (1997) have shown that an individual's L:M cone ratio tends towards a combination of the ratios of the two parents by investigating carriers of red-green colour vision deficiency (CVD), for whom half of their cones would express the L and M genes of one

parent's X chromosome, and half of the cones would express either just one L or M gene from the X chromosome of the other parent with CVD, or two genes with very similar peak wavelengths. If the carrier's resulting L:M cone ratio is the mean of that of the two parents, carriers of a deutan deficiency would have an estimated mean ratio of 5:1, while the mean across protan carriers would be 1:2 (Hood et al., 2006), pushing carriers' ratios into the tails of the distribution of L:M cone ratios. Carriers of red-green CVD therefore provide a useful opportunity for recruitment of participants with extreme cone ratios for the investigation of the perceptual effects of L:M cone ratio variation (Gunther & Dobkins, 2002; Jordan & Mollon, 1997).

1.2.3 Assessment of L:M cone ratios

Two L:M cone ratio estimation techniques that directly test the cones in human observers are (1) microspectrophotometry, in which beams of narrowband light of various wavelengths are passed through the photoreceptor pigments to determine the absorbance spectrum of each cell; and (2) the quantification of mRNA (a substance involved in the synthesis of genes into proteins) in patches of retina containing multiple cells, here to determine the relative degrees to which the L or M gene is actually expressed in the tissue. A microspectrophotometric study found a mean L:M cone ratio across four individuals of 2.66:1 (Bowmaker, Parry & Mollon, 2003), while a mRNA quantification study calculated a mean ratio of 1.50:1 in the central retinas of participants aged 20-60 years (Hagstrom, Neitz & Neitz, 1998). However, such techniques require the eyes to have been removed from the observer, and are expensive and time-consuming. *In vivo* methods are to be preferred for testing large numbers of participants and combining L:M cone ratio estimation with other measures of colour vision.

The classic non-invasive, behavioural measure of L:M cone ratios in human individuals is flicker photometry. Two lights are superimposed and flickered in counterphase such that, if the superimposed lights do not have the same luminance, the stimulus appears to be a field of steady colour but flickering intensity. The intensity of one or both lights may be adjusted such that the luminance flicker is reduced to a null or minimum, and an estimate of L:M cone ratio can be derived from knowledge of the light spectra, the cone fundamentals, and the model of luminance as an additive combination of L and M cone signals. It can be combined with other techniques such as electroretinography (ERG) and can be improved by adjustments to an individual's cone fundamental peaks using the results of genetic testing. Flicker photometric ERG has been validated against the highest standard of *in vivo* L:M cone measurement, the direct measurements obtained via densitometry in cone-typed retinal images (Brainard et al., 2000; Hofer et al., 2005) described in

§ 1.4.2, and in turn estimates obtained from heterochromatic flicker photometry have been found to broadly agree with flicker photometric ERG (Kremers et al., 2000; Kremers, Usui, Scholl & Sharpe, 1999).

Deriving estimates of L:M cone ratio from flicker photometry has been strongly challenged by Stockman, Jägle, Pirzer and Sharpe (2008), who used the technique to measure the relative impact of a range of wavelengths on apparent luminance. First, they found that the background against which the flickering stimuli are presented has a large effect on the M/L cone excitation ratio. The luminosity function they presented, $V_{D65}^*(\lambda)$, had an L cone weighting of 1.89 compared to the M cones, close to the 2:1 estimates of L:M cone ratio commonly obtained from flicker photometry, but ultimately a special case obtainable only against a background of the CIE standard illuminant D_{65}. Second, they argued that the use of flicker photometry without any background in extremely influential studies such as Brainard et al. (2000) and Hofer et al. (2005) was even more inappropriate due to the higher overall adapting chromaticity changes with test wavelength; however, the force of this argument is weakened by the fact that the flicker photometric ERG in these two studies agreed so well with their direct L:M cone ratio measurements in cone-typed retinal images.

1.2.4 Impact of L:M cone ratio

Subjective colour experience

In spite of these individual differences, it has been difficult to discover any convincing inter-individual variance in subjective chromatic judgements such as hue naming and unique hue settings (Brainard et al., 2000; Miyahara et al., 1998; Schmidt et al., 2015). Miyahara et al. (1998) explored the colour vision of BB, a colour-normal, confirmed protan heterozygote with an L:M cone ratio estimated from HFP matches of 0.09:1 (or 1:11.11), and seven normal trichromats. They carried out hue estimation, in which observers responded to repeated presentations of a range of wavelengths by reporting whether the sensation was best described by 'blue', 'green', 'yellow' or 'red', and measured equilibrium (unique) yellow using a staircase procedure to identify the wavelength of a hue that was neither reddish nor greenish. In both cases, the resulting data showed that BB's responses were consistent with those of normal trichromats. One possible explanation for the lack of impact of L:M cone ratio is that the effects of L and M signals on later cells such as the midget ganglion cells are brought to an approximate equilibrium by gain control (Kremers et al., 2000).

Further work has been done on L:M cone ratios and unique yellow, which might be expected to be a prime candidate for showing a difference in the relative weighting of L and M inputs to colour appearance. Brainard et al. (2000) established the L:M cone ratios of two observers by high-

resolution retinal densitometry, and by flicker electroretinogram (ERG). The two methods returned similar ratios (1.15:1 and 3.79:1; and 1.06:1 and 3.38:1), both of which indicated a inter-observer difference of at least threefold. The observers adjusted a flashing test light such that it appeared neither reddish nor greenish; the resulting unique yellow wavelengths were very similar, which was not in line with the predictions modelled using the observers' L:M cone ratios. Schmidt et al. (2015) used genetic sequencing of exons from L and M genes together with ERG to estimate the L:M cone ratios of 14 normal trichromats, and compared the results with measurements of unique yellow and green. In line with other data, unique yellow was not significantly correlated with L:M cone ratio.

Strikingly, L:M cone ratio did significantly, moderately correlate with unique green. However, it is unfortunate that they do not appear to have controlled the luminance of the stimuli (either in this publication or in their earlier study on unique yellow; Neitz, Carroll, Yamauchi, Neitz & Williams, 2002), which is known to alter the hue of a given chromaticity (Purdy, 1931).

Elevated perceptual thresholds in carriers of CVD

Several studies have investigated the consequences to colour perception of carrying a gene for red-green CVD, and by extension having an L:M cone ratio near the tails of the distribution. In addition to estimating cone ratio, Gunther and Dobkins (2002) obtained detection thresholds for briefly presented, low-frequency red-green and luminance gratings in the parafovea. Two-factor analysis showed that L:M cone ratio asymmetry and red-green contrast thresholds both loaded onto the same factor, the first negatively and the second positively: that is, L:M asymmetry was associated with poorer chromatic sensitivity. Likewise, Hood et al. (2006) found that discrimination of red-green gratings was attenuated in deutan carriers compared to controls, but not for protan carriers compared to controls. They attributed this to the likely L:M cone ratio of the carriers: the deutan carriers, who may have an average L:M cone ratio of 5:1, have a particular lack of M cones. On the other hand, the protan carriers, with a probable average L:M cone ratio of 1:2, have a more even ratio that is similar to the normal average of 2:1. It therefore seems to be the case that having a highly uneven cone ratio leads to a reduction in colour discrimination, even in conditions that should provide the maximum opportunity for the minority cone class to sample the stimulus – that is, when a spatially extensive grating is used.

Spatial resolution

The differing packing densities of the L and M cone classes strongly suggests that spatial resolution should be different for stimuli that isolate the L and M cones. Here, too, it has been challenging to

find any evidence that this is the case. In tests that primarily used red and green gratings, Green (1968) and Cavonius and Estévez (1975) failed to uncover any difference between red and green grating contrast which would be expected from the gratings' different impacts on the L and M cones. Only Miyahara et al. (1998) succeeded in this in a single observer, a colour-normal protan heterozygote with an extremely asymmetrical estimated L:M cone ratio of 0.03:1. This may be explained by the modelling of Garrigan et al. (2010) which may indicate that optical blur renders redundant the addition of further L or M cones to a mosaic, meaning that the human retina can tolerate variation in L:M cone ratio with little impact.

Alternative explanations for the lack of effect of L:M cone ratio on spatial resolution can be made with reference to the fact that all of the stimuli in the above tasks were gratings. First, neural integration along the bars could mitigate the effect of L:M cone ratio. When viewing gratings that isolate the minority cone class, there may be no or few pairs of adjacent or nearby cones from the minority class that could facilitate direct comparison across transitions from high to low amplitude. However, if signals from these cones are integrated along bars, a difference may be apparent between the signals of bars oriented in parallel to the grating, while there would be no signal difference between neurally integrated bars of other orientations. Danilova et al. (2013) have additionally highlighted the effect of the clumping of cones of the same class in the sampling array, which will occur in any truly randomly distributed cone mosaic. Such clumping can render the resolution of a cone class the same as the overall sampling resolution across classes (Otake, Gowdy & Cicerone, 2000) without the need for neural integration. To reduce the exploitation of clumping or neural integration in a spatiochromatic resolution task, the spatial extent of the target should be small. Additionally, any effect of spatial resolution may be seen most strongly in the parafovea, because optical factors limit resolution in the fovea (Thibos, 1998) and cone-to-ganglion cell signal convergence limits resolution in the periphery (Curcio & Allen, 1990). Danilova et al. (2013) hence presented Landolt C optotypes to the parafovea at an eccentricity of 5 deg. With this paradigm, they were able to find correlations between OSCAR test settings, a proxy for L:M cone ratio, and both size and contrast thresholds for L-isolating and M-isolating targets. They were therefore the first to find a clear link between L:M cone ratio and spatiochromatic resolution.

1.2.5 Interim summary

To conclude, the cause of the extraordinary range in L:M cone ratios in the human population remains unclear, although some can be attributed to heterozygosity for red-green colour deficiency. Ratio estimates from various techniques tend to be close to 2:1, although the random distribution

of the cone mosaic means that some retinal areas will see local exacerbation of L:M asymmetry due to clumping of cones of the same class. Variation in ratio seems to have little to no impact on subjective colour tasks such as hue estimation and unique yellow settings, although particularly asymmetric ratios are associated with elevation of red-green chromatic contrast thresholds. Only recently has the difficulty in discovering an effect of ratio on spatial resolution been explained by the historic use of spatially extensive gratings; there is indeed a relationship with spatial resolution for L- and M-isolating targets if they probe small areas of retina in the parafovea. The limited overall impact of L:M cone ratios on this range of colour tasks has yet to be explained, and the question of whether optical factors prevent foveal spatiochromatic resolution from being impacted by cone ratio remains unanswered.

1.3 Fixational eye movements

Vision is an active process, with eye movements constantly spatially translating the image across the photoreceptors. These fixational eye movements (FEM), usually too small to be consciously noticed, consist of microsaccades, drift and tremor. FEM are known to be useful in vision: microsaccades reduce fading during fixation (Martinez-Conde, Macknik, Troncoso & Dyar, 2006), and fine spatial processing is better when FEM are allowed to occur compared to when the image is artificially stabilised (Rucci, Iovin, Poletti & Santini, 2007). Furthermore, Intoy and Rucci (2020) have presented data indicating that FEM are tuned to the task at hand, positioning the stimulus on the retina and adjusting the drift speed and curvature to increase the temporal power at the relevant spatial frequency. This resampling of the stimulus by the retina, and the visual performance benefits thereof, have interesting implications for the different sampling densities of the cone classes, particularly in the case of the L and M cones as their relative numbers vary between as well as within individuals. Chapter 5 presents a set of formal tests of the relationship between L:M cone ratio, FEM, and performance on a high-acuity chromatic task.

The reported metrics of the various types of FEM have all varied widely (Martinez-Conde, Macknik & Hubel, 2004), partly due to differences in measurement technique and partly the frequent use of highly-trained observers who are able to keep their eyes stiller than typical observers. Microsaccades are fast, flick-like motions that occur between periods of drift. Unlike saccades, they are typically reported to be involuntary (Horwitz & Albright, 2003). They have been reported to be as small as one arcmin (Ditchburn & Ginsborg, 1953), the width of two foveal cones, and as large as 65 arcmin (Møller, Laursen, Tygesen & Sjølie, 2002), with many mean and median values below 10 arcmin (see Table 3, Martinez-Conde et al., 2004). Microsaccades puncture drift, a slow,

meandering motion. When measured from the start until the end of the intersaccadic interval, the median drift amplitude for an experienced observer can be as low as 1.5 arcmin (West & Boyce, 1968); investigating a larger number of participants, (Schulz, 1984) observed that the largest drifts were slightly more than 8 arcmin. Tremor is the small, high-frequency motion superimposed on drift, falling in the frequency band 50-100 Hz and with an average amplitude ranging from six arcsec (Eizenman, Hallett & Frecker, 1985) to one arcmin (Ratliff & Riggs, 1950).

1.3.1 Function of fixational eye movements

Eye movements in general have been found to support vision. Rucci et al. (2007) showed that performance on a high-spatial frequency grating discrimination task was hampered by stabilising the stimulus on the retina. However, it has been difficult to separate the relative effects of different types of fixational eye movement (Rolfs, 2009). Historically, microsaccades rates were found to be lower during high-acuity tasks relative to the rates seen in fixation (Winterson & Collewijn, 1976), and during the microsaccade itself vision is suppressed (Ditchburn, 1955); they were therefore deemed potentially not necessary or even disruptive when such tasks were being performed. However, more recently it has been demonstrated that although microsaccades are indeed suppressed during free viewing in a needle-threading task relative to fixation on a marker, they are elevated relative to free viewing of a natural scene (Ko, Poletti & Rucci, 2010). The researchers found evidence that the microsaccades did in fact have a functional role by classifying the microsaccade landing points as the needle, thread or background. Fixation landed on the needle or thread significantly more often than on the background, and also significantly more often on the thread – the object the participant was moving – than on the needle. This indicated that the microsaccades tended to move vision onto the relevant parts of the image (needle and thread) more often than the irrelevant parts (background), and onto the moving object (thread) more often than the static object (needle).

Intoy and Rucci (2020) further investigated the behaviour of microsaccades and drift during a high-acuity task by presenting participants with a Snellen chart of tumbling Es and requiring the observer to report the orientation of each optotype. Comparing task behaviour to sustained fixation, they found drift during the Snellen task to be slower and more curved. They further analysed drift in terms of effective temporal power by taking the power versus temporal frequency spectrum of all FEM that occurred during a given condition and weighting it by the temporal contrast sensitivity function of the parvocellular (midget) ganglion cells. From this, they then determined the temporal power for a range of spatial frequencies from 5 to 60 cycles per degree (cpd) for each condition, resulting in two temporal power versus spatial frequency curves, one each for the FEM seen during

the Snellen task and during sustained fixation. The peak of the temporal power versus spatial frequency curve during the Snellen condition shifted closer to the spatial frequency of the optotype compared to sustained fixation. Similarly, the amplitudes of microsaccades were lower during the task compared to sustained fixation – close, in fact, to the centre-to-centre spacing of the optotypes in the Snellen chart – and they were more likely to be directed to the right, as would be useful in the left-to-right reading required by Snellen task. This signifies that microsaccades might be optimised for the distance required to shift a different, nearby target onto the preferred retinal locus (PRL), the location of the retina onto which observers typically bring targets to bear. As Ko et al. (2010) found for microsaccades, both drift and microsaccades here seem to be tuned to the requirements of the high-acuity task. Although retinal stabilisation reduced acuity relative to non-stabilised performance, indicating the functional importance of FEM, their eye movement manipulations did not attempt to separate the contributions of microsaccades and drift.

Meanwhile, when viewing a Troxler fading stimulus, microsaccade probability, rate, and magnitude decreased prior to fading, and vice versa prior to increases in visibility (Martinez-Conde et al., 2006). The same held true when a similar experiment was carried out with the stimulus falling only on the fovea (Costela, McCamy, Macknik, Otero-Millan & Martinez-Conde, 2013). This work suggests that, should drift and tremor functionally contribute to vision, they alone do not suffice to prevent fading. When fading prevention and recovery are disentangled, it has been found that both microsaccades and drift contribute to each of these processes, but microsaccades are more important for recovery from fading, while greater drift speed and variation in speed increase drift's efficacy in fade prevention (McCamy, Macknik & Martinez-Conde, 2014).

Tremor has a particularly low amplitude of eye rotation and as such is more vulnerable to measurement error than microsaccades and drift. This makes tremor difficult to separate from the other types of FEM, and therefore more difficult to study in isolation (and in fact, in the study of function it is sometimes simply rolled in with drift; Rucci & Poletti, 2015). Recent advances in eye tracking, described in § 1.3.2, have allowed for measurements of eye position to be made with extremely high spatial and temporal resolution in less uncomfortable and unnatural viewing conditions. A study exploiting this has yielded an estimate of tremor size at the low extreme of previously reported values (Bowers, Boehm & Roorda, 2019), which may be due to smaller movements being revealed by the lower measurement error in higher-resolution technology. The standard deviation of motion in the 50 to 100-Hz tremor band of the eye movement amplitude spectrum was around 5 arcsec, just one-sixth the size of a foveal cone (Martinez-Conde et al., 2004), and therefore less likely to be able to play a functional role in vision.

1.3.2 Interim summary

Taken together, the studies described here indicate a role for FEM both in preventing and recovering from fading during sustained fixation, and in performance on high-acuity discrimination tasks. They also point to the tailoring of such movements to the requirements of the task. Microsaccades and drift both help to prevent stimulus fading (McCamy et al., 2014). Microsaccades have particular importance in recovery from fading (McCamy et al., 2014) and in the precise positioning of the stimulus close to the PRL in high-acuity discrimination (although the relative contributions of microsaccades and drift to discrimination performance have not been disentangled; Intoy & Rucci, 2020). Drift is less important for fade recovery, but the Intoy and Rucci (2020) study has pointed to a role for drift in the undertaking of the Snellen test, via the increase in the magnitude of intensity modulations at task-relevant spatial frequencies. FEM in general may also have a role in assisting the resampling of a single location by multiple cells.

There is evidence that FEM improve visual perception by allowing or causing a slightly different part of retinal mosaic to resample the visual scene (Park, Park & Kang, 2003; Patrick, Roach & McGraw, 2017; see also § 5.1). Given this, resampling may be particularly useful to colour vision: the interleaved nature of the cone classes means that if a small stimulus falls on a patch of the retina without M cones, a tiny movement may shift the stimulus to a patch that does contain M cones, along with the appropriate post-receptoral machinery to extract the spatiochromatic signal. FEM may improve chromatic performance by increasing the number of cone classes that can sample a single point of an image.

Any FEM may only vary by chance between individuals and tasks, but they may also be increased in individuals who would benefit most (e.g. in those with a strongly skewed L:M cone ratio), and during tasks that would benefit most (e.g. for tasks that isolate a minority class of cells). Therefore, Chapter 3 presents data demonstrating that L:M cone ratio is functionally relevant at the fovea, while Chapter 5 investigates the hypothesis that visual performance driven by the comparison of signals from the L and M cones is aided by FEM, which may be able to overcome the variation in retinal distribution of cone cells, both within and between individuals.

Eye tracking

The study of FEM requires eye tracking equipment that returns eye positions with good spatial and temporal resolution while allowing the participant to view and respond to stimuli comfortably, but techniques used to date have generally required significant trade-offs between these aims. Some researchers measured reflections from mirrors, originally using a plane mirror attached to

the eye itself (Adler & Fliegelman, 1934) and later affixing the mirror with contact lenses (Riggs, Armington & Ratliff, 1954), resulting in high resolution but uncomfortable and unnatural viewing conditions. Boundaries between different parts of the eye have been non-invasively tracked by taking the difference of the reflection intensities of the iris and sclera (Zuber, Stark & Cook, 1965), which can only return low-resolution estimates of horizontal eye position, and by measuring the differential reflection of the pupil and the cornea using a camera such as the EyeLink (SR Research Systems), capable of measuring eye position in both the horizontal and vertical axes but limited in accuracy by factors such as variation in pupil size. The Dual Purkinje Image (DPI) eye tracker, on the other hand, measures the relative positions of the Purkinje reflections from the anterior cornea and posterior lens, and is both a non-invasive and a high-resolution method. Yet none of the above methods are capable of measuring the shift of the image on the retinal mosaic rather than the position of the eye. This requires a much more recent methodological addition to the FEM researcher's toolkit: the adaptive optics scanning laser ophthalmoscope (AOSLO).

1.4 Adaptive optics

The scanning laser ophthalmoscope (SLO) was first developed in 1980 (Webb & Hughes, 1981; Webb, Hughes & Pomerantzeff, 1980), providing higher-contrast retinal images compared to other available imaging methods. It does so with a confocal microscope design, illuminating one point of the (x, y) plane of the sample at a time and using a pinhole to block scattered light from out-of-focus layers along the z axis that would reduce the image contrast. It uses a pair of scanning mirrors to raster-scan a laser beam over the *fundus oculi* and collects all the reflected light within a given temporal window, which is then used to build up a whole image frame.

One limitation to all *in vivo* retinal imaging is the passage of light through the eye's optics. The eye experiences a range of aberrations, including defocus and higher-order spherical aberrations such as coma (Liang & Williams, 1997). These aberrations can be measured and compensated for by adaptive optics (AO). Originally developed to allow ground-based telescopes to obtain clear images of space through Earth's atmosphere, it was notably incorporated with an SLO by Roorda et al. (2002). It requires a wavefront sensor to measure the local angles of reflection within a beam of light returned by the retina. This can be compared with the expected angle had the eye not introduced aberrations and then compensated for using a deformable mirror which corrects the angle of the returned light before it is imaged.

1.4.1 Cone reflectance

A long-term aim of the Oxford Perception Lab is to unite the study of individual differences in L:M cone ratios with the ability to relate performance on a chromatic task to how the stimulus has moved over different classes of cone during that task. Cone classification (see § 1.4.2), the assessment of whether a cone in a retinal image is long-, middle- or short-wavelength sensitive, first requires accurate cone localisation, the determination of the location of each cone in an AO retinal image. In turn, cone localisation requires a model – either on the part of a human or an algorithm – of a cone's appearance in these images. That cones look like light gaussian blobs on a grey background has not been disputed, yet even in images of healthy retinas, reflectance can vary substantially from cone to cone even when fully-bleached, and in the same cone imaged under the same conditions at different time points.

Pallikaris, Williams and Hofer (2003) used flood illumination AO ophthalmoscopy to measure cone reflectance changes every hour for 24 hours, and as often as every 30 seconds for three hours. They observed some large reflectance changes of as much as a factor of 3.5, including one change by a factor of 2.4 in just 13 minutes. Although cone autocorrelation was slightly negative after approximately 12 hours, which might suggest the influence of circadian rhythms, this is disputed by the fact that autocorrelation was close to zero after 24 hours. Correlation between pixels as a function of pixel separation dropped to close to zero after the average cone spacing, indicating that cone intensity changes are independent of one another. Furthermore, they imaged cones from multiple angles to measure their directional sensitivity, and found little change in directionality after 24 hours, but large changes in reflectance over this time, indicating that directionality does not drive changes in cone reflectance in healthy eyes.

Bruce et al. (2015) used AOSLOs to investigate the relationship between persistent low reflectance on perceptual sensitivity. They selected 10 cone-sized dark spaces surrounded by normal cone mosaics in five subjects and monitored their reflectance longitudinally. The effects of eccentricity and amount of light delivered by the imaging system were controlled for by pairing the dark patches with nearby cones and studying dark cone reflectance relative to the mean intensity of the cone pair. Three of the dark cones remained dark throughout the testing period, while seven were occasionally visible. To measure functionality, increment thresholds for small spots of light were obtained for the ten cone pairs during sessions when the dark cone appeared dark. To target a single cone, it was necessary to track the participant's eye movements and adjust the stimulus in real time to maintain position on the target cone throughout light delivery. Dark cones' increment thresholds were not significantly elevated compared to their paired, normally reflective cone, and the variability in dark

cone thresholds did not differ between persistently and intermittently dysflective cones. Therefore, a dark gap in the arrangement of the cones in an AO retinal image may contain a functional cone, which should be taken into account by human image raters and automatic algorithms.

However, the lack of relationship found between reflectivity and function in normal retinas does not indicate that reflectivity cannot be a marker of function in diseased retinas. For example, Carroll, Neitz, Hofer, Neitz and Williams (2004) used flood illumination AO ophthalmoscopy to study a dichromat with a gene encoding a nonfunctional pigment rather than an M pigment. One third of his cones had low reflectance, and the dark spaces were randomly interleaved amongst the bright cones, as would be expected of the packing arrangement of different cone classes. This was taken to indicate that the dysflective cones were the cause of the participant's dichromacy.

1.4.2 Cone typing

As described in § 1.1.1, there are three classes of cone photoreceptor, maximally sensitive to either long-, middle- or short-wavelength (L, M and S) light. The classes of individual cones have been identified in images of human retinas taken by AO imaging systems, including flood illumination ophthalmoscopy (Hofer et al., 2005; Roorda & Williams, 1999), scanning laser ophthalmoscopy (Sabesan, Hofer & Roorda, 2015; Sabesan, Schmidt, Tuten & Roorda, 2016; Schmidt, Boehm, Foote & Roorda, 2018; Schmidt, Sabesan, Tuten, Neitz & Roorda, 2018; Tuten, Harmening, Sabesan, Roorda & Sincich, 2017) and optical coherence tomography (OCT; Zhang, Kurokawa, Lassoued, Crowell & Miller, 2019).

In vivo cone classification in retinal images generally relies on densitometry: measurement of the effect of exposure to different wavelengths of light on the brightness of the cones. When the retina is adapted to white light, the retinal in the opsins undergoes a conformational change to all-*trans*-retinal to begin the phototransduction cascade to signal the presence of light. In this state, the opsin cannot absorb any more light, so the cones will appear bright in a retinal image. When the retina is dark-adapted, the retinal in the opsins return to their 11-*cis*-retinal state. They are therefore more likely to absorb imaging light when compared to their fully-bleached state, so the cones in the image will appear darker. However, the wavelength of the imaging light itself has an effect on cone brightness: many of the above studies were carried out with an imaging light near the peak absorptance of the L and M cones (e.g. 550 nm, Roorda & Williams, 1999, Hofer et al., 2005; 543 nm, Sabesan et al., 2015), which the L and M cones would absorb (leading to a dim appearance in the dark-adapted reflectance image) but the S cones would instead reflect (thus appearing bright).

Because cones have different reflectances even when exposed to white light, it is not sufficient to simply take a single such dark-adapted image and divide the cones into dim and bright ones to classify them as L/M and S. Some L or M cones may be as bright as some of the dimmer S cones simply because at time of imaging they had a higher baseline reflectance; likewise, some S cones with a lower baseline reflectance may appear as dim as the brighter L and M cones. Rather, it is necessary to take the *difference* of the brightness in the dark-adapted and the corresponding fully-bleached image, thus taking into account the differences in baseline reflectance. In a 550-nm absorptance image of 1 − (dark-adapted image / fully-bleached image), the S cones will appear as a few dim cones in an array more heavily populated with bright L and M cones (Hofer et al., 2005; Roorda & Williams, 1999).

Having separated out the S cones, adapting light of different wavelengths – notably 470 and 650 nm (Hofer et al., 2005; Roorda & Williams, 1999) or 470 and 680 nm (Sabesan et al., 2015) – can be used to classify the remaining cones as L and M, which is a harder task due to their heavily overlapping absorptance spectra. When the brightness of the cones in pairs of images taken following the two different selective bleachings are plotted against each other, either using the raw selectively-bleached images (Roorda & Williams, 1999) or using 1 − (selectively-bleached image / fully-bleached image) (Hofer et al., 2005), two slightly overlapping populations emerge that are taken to represent the L and M cones.

Other similar approaches include dynamic densitometry (Sabesan et al., 2015), in which AOSLO videos capture the time course of cone reflectance changes during full and selective bleaches, and the phase dynamics of changes in optical path length in response to brief flashes of light in AO-OCT images (Zhang et al., 2019), which has also been linked to the proportion of bleached photopigment.

1.4.3 The Oxford AOSLO

The work in this thesis uses the Oxford AOSLO described by Young et al. (2018). A key advantage of this instrument is that it uses a periscope to elevate the participant's eye above the level of the instrument. A dichroic is used to separate the imaging wavelength (850 nm) from the visible wavelengths. This allows a wide view, that does not pass through the optical system, of fixation targets and stimuli on a display beyond the instrument. However, the deformable mirror in the Oxford AOSLO has limited stroke, meaning that a participant's own glasses or contact lenses, or a set of trial lenses, are sometimes necessary to bring the inner segments into focus. Additionally, frame reconstruction is performed offline, with only spatially distorted and temporally downsampled video streams available with a short delay during the imaging session.

1.4.4 Interim summary

The introduction of AO to retinal imaging has allowed us to study the human retina *in vivo* on the smallest scale to date. The meaning of the varying intensities of the bright blobs taken to represent cone photoreceptors has been systematically studied at different timepoints and with functional testing enabled by real-time correction for eye movements. Furthermore, cone reflectance has been manipulated to provide the first *in vivo* assessments of human cone photoreceptor class and L:M cone ratio. The ability to use AO retinal imaging for the measurement of eye movements at single-cone resolution and referenced to the photoreceptor mosaic provides an important opportunity to advance the study of FEM and their effect on vision.

1.5 Thesis overview

The combination of cone inputs by the retina, LGN and visual cortex has been extensively studied, and this combination has been found to be linear up to the early cortex (Derrington et al., 1984; Solomon & Lennie, 2007; Sun et al., 2006). Nonetheless, some aspects of colour perception defy attempts to express it in a Euclidean colour space. This includes the super-importance of hue, an effect in which discrimination of differences of hue is finer compared to that of saturation than would be predicted by a colour model of linear combination of the cardinal mechanisms. The possibility that this arises from the existence of separate neural channels for hue and saturation is tested in Chapter 2.

Evidence for the hypothesis that individual differences in the relative densities of different cone classes affects spatial resolution for L- and M-isolating stimuli has finally been uncovered in parafovea (Danilova et al., 2013), yet the longstanding reliance on spatially extensive gratings meant that the functional relevance of L:M cone ratios in the fovea had yet to be ascertained. The denser photoreceptor packing in central retina may mean that the limits of the eye's optics may prevent any such effect. Chapter 3 presents the first experiment to use the more appropriate spatially local targets in the fovea.

Although some effects of L:M cone ratio on chromatic vision have been discovered, both the challenge of uncovering them and their limited number suggests that some further processes are at work to counteract any effects. One potential such mechanism may be resampling the scene with FEM. By shifting the image across the retina, a greater number of cones of a useful class may be able to sample crucial parts of the stimulus and improve perception, particularly in the case of small targets with fine spatial detail most easily seen by a minority cone class where there may be a very low number of the relevant class of cones in the first instance of sampling. In Chapter 5, the

AOSLO is leveraged to provide extremely high-resolution measurements of FEM while gathering performance data for a high-acuity spatiochromatic task. These data allow us to advance the hypotheses of the necessity of FEM for vision further than any work to date by investigating the possibility that levels of eye movements may be personalised to an individual's retina.

Finally, further development of using AO retinal imaging to track single cones requires (i) excellent automatic cone localisation, and (ii) greater understanding of the quality of the assessment of such algorithms. The best-performing algorithm may vary between instruments, and machine learning approaches, while a rapid entry point into localising cones, may be outperformed by theory-driven programs. Meanwhile, no ground truth of cone locations exists so instead manual cone localisation is generally used to provide a gold standard against which to test the algorithms. These gold standards have never been evaluated for agreement between manual raters, meaning that different published approaches may have been tuned to differing beliefs about what cone cells look like in AO retinal images. Differences between manual raters may go some way towards explaining why the same algorithm appears to perform differently when re-evaluated in other papers (Mariotti & Devaney, 2015). A new algorithm is presented in Chapter 6 and its performance measured against the grouped assessments of multiple manual raters. Individual manual raters' assessments are also compared to the group gold standard to highlight the range of possible differences between raters.

Chapter 2

Are hue and saturation carried in separate neural channels?

Declaration: The work contained in this chapter is entirely that of the author (SER). It is adapted from a publication whose first author is the author of this thesis: Regan et al. (2018).

2.1 Introduction

Hue and saturation are important perceptual qualities of colour. As described in Chapter 1, the subjective quality of appearing blue, bluish-green or green is known as hue, while colorimetric purity's subjective counterpart is known as saturation. It is useful both in vision science and commercially to be able to quantify colours and the perceptual relationships between them. Colour spaces are models that attempt to achieve these aims.

Such spaces have been designed for a range of motivations. For example, the CIE 1931 colour spaces were introduced to provide a common standard in science and industry for the measurement and communication of colour, and for the translation of physical light measurements to colour appearance in a standard observer (Smith & Guild, 1931). This latter aspect rested on the trichromacy of the human visual system, which means that any visible light can be matched by adjusting the intensities of three independent primaries. Colour spaces such as CIE XYZ and CIE xyY arise from linear combinations of the three photopic light detectors in the human eye. Meanwhile, MacLeod-Boynton color space and Derrington-Krauskopf-Lennie (DKL) color space made representation of the underlying physiology a more central aim (Derrington et al., 1984; MacLeod & Boynton, 1979). Taking the $\bar{l}(\lambda)$, $\bar{m}(\lambda)$ and $\bar{s}(\lambda)$ spectral sensitivity curves of the L, M and S cones of the standard observer as their starting point, one axis of these spaces represents the comparison of long and middle-wavelength (L and M) cone signals, thought to be implemented in the midget ganglion cells

(Lee et al., 1998). The other axis represents the inhibition of excitatory short-wavelength (S) cone signals by the summed signals of L and M cones, thought to be implemented in the small bistratified ganglion cells (Dacey & Lee, 1994). MacLeod-Boynton space considers only the equiluminant plane, where the sum of L and M cone signals is held constant; DKL space additionally represents intensity variation (summed L, M and S) on a third axis and places the white point at the origin of the space.

Another desirable type of colour space is one that represents subjective perceptual differences between colours in an isomorphic manner. In such a space, the difference between two colours in any one direction would be the same as that in any other direction. MacAdam (1942) approached this matter by studying the standard deviations of isoluminant colour matches in CIE 1931 colour space. Rather than the circles of equal size that would be expected in a perceptually isomorphic space, the standard deviations were instead found to be arranged in ellipses in different sizes; furthermore, there was no possible Euclidean projection of CIE 1931 space that could rectify this. No attempt (e.g. CIE L*a*b*, CIE Luv) to construct a perceptually isomorphic Euclidean colour space has fully succeeded. Farup (2014) argues such a feat is impossible.

One notable exploration of these issues began in the field of textile colour. Sensitivity to perceptual differences can also be measured experimentally using the concept of a just-noticeable difference (JND), the point at which the difference between two samples is of the smallest possible amount that still allows an observer to see that they are different. In work to produce an Index of Fading for the fabric industry, Nickerson (1936) made JND measurements using Munsell papers along the dimensions equivalent to lightness, hue and saturation. She reported that one lightness step was equivalent to three hue steps and two chroma steps. These findings were later echoed by Judd (1968), who used Nickerson's fading formula to calculate the relationship between the number of hue JND steps required to trace out a circle around the neutral point of radius r saturation steps. In a perceptually isomorphic space, the value ought to be equivalent to $2\pi r$; in fact, it is $4\pi r$. The human visual system is therefore 1.5-2 times more sensitive to changes in hue than in saturation. This prompted Kuehni (2003) to suggest the existence of separate neural channels for hue and saturation, the former of which would operate on lines tangential to circles around the white point at a given intensity. The present experiments investigate the neural encoding of colour, with specific reference to the nonlinearities implied by hue super-importance.

2.1.1 Colour mechanisms

Understanding the processing of cone responses by later stages of the visual system (Stockman & Brainard, 2010), and ultimately understanding the link between these processing stages and

colour appearance, has been a central aim of colour vision research. As described in Chapter 1, the L/M-opponent and S-opponent mechanisms, represented in the MacLeod-Boynton and DKL colour spaces, were described as 'cardinal' mechanisms by Krauskopf et al. (1982). The special status given to these mechanisms stems from the highly-selective losses of chromatic sensitivity that were produced by adaptation to a temporally-varying stimulus modulated along one of the S, L–M, or L+M+S axes, compared to the broader losses produced by adaptation to intermediate axes of modulation. The chromatically opponent cardinal mechanisms identified by this psychophysical method have been strongly linked with the separate populations of cells in the macaque LGN (Derrington et al., 1984) that inherit their tuning from the midget and small bistratified retinal ganglion cells (Dacey & Lee, 1994; Lee et al., 1998).

Unlike the putative hue mechanisms (Kuehni, 2003), the other post-cardinal mechanisms whose existence has been hypothesised have been suggested to pass radially through the white point. These 'higher order' colour mechanisms tend to be described as recombinations of the cardinal mechanisms, maximally sensitive to intermediate directions between the cardinal axes (Eskew, 2009; Krauskopf et al., 1986). Effects ranging from non-cardinal adaptation (Krauskopf et al., 1986) to unique hues (Mollon & Jordan, 1997) have been attributed to such mechanisms. However, the tuning directions of cortical neurons show no sign of grouping into a limited number of distinct mechanisms, cardinal or otherwise (Lennie et al., 1990).

2.1.2 Related work

Recently, Danilova and Mollon (2016) compared thresholds for step-changes either in hue or in saturation to investigate the super-importance of hue. They used a 4AFC discrimination task on a neutral D65 background where a briefly-displayed target was differentiated from three distractors either by hue or saturation. They defined their stimuli in MacLeod-Boynton colour space, scaling it for individual participants by detection thresholds along the opponent axes. Reference points were selected along a +45° or −45° line in scaled MacLeod-Boynton space at six radial distances from the white point in all four quadrants of the colour space. Chromaticities for the targets and distractors were also defined along a +45° or −45° line through the reference points; therefore, the underlying trichromatic and opponent mechanisms had the same modulation across conditions. At most distances from the white point, they replicated the super-importance of hue. Very close to the white point, there was no significant hue super-importance. However, they did not directly test whether hue and saturation are carried in separate neural channels.

2.1.3 The present study

A classic method of determining whether a perceptual characteristic is carried in a discrete perceptual mechanism is to use an adaptation paradigm. Prolonged viewing of a stimulus can affect subsequent sensitivity to stimuli of the same or a similar kind. The degree of threshold elevation of test stimuli can indicate the extent to which they share adaptable mechanisms. Such a paradigm was responsible for the identification of the cardinal chromatic channels: the L vs M and S vs (L+M) channels that underlie the axes of the MacLeod-Boynton and DKL colour spaces described in § 2.1. Krauskopf, Williams and Heeley presented equiluminant stimuli that were temporally modulated through the white point in colour space while keeping the time-averaged stimulation of the L, M and S cones constant across conditions, an innovation that allowed post-receptoral mechanisms to be targeted. Two kinds of flicker elevated detection thresholds for only the opponent combination that had been habituated. The first of these was flicker that modulated the L and M cones in opposition while holding S excitation constant. The second was flicker that modulated the S cones while holding L vs M excitation constant. Conversely, flicker that varied in terms of a combination of both opponent mechanisms raised detection thresholds in all directions (Krauskopf et al., 1982).

In this study, stimuli were presented that were matched in their time-averaged trichromatic and cardinal modulations, but that differed in their temporal modulation of hue and saturation. To achieve this, the stimuli modulated both of the cardinal axes over time with a triangular waveform centered on the white point, but in different conditions the waveform of the S/(L+M) modulation was phase-shifted relative to that of the L/(L+M) modulation (see § 2.2.3: Chromatic modulation and Fig 2.1, left panel). Modulation around the white point ensured that the time-averaged excitation of the trichromatic and cardinal mechanisms was matched for all stimuli; thus, any differences between conditions could truly be attributed to post-receptoral, non-cardinal mechanisms. Different phases of combination produced modulations that were tangential or radial to the time-averaged white-point, varying predominantly in hue or saturation respectively.

In Experiment 1, discrimination thresholds were collected for stimuli that modulated in hue or in saturation. In Experiment 2, discrimination thresholds were collected for hue or saturation modulations following adaptation to either hue or saturation modulations. Thresholds following adaptation were analysed in terms of elevation compared to non-adaptation test thresholds. If hue and saturation were carried in separate neural channels, it would be expected that the elevation of flicker-detection thresholds would be higher for within-condition adaptation than for between-condition adaptation. That is, threshold elevation for adapting to hue and testing hue and for adapting to saturation and testing saturation should be higher than adapting to hue and testing

saturation and for adapting to saturation and testing hue.

2.2 Methods

2.2.1 Ethics

This study was approved by the Medical Sciences Interdivisional Research Ethics Committee at the University of Oxford.

2.2.2 Apparatus

All procedures were carried out in a dark room. Stimuli were presented on an NEC colour monitor (MultiSync FP2141SB) controlled by a Cambridge Research Systems ViSaGe graphics system running in palette mode giving an 8-bit index and a 14-bit per channel RGB-encoded output value. The monitor had a spatial resolution of 800 x 600 pixels and its refresh rate was set to 160 Hz. Following data collection, it was confirmed that measured thresholds were not instrumentally limited. The experiment was run from MATLAB R2011a. Staircases were controlled by the Palamedes toolbox (Prins & Kingdom, 2009). The monitor was gamma-corrected with a Cambridge Research Systems ColourCAL II colourimeter, and the spectral energy outputs of the CRT phosphors were measured by a Cambridge Research Systems SpectroCAL MKII spectral radiometer. Background luminance was confirmed using the ColourCAL. Viewing distance was maintained by a chin and forehead rest positioned 92 cm from the CRT monitor. Observers viewed the stimuli binocularly.

2.2.3 Stimuli

Colour space

All chromaticity coordinates were defined in terms of their location on a MacLeod-Boynton diagram (MacLeod & Boynton, 1979), as constructed by multiplying the Stockman and Sharpe 2-deg cone fundamentals (scaled so that L(λ)+M(λ) is equal to the luminosity function V(λ) and that S(λ) peaks at unity (Stockman & Sharpe, 2000)) by the spectral energy output of the CRT phosphors. All chromaticity coordinates were defined relative to the coordinates of Equal Energy White (EEW) at $14 cd/m^2$.

Chromatic modulation

All chromatic modulation was based on triangle-wave deviation from the coordinates of EEW along a given axis of the MacLeod-Boynton diagram (see Fig 2.1). Triangle waves were selected rather

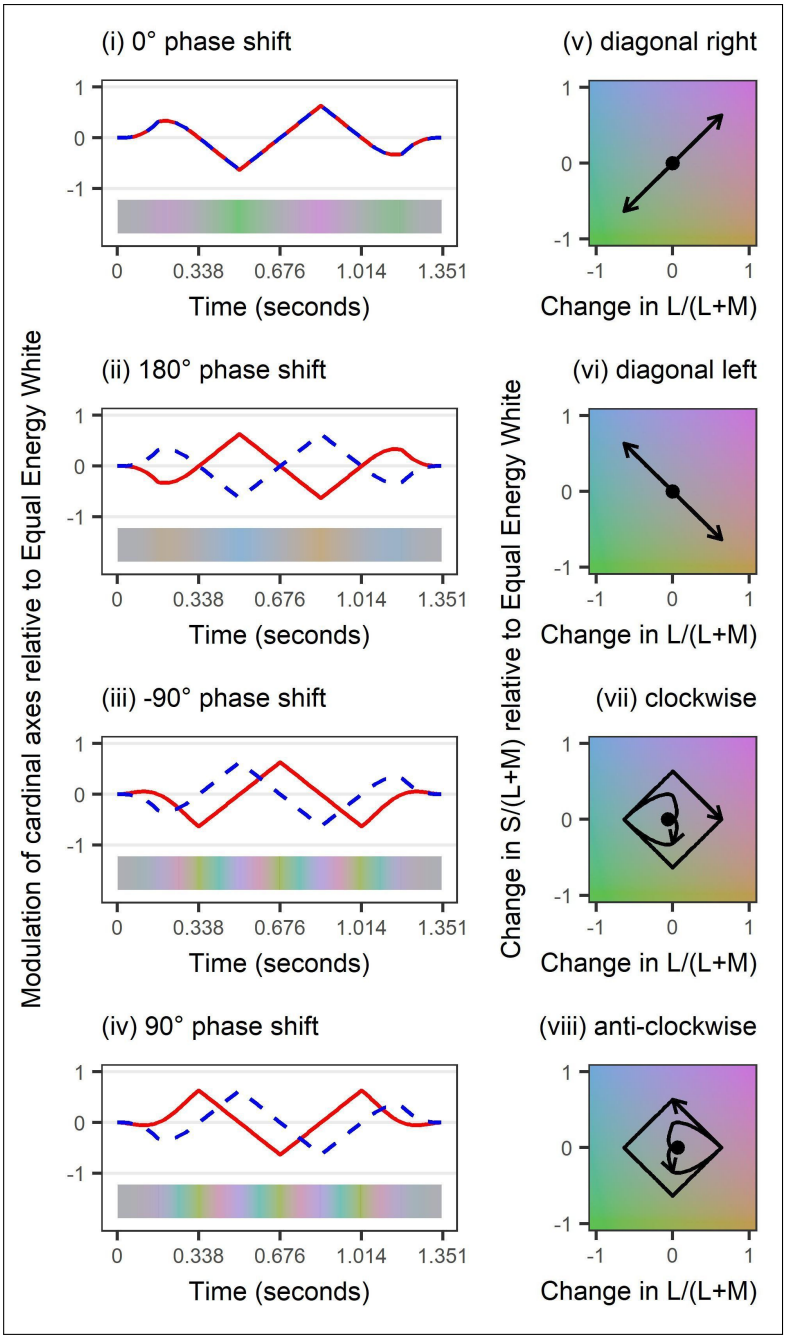


Figure 2.1: Chromaticity of stimuli temporally varying in hue or saturation as defined in terms of the cardinal mechanisms. For illustrative purposes, a hypothetical subject with extremely high cardinal thresholds is used so that the colours are clearly visible. The maximum modulation of the component mechanisms is set to 63.4% of cardinal flicker thresholds (see § 2.4.3 for why this value was chosen). Panels (i) to (iv) show the L/(L+M) (solid red lines) and S/(L+M) (dashed blue lines) modulations relative to Equal Energy White (EEW) used to create the four non-cardinal flicker conditions. Each test stimulus is composed of two periods of modulation; a cosine ramp has been applied to the first and last half-period. The adaptation stimuli use the same phase combinations, but modulations continue for two minutes at the start of each session, with six seconds of top-up adaptation before each trial. The starting phase is randomised for each stimulus. Corresponding colour appearance over time is inset at the bottom of each plot. Panels (v) to (viii) plot the resulting paths of the chromaticity coordinates (solid black lines with arrows) in MacLeod-Boynton space, zoomed in to the cardinal thresholds. When the phase of the S/(L+M) modulation is displaced relative to L/(L+M) modulation by (i) 0°, the path in MacLeod-Boynton space traces out (v) a diagonal line through the white point, slanting rightwards. When the phase shift is (ii) 180°, the path is (vi) a left-slanting diagonal line through white. When the phase shift is (iii) −90°, the path is (vii) a clockwise diamond. When the phase shift is (iv) 90°, the path is (viii) an anti-clockwise diamond. The mean chromaticity (black dots) for each stimulus is, on average, only 6.5% of maximum deviation from the white point. See also Visualisations 1-7 for movies of the test and adaptation stimuli in the Additional materials of this thesis.

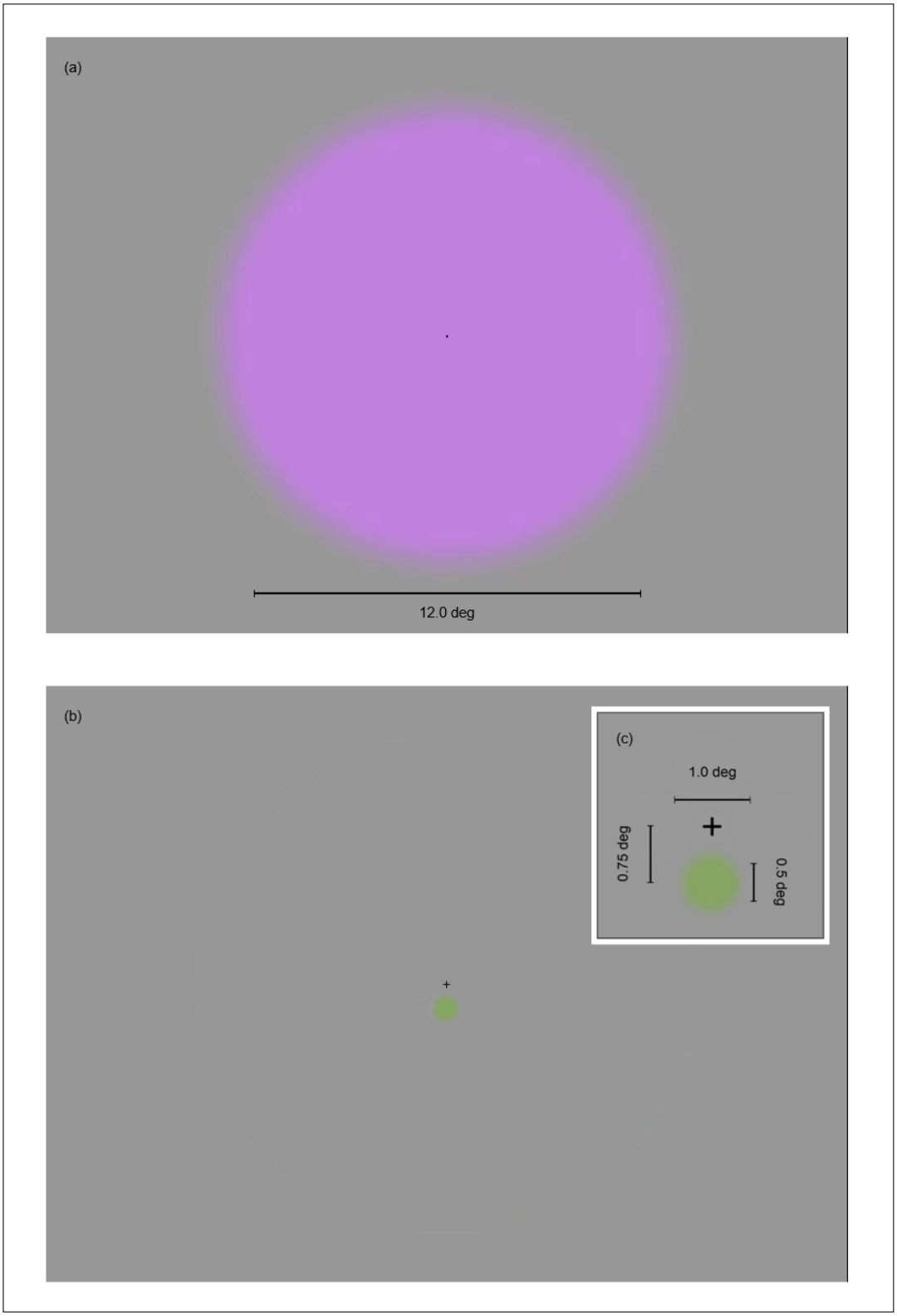


Figure 2.2: Spatial characteristics of the adapting and test stimuli. Panel (a): The spatial characteristics of the adapting stimulus. The central region of maximum chromatic contrast was 12 deg in visual angle in diameter. Its total diameter was increased by 3 deg by blur into the background chromaticity defined by a cosine function. Panels (b) and (c): The spatial characteristics of the test stimulus on an occasion when it has been presented in the bottom spatial quadrant. Its center is displaced from the fixation point by 0.75 deg of visual angle. The central 0.5 deg of the stimulus is of maximum chromatic contrast for a given frame. The edge of the stimulus blurs to the background chromaticity in a cosine ramp such that the total diameter of the stimulus is 1.0 deg. See also Visualisations 1-7 for movies of the test and adaptation stimuli in the Additional materials of this thesis.

than sine waves so that the wave spends the same amount of time at all cardinal contrasts, which would not be possible with waves that do not have gradients of a constant magnitude, such as sine waves.

Where the test stimulus was presented briefly for chromatic threshold detection (see §§ 2.2.5: Pre-experiment 2, Experiment 1 and Experiment 2 below), two cycles were presented and the starting phase was randomised. Where variation in terms of only one of the cardinal colour mechanisms L/(L+M) and S/(L+M) was desired (see § 2.2.5: Pre-experiment 2), coordinates of one axis were kept constant at the value of EEW while those of the other axis were temporally modulated in a triangle wave. During the adaptation experiment (see § 2.2.5: Experiment 2), 177.6 cycles (two minutes) of the adaptation stimuli were presented at the start of each session and 8.88 cycles (six seconds) were additionally presented prior to the onset of each trial.

Stimuli that varied predominantly in terms of saturation were created by shifting the phase of S/(L+M) modulation relative to L/(L+M) modulation by either 0 or 180° (such that the locus of chromaticities visited lay on a diagonal line passing through EEW on the MacLeod-Boynton diagram, slanted right or left respectively). Stimuli that varied predominantly in terms of hue were created by shifting the phase of S/(L+M) modulation relative to L/(L+M) modulation by either −90 or +90° (tracing out a diamond centered on EEW on the MacLeod-Boynton diagram in the clockwise or anti-clockwise direction respectively).

If the stimuli were allowed to modulate forever, the average chromaticity in all cases would be EEW. However, the test stimuli were presented briefly. In the case of hue there was no phase at which the chromaticity was EEW, and sudden saturation changes from and to EEW at the start and end of the stimulus presentation might facilitate detection. Therefore, for all test stimulus types the starting phase was randomised and a cosine ramp was applied to the first half-period and the last half-period of each two-period stimulus such that the stimulus began and ended at the coordinates of EEW and ramped smoothly up to and down from the maximum contrast of the component modulations. The ramped onsets and offsets of the modulations did cause the mean chromaticity on each trial to be slightly offset from the white point. However, the deviation of the mean chromaticity from EEW on each trial was not large (never more than 10.1%, and on average 6.5%, of the maximum contrast on any given trial), which ought to be imperceptible given that the maximum contrast was generally at or around threshold. Cosine ramps were also applied to the first and last half-periods of the adaptation stimuli, but as these stimuli were presented for a long period of time, the offsets of the mean chromaticity from EEW were even smaller than those of the test stimuli.

Temporal properties

The hue and saturation stimuli modulate at 1.48 Hz (each cycle is specified over 108 frames at the frame rate of 160 Hz). For the adaptation stimuli, modulations continue for two minutes at the start of each session, with six seconds of top-up adaptation on each trial. For test stimuli, two cycles are presented in 1.35s, including a 0.338s ramped onset and offset, and with peak contrast presented for 0.675s. Hue stimuli cycle through the full range of hue-angles. They also necessarily differ from the background in saturation and, for test stimuli, the ramp additionally introduces a slow saturation modulation. Saturation stimuli vary in their proximity to the white point, with two saturation minima per cycle. They include a change in hue at the two poles of the excursion. Therefore, the 'hue' stimuli are not exclusively modulations of hue, and the 'saturation' stimuli are not exclusively modulations of saturation. However, the stimulus design offers substantially different modulations to the putative hue and saturation mechanisms, while ensuring that hue and saturation stimuli are perfectly matched in the modulation they offer to the cardinal mechanisms.

Spatial arrangement

The spatial layouts of the adaptation and test stimuli are presented in Fig 2.2. All stimuli were discs with a blurred edge to limit use of edge cues. Adaptation stimuli had central region of maximum saturation that was 12 deg of visual angle in diameter, with an additional 3 deg in diameter of blur into the background chromaticity where the proportion of contrast from center to background was defined by a cosine function. Test stimuli had a central region of maximum saturation 0.5 deg in diameter, with a cosine ramp blurring the stimulus into the background chromaticity and adding an additional 0.5 deg to the stimulus diameter. Stimuli were centered 0.75 deg above, below, to the left or right of a fixation cross such that the central 0.5 deg of all stimuli fell within the smallest typical area of macular pigment, which is 2 deg across in 18-year-olds (Chen, Chang & Wu, 2001) and avoided the blind spot, 15 deg nasally medial to the fovea, and the S-cone free area in the central 0.4 deg of the fovea (Curcio et al., 1991). These locations will hereafter be referred to as the top, bottom, left and right spatial quadrants.

2.2.4 Task

For §§ 2.2.5: Pre-experiment 2, Experiment 1 and Experiment 2, triangle-wave flicker-detection thresholds were measured for the relevant conditions with the luminance adjustments obtained from heterochromatic flicker photometry (HFP). Each trial was four-alternative spatial forced-choice, presenting a stimulus modulating at 1.48 Hz in one of the four spatial quadrants on a

$14cd/m^2$ background metameric to EEW. Participants adapted for two minutes to the background. Observers were then asked to look steadily at the fixation cross and report the stimulus location by pressing buttons on a response box.

Modulation amplitude on each trial was determined by two interleaved 25-trial Palamedes QUEST staircases with a prior mean and SD obtained in pilot testing. The trial and final threshold chromatic contrasts were the mean of the posterior probability density function (King-Smith, Grigsby, Vingrys, Benes & Supowit, 1994). Because perceptual space is logarithmic (Fechner, 1966), the staircases varied modulation amplitude on a log scale.

The order of conditions was counterbalanced. Thresholds became lower with practice; therefore, to ensure that participants had adequately learnt to do the task before measurements were taken, all participants completed at least one practice run, following which the final threshold for each of the two conditions was the mean of the last six runs after the observer had reached an asymptotic level of performance for each.

2.2.5 Procedures

Experimental phases

In order to reduce likelihood that luminance cues could contribute to any of the thresholds measured, the experimental stimuli were defined on an equiluminant plane. To ensure stimuli were always equiluminant for all observers, heterochromatic flicker photometry (HFP) was first carried out in Pre-experiment 1 to establish the corrections needed for any individual differences in observers' sensitivity to luminance.

To create stimuli that modulated in terms of either hue or saturation, but whose time-averaged trichromatic and cardinal mechanism modulations were held constant, the two cardinal mechanisms were simultaneously modulated with relative phase shifts of 0° and 180° (saturation stimuli) and -90° and 90° (hue stimuli). It was necessary that the time-averaged modulation to the cardinal mechanisms be equal within each hue or saturation stimulus. Therefore, thresholds for triangle-wave modulation along the L/(L+M) and S/(L+M) axes of MacLeod-Boynton colour space were measured in Pre-experiment 2. When the modulations of these mechanisms were combined for the hue and saturation stimuli, the maximum deviation from EEW along each cardinal axis was a given proportion of the flicker threshold for that same axis; for example, if the maximum amplitude of the L/(L+M) wave for a given stimulus was 0.7 times the L/(L+M) flicker threshold, the maximum amplitude of the S/(L+M) wave was 0.7 times the S/(L+M) flicker threshold.

Experiment 1 measured modulation thresholds for these phase-shifted non-cardinal stimuli

and evaluated the differences in thresholds both between all four stimuli and between the combined thresholds of the hue stimuli and of the saturation stimuli. However, any simple differences in modulation thresholds would not indicate whether hue and saturation were carried in different neural channels. Experiment 2 measured thresholds for hue or saturation modulation following adaptation to either hue or saturation modulation and analysed for a difference between thresholds following within-condition and between-condition adaptation. If hue and saturation were carried in different neural channels, adaptation to a hue stimulus should raise thresholds for hue test stimuli but not for saturation test stimuli, and vice versa.

Pre-experiment 1: Heterochromatic flicker photometry

Equiluminant stimuli in different parts of the MacLeod-Boynton diagram were derived via subjective HFP (Pokorny, Smith & Lutze, 1989) to control for individual differences in spectral sensitivity of the luminance mechanism. HFP was carried out on the CRT monitor described in § 2.2.2. Chromatic modulation constituted a triangle wave along one of the two cardinal axes, and relative luminance of the poles was adjusted until the flicker was minimal. In order that the participants did not match the appearance of the flickering stimulus to the background, the background was black. The nominal target luminance was $14cd/m^2$. Prior to the task, participants adapted for two minutes to a $14cd/m^2$ background metameric to EEW. To limit rod intrusion (at least to the levels that would occur in the threshold measurements), staircases were very brief and each setting was preceded by a 10 s re-adaptation period to the background. Although little adjustment should be necessary for S/(L+M) variation, HFP was performed for chromatic temporal modulation along both the L/(L+M) and the S/(L+M) axis. Individual measurements were carried out for each of the four spatial quadrants in the spatial arrangement described in § 2.2.3: Spatial arrangement due to the possibility of retinal heterogeneity (Wyszecki & Stiles, 1982).

Observers were asked to look at a fixation cross and respond to the apparent flicker of one spatial quadrant. Constant-chromaticity stimuli metameric to EEW were placed in the three spatial quadrants that did not contain a flickering stimulus to facilitate steady fixation. The flicker frequency was 16 Hz, which is lower than typical HFP but suitable for the lower luminances achieved by a CRT monitor (Moreland, Robson & Kulikowski, 2001). Maximum chromatic deviation was set to three times the estimated threshold from pilot tests. Observers pressed up and down buttons on a response box to increase the luminance of one pole of the chromatic axis while decreasing the other on an up-down staircase with seven reversals, where the point of least flicker was defined as the mean of the final four reversals. They were instructed to explore the range of flicker until

they had 'changed direction' in the flicker space twice, at which point they should move back and forth in the zone of no apparent flicker, changing direction when and only when the flicker began to increase. Following two practice runs, participants completed this exercise six separate times for each of the four spatial quadrants for each axis, and the mean of those six runs for each quadrant was taken to be the luminance correction.

It was important to ensure that the corrections resulting from the HFP settings made along the cardinal axes could successfully generalise to equate the apparent luminances of chromaticities that did not lie along the cardinal axes. The corrections were therefore applied to intermediate chromaticity pairs and shown to the observers, flickering in counterphase at 16 Hz, before proceeding with the next stage of the experiment. The resulting stimuli appeared to be of steady luminance, indicating that the corrections did generalise to non-cardinal chromaticities.

Pre-experiment 2: Cardinal axis scaling

With the luminance adjustments obtained from § 2.2.5: Pre-experiment 1 and the task described in § 2.2.4 Task, triangle-wave flicker-detection thresholds were measured for the chromatic cardinal axes. The staircases were run in log steps in peak modulation contrast.

Experiment 1: Phase-shifted non-cardinal thresholds

The cardinal flicker thresholds obtained in § 2.2.5: Pre-experiment 2 were used to scale chromatic modulation along both axes such that the maximum deviation from EEW of the triangle wave of each chromatic axis was the same proportion of its cardinal flicker threshold. The staircases were therefore run in log fractions of cardinal threshold (that is, log(contrast / contrast at cardinal threshold). Flicker detection thresholds were measured for chromatic modulations of saturation (phase shifts of 0°, diagonal right; and 180°, diagonal left) and of hue (phase shifts of −90°, clockwise; and 90°, anti-clockwise; see Fig 2.1) using the task described in § 2.2.4 Task. Contrast of component modulations varied in terms of log threshold units. See also Visualisations 1-4 for movies of the test stimuli in the Additional materials of this thesis and § 2.5 for captions.

Experiment 2: Adaptation

Using the cardinal scaling as obtained in § 2.2.5: Pre-experiment 2 and the task described in § 2.2.4 Task, hue and saturation thresholds were gathered following adaptation to hue and saturation stimuli. To reduce the number of conditions, only the 0° phase shift (diagonal right) was used for saturation modulation and only the −90° phase shift (clockwise) was used for hue modulation.

Therefore, the four conditions were as follows: adaptation to diagonal-right saturation modulation, test diagonal-right saturation modulation; adaptation to diagonal-right saturation modulation, test clockwise hue modulation; adaptation to clockwise hue modulation, test diagonal-right saturation modulation; and adaptation to clockwise hue modulation, test clockwise hue modulation. As before, the background was metameric to EEW at $14cd/m^2$. Instead of beginning each session with two minutes of adaptation to the background, the participant adapted to a disc 12 deg in diameter modulating at 1.48 Hz in terms of either hue or saturation. The chromatic contrast of the adaptation modulation was the maximum that could be achieved within the gamut of the monitor, which was approximately 2.45 log threshold units above the pilot thresholds for hue and saturation. Each trial was preceded by six seconds of top-up adaptation with the same stimulus. See also Visualisations 5-6 for movies of the adaptation stimuli and Visualisation 7 for a movie of an adaptation stimulus followed by a test stimulus in the Additional materials of this thesis and § 2.5 for captions.

2.2.6 Observers

Four female and two male participants with normal or corrected-to-normal acuity and no known colour vision deficiencies completed the experiment. Observers A and B are authors of the associated publication.

2.3 Results

2.3.1 Experiment 1: Phase-shifted non-cardinal thresholds

The mean non-adaptation test detection thresholds for individual participants and across participants are shown in Fig 2.3. Thresholds are given as the log fraction of the individual subject's cardinal threshold, or log threshold units. Any negative value in log threshold units indicates that the threshold is lower than the individual subject's cardinal threshold. In two runs of one condition in one participant, the staircase did not converge to a stable position in the number of trials available, and the data points were excluded.

Mauchly's test of sphericity was violated (p = .022) and the standardised residuals were not all normally distributed (Shapiro-Wilk test of normality, p = .031 for diagonal left), so a repeated-measures ANOVA could not be run. Instead, a non-parametric Friedman test was run with four groups: diagonal right, diagonal left, clockwise and anti-clockwise.

There was a statistically significant difference in log threshold depending on the phase of combination of the L/(L+M) and S/(L+M) modulation, $\chi^2(3) = 11.400$, p = .010. Therefore, post

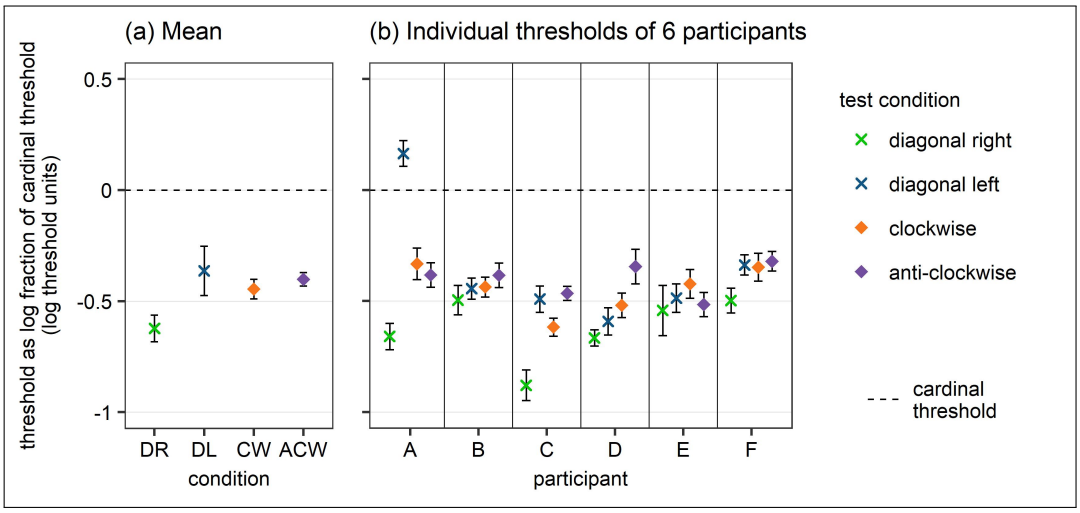


Figure 2.3: Detection thresholds in log threshold units (log fraction of the individual subject's cardinal threshold, black dashed line), for phase shifts of 0° diagonal right, 180° diagonal left, −90° clockwise and 90° anti-clockwise. Panel (a) plots the mean results across participants for Experiment 1; the error bars (black lines) represent the standard error across participants. Panel (b) plots the thresholds for each individual participant; the error bars (black lines) represent the standard error across runs. There was a statistically significant effect of phase of combination of S/(L+M) modulation with L/(L+M modulation ($p = .010$) for the four phase combinations; however, none of the pairwise comparisons were significant after Bonferroni corrections. The significant effect of phase could have been driven by the lower thresholds for 0° diagonal right compared to the other conditions for every participant.

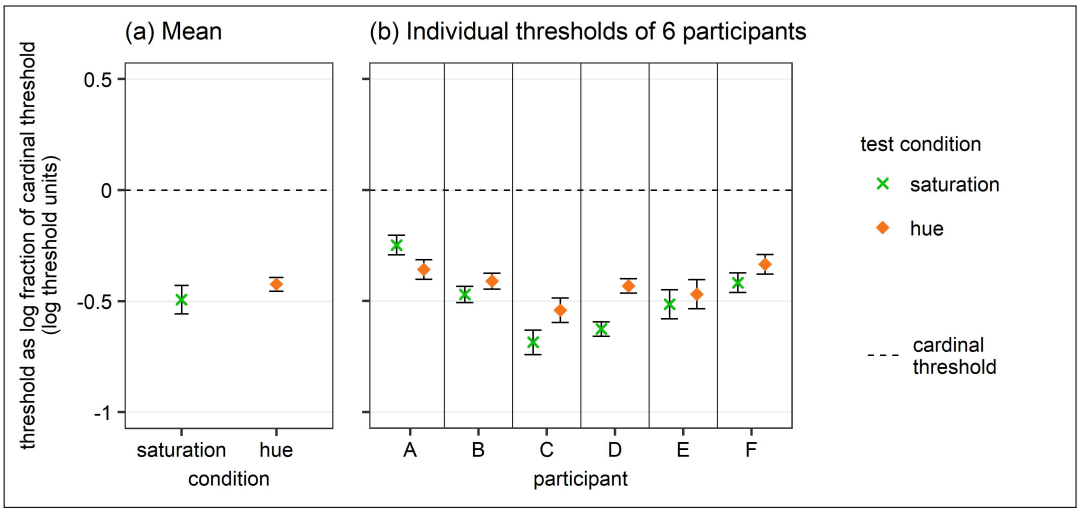


Figure 2.4: Detection thresholds in log threshold units (log fraction of the individual subject's cardinal threshold, black dashed line), expressed in terms of the mean thresholds for saturation and hue, where mean saturation threshold is the mean of 0° diagonal right and 180° diagonal left, and mean hue threshold is the mean of −90° clockwise and 90° anti-clockwise. Panel (a) plots the mean thresholds across participants; the error bars (black lines) represent the standard error across participants. Panel (b) plots the individual thresholds; the error bars (black lines) represent the standard error across runs. A paired difference test showed no significant difference between hue and saturation thresholds ($p = .163$).

hoc tests could be carried out. The assumption of symmetrical distributions was also violated for diagonal right − diagonal left (Shapiro-Wilk $p = .013$) and diagonal left − clockwise ($p = .027$), so sign rank tests were used; Bonferroni corrections set the two-tailed significance level at $p < 0.008$. None of the differences survived the Bonferroni corrections (diagonal right and diagonal left, $p = .031$; diagonal right and clockwise, $p = .031$; diagonal right and anti-clockwise, $p = .031$;

diagonal left and clockwise, $p = 1.000$; diagonal left and anti-clockwise, $p = .687$; clockwise and anti-clockwise, $p = .687$).

Our primary hypothesis concerned the difference between hue thresholds and saturation thresholds. The Shapiro-Wilk test indicated that the combined hue and saturation thresholds were normally distributed (saturation, $p = .912$; hue, $p = .891$), so a paired t-test was run. The mean hue thresholds (-0.424 log threshold units) were higher than mean saturation thresholds (-.493) contrary to hue super-importance, however the difference between the two thresholds did not prove significant ($p = .163$).

2.3.2 Experiment 2: Adaptation

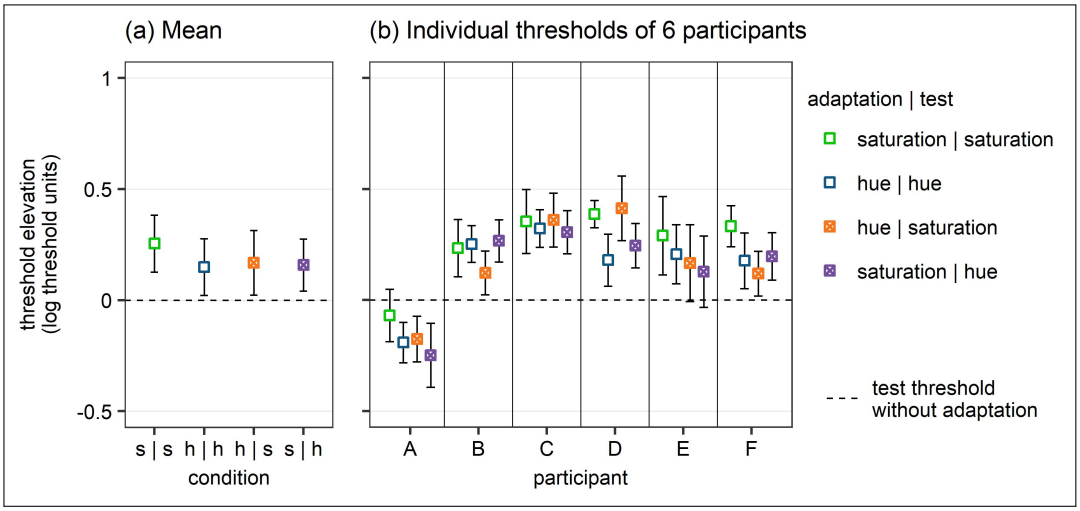


Figure 2.5: Post-adaptation detection thresholds, in log threshold elevation (threshold after adaption minus threshold without adaptation, black dashed line, in log threshold units), for the following conditions: adaptation to diagonal-right saturation modulation, test diagonal-right saturation modulation; adaptation to clockwise hue modulation, test clockwise hue modulation; adaptation to clockwise hue modulation, test diagonal-right saturation modulation; and adaptation to diagonal-right saturation modulation, test clockwise hue modulation. Panel (a) plots the mean results across participants for Experiment 2; because the data are presented in terms of the threshold elevation from pre-adaptation threshold, the error bars (black lines) represent the sum of the standard error across participants for the post-adaptation threshold and the standard error across participants for the corresponding pre-adaptation threshold. Panel (b) plots the post-adaptation detection thresholds for each individual participant; the error bars (black lines) represent the sum of the standard error across runs for the post-adaptation threshold and the standard error across runs for the corresponding pre-adaptation threshold.

The mean post-adaptation test detection thresholds for individual participants and across participants are shown in Fig 2.5. Thresholds are given in terms of their elevation compared to non-adaptation test thresholds, again as a log fraction of the individual subject's cardinal threshold. In one run of one condition in one participant, the staircase did not converge to a stable position in the number of trials available, and the data point was excluded.

If hue and saturation were carried in separate neural channels, the within-condition adaptation would be expected to raise thresholds more than between-condition adaptation. As seen in Fig 2.6,

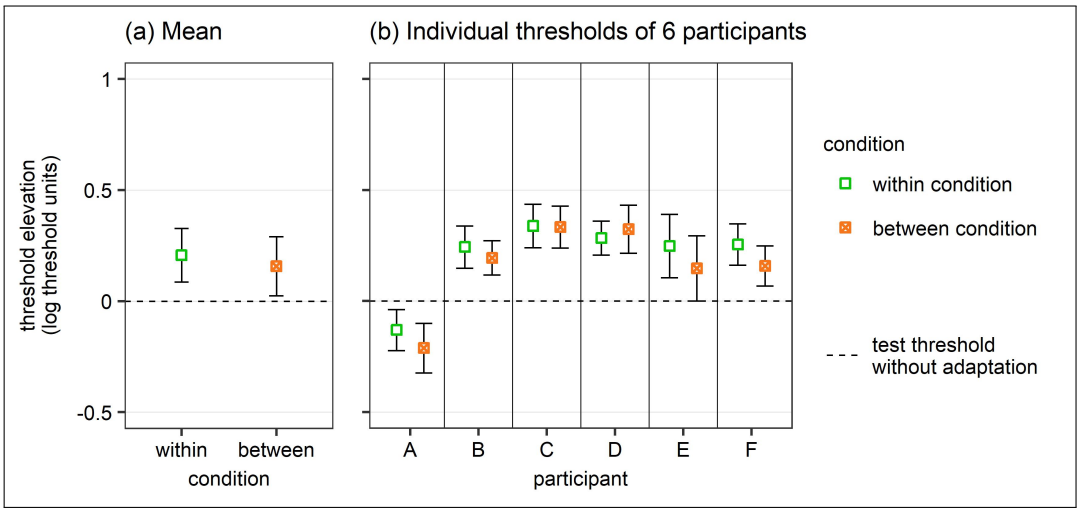


Figure 2.6: Post-adaptation detection thresholds, in log threshold elevation (threshold after adaption minus threshold without adaptation, black dashed line, in log threshold units), expressed in terms of the means of within-condition thresholds and between-condition thresholds, where mean within-condition threshold is the mean of adaptation to diagonal-right saturation modulation, test diagonal-right saturation modulation and adaptation to clockwise hue modulation and test clockwise hue modulation; and mean between-condition threshold is the mean of adaptation to hue modulation, test saturation modulation and adaptation to saturation modulation, test hue modulation. Panel (a) plots the mean threshold elevations across participants; the error bars (black lines) represent the sum of the standard error across participants for the post-adaptation threshold and the standard error across participants for the mean of the corresponding pre-adaptation thresholds. Panel (b) plots the individual threshold elevations; the error bars (black lines) represent the sum of the standard error across runs for the post-adaptation threshold and the standard error across runs for the mean of the corresponding pre-adaptation thresholds. A Wilcoxon signed-rank two-tailed test showed that the difference between within- and between-condition thresholds was not significant ($p = .075$).

the saturation-saturation and hue-hue adaptation-test thresholds were combined into within-condition thresholds (green open squares) and the hue-saturation and saturation-hue thresholds were combined into between-condition thresholds (orange crossed squares). The Shapiro-Wilk test of normality showed that the within-condition thresholds were not normally distributed ($p = .005$), but the distributions were symmetrical (Shapiro-Wilk for within − between, $p = .349$). Therefore, a Wilcoxon signed-rank two-tailed test was run. Although median within-condition adaptation thresholds (0.252 log threshold elevation) were higher than median between-condition thresholds (0.176 log threshold elevation), the difference was not significant ($p = .075$).

2.4 Discussion

Thresholds for hue differences have previously been found to be lower than those for saturation differences (Danilova & Mollon, 2016; Judd, 1968). Some have cited this as evidence for separate neural channels for hue and saturation (Kuehni, 2003). In this study, a direct test was carried out of the hypothesis that hue and saturation are carried in different neural channels.

2.4.1 The super-importance of hue

Colour spaces are subject to a number of striking perceptual nonlinearities. One of these is the super-importance of hue: when just-noticeable difference steps are measured along the Munsell hue and saturation scales, Judd indicated that there are at least twice as many just-noticeable steps in traversing the hue circle between any two colour points on either side of the desaturated center as would be expected from Euclidean geometry (Judd, 1968).

Hue super-importance could be explained by the existence of dedicated neural channels for detection of changes in hue that lie along lines tangential to circles centred on the white point in colour space. These channels would be recruited in succession by this study's hue stimulus as it modulates around the white point. The saturation stimulus would be unlikely to affect more than two hue mechanisms, one for each pole of the bipolar saturation modulation. Experiment 1 of this study found that sensitivity to hue and saturation modulation under threshold conditions was similar, which is consistent with a Euclidean model in which the isoluminant colours occupy a flat surface; there was therefore no evidence for the super-importance of hue under these conditions. However, the findings do not directly contradict the super-importance of hue because a lack of measurable difference in saturation and hue modulation thresholds does not preclude the existence of differences for sensitivity to colour modulations around a mean that is itself displaced away from the white point.

If the only threshold-determining chromatic channels were those modelled by the cardinal axes, the relative phases of the axes' modulations should not affect the resulting thresholds. In the present experiment, it was found that saturation thresholds were lower than hue thresholds, although not significantly. This replicates the findings of Danilova and Mollon (2016), who did not find a significant advantage for hue over saturation close to the white point.

When thresholds for hue and saturation were analysed separately as diagonal right, diagonal left, clockwise and anti-clockwise (see Fig 2.3) with a Friedman test, there was a significant effect of relative phase of combination; however, none of the post-hoc tests survived Bonferroni corrections. Inspection of the data indicates that the significance of the effect of relative phase was driven by the lower thresholds obtained for diagonal right compared to all other conditions. These results echo those of Fig 6 in Danilova and Mollon (2016): for quadrants 1 and 3 of the MacLeod-Boynton diagram, which correspond to this study's diagonal right, saturation thresholds may actually be lower than hue thresholds when very near the white point. They also found the difference to be non-significant.

Lack of difference between thresholds does not disprove the hypothesis that hue and saturation

are carried in different neural channels. An adaptation experiment was therefore carried out to directly test this possibility.

2.4.2 Separate neural channels for hue and saturation?

The hypothesis that hue and saturation are carried in different neural channels was directly tested with an adaptation paradigm. If hue and saturation were carried in separate neural channels, it would be expected that within-condition adaptation would raise thresholds more than between-condition adaptation. Furthermore, as Judd showed that Nickerson's measurements predicted an extremely large advantage for hue thresholds over saturation thresholds (Judd, 1968; Nickerson, 1936), the suprathreshold hue adaptation stimulus used in Experiment 2 should be a very strong adapting stimulus. However, there was no significant difference between within- and between-condition adaptation. Therefore no evidence was found for separate neural channels for hue and saturation. Moreover, the results provide no support for hue super-importance at suprathreshold levels, when the suprathreshold stimuli are considered in terms of their adapting effects rather than their discriminability from other clearly non-white hues.

2.4.3 Detection by cardinal mechanisms

If changes in hue angle of the stimulus were detected by a (highly sensitive, nonlinear) hue channel, it would be expected that hue and saturation thresholds would be different. However, hue and saturation thresholds did not significantly differ from one another, consistent with performance that was based on excitation of only the cardinal mechanisms or linear combinations of these mechanisms.

One explanation of the results that should be ruled out is that detection of the hue and saturation stimuli was in fact supported by only one of the cardinal mechanisms. If this were the case, the relative phase of the two modulations would be expected not to affect thresholds. This is indeed what was seen: with the exception of the Friedman test for the four detection thresholds, all differences between thresholds were non-significant. However, the measured hue and saturation thresholds are lower than the cardinal thresholds measured in isolation: the mean linear fraction across conditions and participants is 0.634 linear threshold units (as shown in Fig 2.1). This indicates that more than one mechanism is involved in the detection of hue and saturation stimuli.

The simplest account of the detection of the hue and saturation stimuli is that the two cardinal mechanisms detect them independently of one another. Since there are two of them, it is more likely that one or the other will meet the criterion for detection on any given trial, so the resulting

thresholds are lower than cardinal thresholds (values less than zero in Figures 2.3 and 2.4). In the case of the saturation stimuli, the two cardinal mechanisms would simultaneously but independently have the opportunity to detect the stimulus. In the case of the hue stimuli, the stimuli may stimulate first one cardinal mechanism than the other. While it is not impossible that stimulation of one cardinal mechanism facilitates detection by the other, this seems unlikely given that the final thresholds are substantially below the thresholds for each of the mechanisms.

Therefore, cardinal threshold scaling was sufficient to allow both cardinal mechanisms to contribute to detection but, even so, there was little effect of the phase of their combination. There was no evidence that the dynamic change in hue was being detected, or for the excitation of any mechanism other than the cardinal mechanisms or linear combinations of these mechanisms.

2.4.4 Correlated neural noise

If hue and saturation are not carried in separate neural channels, an alternative explanation is required for the differences in hue and saturation thresholds that have previously been found away from the white point (Danilova & Mollon, 2016; Judd, 1968). Danilova and Mollon (2016) have advanced an alternative hypothesis: that of correlated neural noise. The firing of small bistratified cells (which are responsible for the S/(L+M) cardinal mechanism, Dacey & Lee, 1994) is correlated with that of spatially close midget ganglion cells (which are responsible for the L/(L+M) cardinal mechanism, Lee et al., 1998), Greschner et al., 2011. Correlated neural noise would make it easier to discriminate between hue stimuli, which would lie obliquely with the axis of correlation, than between saturation stimuli, which would lie along that axis (Danilova & Mollon, 2016).

2.4.5 Conclusions

Previous work has found striking differences between hue and saturation thresholds (Danilova & Mollon, 2016; Judd, 1968). Although these threshold differences had sometimes been attributed to separate neural channels (Kuehni, 2003), no one had directly tested this hypothesis. An adaptation paradigm, the classic test of whether perceptual mechanisms can be attributed to discrete neural populations (Krauskopf et al., 1982), yielded no evidence of separate channels for hue and saturation.

2.5 Appendix: Visualisation captions

The following visualisations are available in the Additional materials of this thesis.

Visualisation 1: Movie of the 0 deg phase-shifted (diagonal right) saturation test stimulus whose chromaticity specifications are given in Fig 2.1.i and v, and whose spatial specifications are given in Fig 2.2.c.

Visualisation 2: Movie of the 180 deg phase-shifted (diagonal left) saturation test stimulus whose chromaticity specifications are given in Fig 2.1.ii and vi, and whose spatial specifications are given in Fig 2.2.c.

Visualisation 3: Movie of the −90 deg phase-shifted (clockwise) hue test stimulus whose chromaticity specifications are given in Fig 2.1.iii and vii, and whose spatial specifications are given in Fig 2.2.c.

Visualisation 4: Movie of the 90 deg phase-shifted (anti-clockwise) hue test stimulus whose chromaticity specifications are given in Fig 2.1.iv and viii, and whose spatial specifications are given in Fig 2.2.c.

Visualisation 5: Movie of the 0 deg phase-shifted (diagonal right) saturation adaptation stimulus whose chromaticity specifications are given in Fig 2.1.i and v, and whose spatial specifications are given in Fig 2.2.a.

Visualisation 6: Movie of the −90 deg phase-shifted (clockwise) hue adaptation stimulus whose chromaticity specifications are given in Fig 2.1.iii and vii, and whose spatial specifications are given in Fig 2.2.a.

Visualisation 7: Movie of the overall time course of a hue-adaptation, saturation-test trial.

Chapter 3

Does L:M ratio affect foveal spatiochromatic resolution?

Declaration: The work contained in this chapter is entirely that of the author (SER) but for the following exceptions. 65% of the participants for this chapter were recruited by the SER and 35% were recruited by another DPhil student, Allie C Hexley (ACH). 50% of the participants were run by SER; 30% by an undergraduate student, Rebecca Williams, under SER's supervision; and 20% by ACH. SER built a prototype of the heterochromatic flicker photometry (HFP) apparatus and some control software. The final version of the HFP apparatus and control software, and the software for the conversion of HFP measurements to L:M ratio, were built and written by ACH.

3.1 Introduction

Amongst eyes in different human individuals, the long- and middle-wavelength sensitive (L and M) cones exist in very different relative numbers. While the average L:M cone ratio is around 2:1, it can vary from 0.4:1 to 16.5:1 (Carroll et al., 2002; Hofer et al., 2005) between individuals. In general, it has proven difficult to find any effect of L:M ratio on colour perception. As described more fully in Chapter 1, both colour naming (Miyahara et al., 1998) and unique yellow settings (Brainard et al., 2000; Miyahara et al., 1998; Schmidt et al., 2015) were unaffected by the L:M ratio of the observer, while the relationship between cone ratio and unique green found by Neitz et al. (2002) was methodologically questionable due to a failure to control stimulus luminance. Hood et al. (2006) did find that in the particularly imbalanced cases of carriers of deutan colour vision deficiency (CVD), modelled to have an average L:M ratio of 5:1, red-green colour thresholds were elevated relative to the control group and the protan CVD carriers, who would be expected to have more evenly balanced average L:M ratios of 2:1 and 1:2 respectively.

3.1.1 Genetics of L:M ratio

There are still unanswered questions as to what causes a particular cone photoreceptor to express either the L or M pigment, and, by extension, why the L and M cones might exist in different relative numbers between individuals. The q-arm of the human X chromosome contains a small array of genes that encode the L and M photopigments, including a locus control region (LCR; Nathans et al., 1989), one copy of OPN1LW (the gene encoding the L pigment), and one or more copies of OPN1MW (the gene for the M pigment; Nathans, Piantanida et al., 1986; Nathans, Thomas et al., 1986) downstream of OPN1LW. The LCR, which is required for the expression of the L and M genes (Nathans et al., 1989; Wang et al., 1992), is thought to bind stably to the promoter of either OPN1LW or the first copy of OPN1MW to form either an L or M cone respectively (Yamaguchi et al., 1997).

The number of copies of OPN1MW does not correlate with the L:M ratio of expressed pigment mRNA (Yamaguchi et al., 1997). Because only the first copy of OPN1MW is expressed in any considerable amount, Winderickx, Battisti, Motulsky and Deeb (1992) proposed that this binding took place in a position-dependent manner, favouring proximal copies. Smallwood, Wang and Nathans (2002) explicitly extended this idea to the binding to either OPN1LW or OPN1MW, in addition to the binding to different copies of OPN1MW. They created lines of transgenic mice carrying human L and M pigment genes in different arrangements. When the space between the L gene and the LCR was held constant and that between the M gene and LCR was increased, the number of photoreceptors expressing a reporter for the M gene decreased, and those expressing a reporter for the L gene increased. When the position of the L and M genes was reversed so that the M was before the L gene, M cones outnumbered L cones. Furthermore, the L:M cone ratio was not simply inverted: 99% of L and M cones were actually M cones, much greater than the proportion of L cones generated by the typical order of L and M genes, suggesting that the M promoter is more effective than the L promoter. The efficacy of the M promoter had previously been shown *in vitro* (Shaaban & Deeb, 1998). In normal vision, this likely offsets some of the effects of the distance from the LCR (Shaaban & Deeb, 1998; Smallwood et al., 2002).

L:M cone ratio has been linked with several genetic factors. Gunther et al. (2008) found at least one haplotype that was significantly associated with L:M ratio, and Deeb et al. (2004) identified five hypersensitive sites in the L and M gene array that may affect gene expression, although whether these affect L:M ratio is unknown. Lawrance-Owen et al. (2014) found a single nucleotide polymorphism (SNP), rs16844995, that accounted for 2.6% of the variance in L:M ratio. The SNP was close to the gene *RXRG* which encodes a receptor implicated in cone differentiation, for example by suppressing the S opsin during cone development (Roberts, Hendrickson, McGuire & Reh, 2005),

highlighting the potential for genes in or near *RXRG* to affect the probability of L and M expression.

In females, who have two X chromosomes, there is an additional stage prior to the development of an L or M pigment: random X inactivation. This leads to half of a female individual's L and M cones expressing the L and M opsins – and probably also the L:M ratio – of one parent, and half expressing those of the other parent, which may therefore lead to the individual's L:M ratio being the mean of the ratios of the two parents. Evidence for this comes from carriers of genes for red-green colour deficiency (Jordan & Mollon, 1997).

Carriers of colour vision deficiency

As touched upon in Chapter 1, carrying a gene for red-green colour vision deficiency (CVD) has been linked to L:M cone ratio. Following random X inactivation, half of a daughter's cones will have an active X chromosome with genes for both L and M opsins, and half will have the other X chromosome activated, with either just one opsin gene or two with very similar peak wavelengths. Across all carriers, the average L:M ratio in the half of cones with normal L and M genes will be 2:1. The other half of a daughter's cones will consist entirely of one class or another, or two sets with very similar spectral peaks. Across carriers of a deutan deficiency, the average resulting L:M cone ratio will become 5:1; across protan carriers, the average ratio will be 1:2 (Hood et al., 2006). Of course, the L:M ratio of the cones with an active normal X chromosome will vary hugely between individuals: colour-normal males (with just one, normal, X-chromosome) have been found to have a 30-fold range in L:M ratio (Carroll et al., 2002), assessed by electroretinogram (ERG) flicker photometry and genetics. However, on average, the two groups of carriers of a red-green CVD should still have ratios that tend towards the tails of the distribution of L:M ratios. This can be seen in the data of Jordan and Mollon (1997), who investigated carriers and their sons using the OSCAR test, a correlate of L:M ratio. The peaks of the distributions of OSCAR test settings of protan and deutan carriers fell on either side of the ratio distribution peak of normal controls, and the individual settings of carriers and controls were strongly correlated with those of their sons (Jordan & Mollon, 1997). Similar results were also obtained by Gunther and Dobkins (2002), who estimated L:M ratios from equiluminance settings for 16-Hz (n = 19, nine with a dichromatic father) and 4-Hz (n = 41, two with a dichromatic father) modulating red-green gratings and found that presumed deutan carriers had, on average, higher ratios than non-carriers, and presumed protan carriers had lower ratios.

3.1.2 Assessment of L:M ratios using flicker photometry

As introduced in Chapter 1, the classic non-invasive, behavioural measure of L:M ratios in awake, behaving humans is flicker photometry. Two lights are flickered in counterphase at a rate at which no colour difference is visible to the relatively sluggish chromatic channels, but, if the lights do not have the same luminance, a flicker of intensity is apparent to the faster luminance channel. If flicker photometry is carried out psychophysically, the observer typically adjusts the intensity of one or both lights until the appearance of flicker is reduced to the lowest possible amount. If it is combined with electroretinography (ERG), the experimenter may either record the ERG signal intensity difference generated by the two lights or adjust the test intensity until the ERG response is equal.

For rapid assessment of L:M ratio, flicker photometry can be carried out with only one pair of lights. Typically, red and green are chosen to maximise the difference in signal to the L and M cones; in the present study, lights with peak wavelengths of 630 and 545 nm were used. Assuming that each L and M cone contributes equally to the luminance channel, to an observer with an unusually high number of L cones, an L-favouring red light of a given objective luminance will appear brighter than to an observer with an average number of L cones, while the M-favouring green light will appear dimmer than to the average observer. This high-L:M ratio individual will therefore use less red and/or more green than average to produce a flicker null. A low-L:M observer will produce the opposite settings. Flicker photometry settings can be used in raw form as a correlate of L:M ratio, as in, for example, the OSCAR test (Le Sueur, Mollon, Granzier & Jordan, 2014). Alternatively, L:M ratio can be estimated from the settings by combining the spectra of the primaries with those of the cone fundamentals, as carried out in this chapter and in Chapter 5. For a more detailed assessment of photometric sensitivity curves, test lights of different wavelengths from across the visible spectrum can be compared to a fixed reference (e.g. Stockman et al., 2008).

Flicker photometry has often been combined with both ERG and genetics. Carroll et al. (2000) flickered test lights from 480 nm to 670 nm with an achromatic reference light at 31.25 Hz. Electrodes measured the retinal response to the stimulus, and the spectral sensitivity at the test wavelength was obtained by adjusting the intensity of the test light such that the response it generated was equal to that of the reference. Observers' L pigment genes were sequenced in order to identify the peak sensitivity of the L pigment, which can have a large effect on estimated L:M ratio (Bieber, Kraft & Werner, 1998), and finally the best weightings were identified to fit the L and M pigment templates to the ERG-derived spectral sensitivity curve, together with the lens density spectrum estimated from observer age. Ratios ranged from 0.6:1 to 12:1 with a mean of 2:1.

Flicker photometric ERG with genetics has been validated by AO retinal densitometry (described in § 1.4) in male, colour-normal observers, first in two observers using the first-generation Rochester AO system (Brainard et al., 2000) and then in eight observers with the second (Hofer et al., 2005). AO-estimated L:M ratios in the colour-normal male participants ranged from 1.1:1 to 16.5:1, while one additional observer, a protan carrier, had a ratio of 0.37:1. The ERG procedure was similar to that of Carroll et al. (2000), and the resulting L:M estimates were highly correlated with those derived from AO densitometry (r^2 = 0.95). However, the slope of the regression of the AO-derived estimate on the ERG-derived estimate was less than one and indicated that the contribution from an M cone to the ERG measure was 1.5 times greater than that of an L cone. Therefore, the indirectly-derived average L:M ratio of 2:1 was argued to equate to an underlying average L:M ratio of 2.5:1.

Kremers et al. (2000) compared a range of techniques for estimating L:M ratios, including flicker photometric ERG and heterochromatic flicker photometry (HFP). They carried out HFP in 13 male and two female participants, superimposing a test light in a range of wavelengths from 400 to 700 nm with a reference light of fixed, 560-nm wavelength. The lights were alternated at 25 Hz against one of two background fields: a white field to saturate the rods, or a 430-nm field to prevent any contribution from rods or S cones. Observers adjusted the intensity of the test light until the flickering of the stimulus reached a null or minimum point. Similarly to the method used with ERG, the mean spectral sensitivities deriving from this method were fitted to the best weightings of combinations of the L and M pigment templates and, additionally, the corrections for lens and macular pigment densities that fit the data best. Against the white, rod-saturating background, L:M ratios ranged from 0.16:1 to 5.7:1 with a mean of 1.34, and against the 430-nm rod- and S cone-saturating background, they ranged from 0.15:1 to 4.99:1 with a mean of 1.85.

Members of the group also carried out flicker photometric ERG in 36 colour-normal observers using a CRT monitor (Kremers et al., 1999), the broad primary spectra of which reduce the impact of individual differences in cone spectral sensitivities (Usui, Kremers, Sharpe & Zrenner, 1998). The two flickering chromaticities were targeted to a small range of L:M cone contrasts, including L- and M-cone isolation. Each L:M cone contrast was tested at four different overall intensities, and a linear relationship was found between contrast in a given cone and response amplitude. This enabled the identification of the absolute contrast provided to each cone that would result in a given ERG response amplitude (termed the 'threshold') for each L:M relative contrast. A model could then be fit to extract the relative weightings of the L and M cones. L:M cone contrast gain ratios ranged from 0.28:1 to 28.16:1 with a mean of around 4:1, although the fitting procedure did not work well for values larger than 10:1 so the upper value may be an overestimate.

Several participants in the flicker photometric ERG test also made HFP measurements, and the resulting ERG L:M estimates were particularly well-correlated (r^2 = 0.98) with those derived from the HFP with a violet background (430 nm) in five colour-normal males and one female protanopia carrier, with a slope of 1.01; however, the HFP-derived ratio estimates were systematically lower than those from ERG, which is unsurprising given the different means obtained in larger samples using the two methods. Nonetheless, given the high correlations between L:M ratios derived from HFP and ERG (Kremers et al., 2000), and those from ERG and AO retinal densitometry (Hofer et al., 2005), HFP is likely to provide an appropriate relative estimate of L:M ratio, preserving rank order.

HFP, V*(λ) and the L-cone weighting

In two key papers, Sharpe, Stockman, Jagla and Jägle (2005) and Stockman et al. (2008) proposed improvements to the luminous efficiency function, V*(λ), for a 2 degree circular patch centred on the fovea under photopic conditions. They used HFP, flickering a range of test wavelengths from 425 to 675 nm (Sharpe et al., 2005) and 420 to 680 nm (Stockman et al., 2008) in counterphase at 25 Hz with a 560-nm reference light. Importantly, they superimposed the test and reference stimuli on a steady background because the flickering targets selectively adapt the L and M cones, distorting the resulting luminous efficiency function; the background therefore reduced the effect of the L and M cone adaptation level changing with different test wavelengths. In 35 male and five female colour-normal observers, Sharpe et al. (2005) employed a 'neutral' background close to CIE standard illuminant D_{65} (metameric for the L and M cones with 566 nm) in an attempt to balance similarly adapting L and M cones at the first adaptation site, keeping neutral the second adaptation site of L vs M, and representing standard viewing conditions. The function could be defined as the sum of weighted L- and M-cone fundamentals (L(λ) and M(λ)) in which the weighting was determined to be 1.55 L(λ) + M(λ) when expressed in terms of relative quantal units. In fact, the resulting function proved to be a special case of V*(λ) for that particular background, which they later termed V_{D65}*(λ) (Stockman et al., 2008).

Further luminous efficiency functions gathered with this method were referred to as V_{μ}*(λ), where μ represents the monochromatic wavelength that excites the L and M cones in the same ratio as the adapting background being used. The later paper (Stockman et al., 2008) presented data in six male, colour-normal observers, flickering the HFP targets (with test wavelengths from 420 to 680 nm and 560-nm reference) against a range of monochromatic backgrounds from 430 to 670 nm, and seven bichromatic backgrounds mixing 478- and 577-nm light in different relative amounts. When the HFP null points for a given background are considered in terms of the ratio

of excitation of M/L cones by both the HFP lights *and* the adapting background, if the adapting light has prevented differential adaptation by the HFP lights at different test wavelengths then the null points should all elicit the same M/L cone excitation ratio. In fact, the excitation ratio varies, increasing with test wavelength. Furthermore, the adapting effect of the HFP lights changes with background wavelength: the M/L ratio curves are shifted downwards relative to the excitation ratio of the adapting backgrounds at short background wavelengths, and upwards at long wavelengths.

The new $V_{\mu}^{*}(\lambda)$ therefore defined the spectral sensitivity (luminous efficiency) when measured against a given adapting field, μ. The data from the previous paper (Sharpe et al., 2005) were reanalysed to take into account the variance of the mean adapting chromaticity with wavelength to give a new $V^{*}(\lambda)$, $V_{D65}^{*}(\lambda)$, in which the weighting of the L-cone fundamental, when expressed in relative quantal units, was 1.89.

Stockman et al. (2008) took pains to emphasise that the variance of the L-fundamental weighting with background meant that flicker photometry could not, unlike previously (Kremers et al., 2000), be trusted as a measure of L:M cone ratio. In particular, they strongly argued against the use of ERG without adapting backgrounds due to the much higher overall adapting chromaticity changes with test wavelength, citing (among others) Brainard et al. (2000) and Hofer et al. (2005) as examples of those who had falsely used it as an L:M ratio indicator. However, immediately beforehand they correctly pointed out that the ERG measurements of Brainard et al. (2000) (and by extension Hofer et al., 2005, who used exactly the same methods) had correlated with the values derived from directly imaging and classifying single cones in patches of retina. The strong correlation seems to call into question the claim that flicker photometric ERG without a background cannot provide a good estimate of L:M ratio.

3.1.3 Spatial resolution

As covered in detail in Chapter 1, any effect of L:M ratio on subjective colour perception and colour contrast thresholds is either absent, questionable, or reduced compared to what might be predicted from the extreme variation in ratio between individuals. Furthermore, the differing packing densities of the L and M cone classes strongly suggests that spatial resolution should be different for stimuli that isolate the L and M cones. Here, too, it has been challenging to find any evidence that this is the case. For a range of spatial frequencies, Green (1968) briefly flashed test stimuli, sine-wave gratings of a given colour, against backgrounds of another colour, in the combinations green test-green background, green test-red background, red test-green background and blue test-yellow background. Primarily making the observations himself but confirming with

two further observers, he estimated the degree to which the background reduced the contrast of the grating. For red and green gratings against backgrounds of the other colour, the resulting contrast sensitivity functions were very similar. Meanwhile, both resolution and contrast sensitivity for blue gratings against yellow backgrounds were depressed. Similarly, Cavonius and Estévez (1975) measured contrast sensitivity functions of the π_5, π_4 and π_3 (L, M and S) mechanisms. They monocularly presented two colour-normal observers with gratings of various spatial frequencies that either modulated the π_5 mechanism while holding π_4 constant, or vice versa; additionally, one colour-normal and one deuteranopic observer were shown π_3 gratings. The gratings modulated sinusoidally at 0.5 Hz to prevent fading while the observer adjusted the irradiance to keep the test pattern at threshold. The contrast sensitivity functions of π_5 and π_4 overlapped, while π_3 had a much lower peak contrast sensitivity than the other two. The two studies both indicated that the different packing densities of the L and M cones does not affect their resolution limit for chromatic gratings, but the particularly low packing density of S cones does impact grating resolution.

The lack of difference between contrast sensitivity functions for gratings that preferentially stimulate either the L or M cones has been attributed by Danilova et al. (2013) to the spatial extent of the gratings and to the way in which the L and M cones are arranged in the mosaic. A truly randomly distributed array of cones of different classes will exhibit clumping, and a number of studies have investigated whether the cones exhibit this level of grouping or whether they tend towards either more or less clumping than would be expected in a randomly distributed retina. Assessed using a range of techniques including histochemical measurements of bleaching response (Marc & Sperling, 1977), electron microscopy (Calkins, Schein, Tsukamoto & Sterling, 1994), microspectrophotometry (Bowmaker et al., 2003; Mollon & Bowmaker, 1992), phototransmittance imaging (Packer, Williams & Bensinger, 1996), and adaptive optics retinal imaging (Hofer et al., 2005; Roorda et al., 2001; Roorda & Williams, 1999) in human and non-human primates, the amount of clumping of L and M cones points to random distribution, although occasionally the level of clumping was higher than expected. Interestingly, this stands in contrast with the S cones. Curcio et al. (1991) compared S cone mosaics in patches of retina from the eyes of six human donors with 100 artificial mosaics with randomly assigned S cones. They found that the peripheral S cone mosaic was significantly different from the random mosaics.

Because the L and M cones exhibit clumping, the retina is able to use clumps of cones of the same class to probe a spatially extended target with the cone array's overall sampling resolution (Otake et al., 2000). Perhaps because the S cones do not exhibit clumping, researchers have found reduced sensitivity for blue gratings against yellow backgrounds (Green, 1968) and for the isolated

π_3 mechanism. Furthermore, the imbalance in the number of S cones is particularly extreme: they comprise only 5% of the overall cone population (Curcio et al., 1991) compared to an average of 62% and 33% for L and M cones. This echoes the work of Hood et al. (2006) described in Chapter 1, where only the starker 5:1 predicted L:M ratios of deutan carriers were different to the control average of 2:1, and not the predicted 1:2 L:M ratios of protan carriers.

Further evidence for an effect of L:M ratio on grating-based chromatic contrast sensitivity functions in a case of great asymmetry arises from the work of Miyahara et al. (1998). They explored the vision of two colour-normal, confirmed protan heterozygotes. One, BB, is described in § 1.2.4; the other was her sister, EB, who had an L:M ratio estimated from HFP matches of 0.03:1 (or 1:33.3), placing the L cones in a minority even smaller than that of S cones. Visual acuity for R- and G-phosphor gratings was measured for her and a normal control against different intensities of long- and middle-wavelength backgrounds. For the protan carrier, regardless of the grating phosphor, the acuity curves for the two background wavelengths overlapped when plotted in terms of M-cone trolands and did not overlap when plotted in terms of L-cone trolands. This suggested that grating detection was always mediated by the M cones in her predominantly-M mosaic, even in circumstances that were expected to favour detection by L cones (all R grating detection) or to equally favour detection by L and M cones (G grating detection against the middle-wavelength background) in a colour-normal observer. This was in contrast to the colour-normal control, for whom R grating detection seemed to be mediated by L cones and G grating detection to be affected equally by each background, suggesting equal detection by L and M cones. It may therefore be that only when a cone class forms a particularly low proportion of the overall number of cones can the clumping derived from their random packing arrangement, in combination with the spatially extensive nature of the gratings, fail to fully overcome their minority status.

3.1.4 L:M ratio affects spatiochromatic resolution

Danilova et al. (2013) were the first to find a clear link between L:M ratio and spatiochromatic resolution outside the extreme case studied by Miyahara et al. (1998) by identifying two important factors. First, any effect of spatial resolution may be seen most strongly in the parafovea, because optical factors limit resolution in the fovea (Thibos, 1998) and cone-to-ganglion cell signal convergence limits resolution in the periphery (Curcio & Allen, 1990). Second, because clumping can render the resolution of a cone class the same as the overall sampling resolution across classes (Otake et al., 2000), the spatial extent of the target should be small. Danilova et al. (2013) hence presented Landolt C optotypes to the parafovea at an eccentricity of 5 deg. With this paradigm, they were able

to find correlations between OSCAR test settings, a proxy for L:M ratio, and both size and contrast thresholds for L-isolating and M-isolating targets.

However, healthy human observers tend to fixate on a target with the preferred retinal locus (PRL), a point near, although not exactly coincidental with, the locus of highest cone density within the fovea (Putnam et al., 2005). To ascertain whether L:M ratio can impact perception of small chromatic targets, the above experiment must therefore be replicated in the fovea.

3.1.5 The present study

A key aim of the present study was to discover whether the large variation in L:M ratio between individuals leads to inter-individual variation in spatiochromatic foveal vision. Additionally, these data were necessary to validate a stimulus that could separately challenge the L or the M cone mosaic and could therefore be used in Chapter 5 to test whether the effect of fixational eye movements on visual acuity extends to the mosaics of individual classes of cone.

To these ends, a large sample of colour-normal participants with normal monocular visual acuity were recruited and their L:M ratios and L:M spatiochromatic threshold ratios were assessed. To test L:M ratio, a rapidly measurable correlate, red-green HFP null (described in § 3.1.2), was gathered. The majority of analyses were carried out using HFP scores as they tend to be more normally distributed. L:M ratio was estimated from HFP scores by combining the spectra of the red and green lights LEDs with the Stockman and Sharpe (2000) cone fundamentals.

As described in § 3.1.4, the most appropriate stimulus design to unearth any effect of cone class density on spatiochromatic vision is a small, local target; to that end, the present study uses a pixelated approximation of the L- and M-isolating Landolt Cs of Danilova et al. (2013). Both size and contrast thresholds were collected. It is not known precisely how size and contrast interact to produce the observer's performance for a given target. For their Landolt C gap location discrimination task, Danilova et al. (2013) reported that as target gap size increased, the contrast at threshold gradually decreased until it reached asymptote. The stimulus size must therefore be small enough to put pressure on the contrast required for spatiochromatic discrimination. Contrast thresholds were therefore obtained using the smallest target size for which a viable contrast staircase could reliably be produced. Meanwhile, to reduce the extent to which size thresholds might be limited by contrast, size thresholds were obtained using the highest contrast that was achievable by the display for both the L and M cones. HFP scores and L:M ratios were then compared with the ratios of L and M size and contrast thresholds, expressed as log(L/M) threshold.

In order to ensure that differences in L and M thresholds were not confounded by chromatic

aberration, a small initial experiment gathered thresholds for both incremental (where the target stimulates the isolated cone class more strongly than the background) and decremental (where the target stimulates the cones more weakly than the background) stimuli and ascertained that the resulting L:M threshold ratios were determined by cone class-isolation rather than stimulus colour (Danilova et al., 2013).

It has lately been found that, compared to Caucasian males, mean estimated L:M ratio is lower in males of African descent (McMahon, Carroll, Awua, Neitz & Neitz, 2008) and, according to conference abstracts, also in African American males, Native American males (Kuchenbecker, Neitz & Neitz, 2014), and Asian males (Kuchenbecker et al., 2014; Yamauchi, Yatsu, Kuchenbecker, Neitz & Neitz, 2013). Yet luminous efficiency functions, for example V(λ) (Stockman et al., 2008) and Y in the very commonly used CIE 1931 xyY colour space (CIE, 1932), make no attempt to reflect the true balance of inputs to the luminance channels of the world population. This impacts, for example, non-white consumers who must use displays whose calibration is based on luminosity functions of mainly white observers. In the present study, observers were asked to report their ethnic group so that any difference in ethnic groups could be analysed and reported in a peer-reviewed publication, although no ethnic group was specifically recruited.

3.2 Methods

3.2.1 Ethics

This study was approved by the Medical Sciences Interdivisional Research Ethics Committee at the University of Oxford.

3.2.2 Observers

100 observers who answered an advert for participants aged 18-35 years with normal visual acuity and normal colour vision participated in this study. 66 observers were female, 33 were male, and one declined to provide their gender. The mean age was 25.4 years old.

3.2.3 Acuity and colour vision deficiency assessment

Participants were required to be aged 18-35 years to limit the effects of age-related changes in lens density, have 20/20 monocular acuity and normal colour vision. Acuity in each eye was assessed using a standard wall chart and excluded if neither eye performed sufficiently. The eye with the best acuity was used for the rest of the experiment; where both eyes had the same acuity, the dominant eye was used.

Colour vision was assessed with the Oculus Anomaloscope, carried out in a dark room, and with the Hardy-Rand-Rittler (HRR) colour plates (Cole, Lian & Lakkis, 2006), which were placed on a book stand under a D65 lamp in an otherwise dark room; in both cases, the test did not begin until the observer had adapted to the illumination conditions for two minutes.

Participants additionally filled out a questionnaire to report: age, to verify that they met the age requirement; date of birth, for a separate investigation; gender, because it correlates with sex which mediates L:M ratio due to the different numbers of X chromosomes in males and females; whether they had a CVD, as some individuals who self-reported a CVD (not mentioned in this chapter) were recruited to the study for ACH's purposes; whether they had any relatives with CVD, as this affects L:M ratio in female observers (Jordan & Mollon, 1997); and ethnic group, as this also affects L:M ratio (Kuchenbecker et al., 2014; McMahon et al., 2008; Yamauchi et al., 2013). The options for the ethnic group question were taken from the recommended ethnic group question for use on a survey in England of the Office of National Statistics (2016) with the addition of a 'Prefer not to say' option.

All further tests were carried out in a dark room.

3.2.4 Heterochromatic flicker photometry

Apparatus and software

A combination of a red LED plus a narrowband interference filter and a green LED plus a narrowband interference filter yielded test and reference lights with peak wavelengths of 545 nm (10 nm full width half maximum; FWHM) and 630 nm (10 nm FWHM). A Maxwellian view system delivered the lights as a superimposed spatially uniform field with a diameter of 2 deg, viewed by the observer through a 2-mm artificial pupil. The LEDs were driven by an Arduino Uno, which was in turn controlled by a program in custom-written in MATLAB such that the lights flickered in counterphase at 25 Hz. Under these conditions, the colour appeared steady and the apparent luminance flickered rapidly. The participant responded to what they saw using a keypad. The conversion from HFP settings to L:M ratio was carried out by ACH using the cone fundamentals of the standard observer (Stockman & Sharpe, 2000; Stockman et al., 1999).

Task

While viewing the superimposed red and green lights, observers were required to start pressing either the right or left arrow key. They were to continue pressing the same arrow key until they noticed that the amount of flicker increased, at which point they should start pressing the opposite arrow key. This continued until nine reversals had occurred, and the estimate of the flicker null

was taken to be the mean of the last four reversals. 10 such settings were made over the course of two 15-min sessions. Because adaptive state influences HFP setting Stockman et al. (2008), observers looked away from the stimulus for one minute between each setting in addition to the break between sessions.

Instrument limitations

In designing the HFP apparatus, it was necessary to choose whether to prioritise separating the scores of those with high L:M ratios or low L:M ratios. Because high ratios were prioritised, many observers with low ratios reached the high end of the instrument's range. This compressed their HFP scores downwards and their estimated L:M ratios upwards. No adjustment was made to the data of such individuals.

3.2.5 Square C optotype thresholds

Apparatus

Stimuli were presented on a Display++ LCD monitor (Cambridge Research Systems Ltd., Riverside, Kent, UK) run by a Cambridge Research Systems ViSaGe graphics system running in palette mode giving a 8-bit index and a 10-bit per channel RGB-encoded output value. The monitor had a spatial resolution of 1440 x 1080 pixels and its refresh rate was set to 100 Hz. The graphics system was controlled by MATLAB on a Dell Optiplex PC[1].

The monitor was gamma-corrected with a Cambridge Research Systems ColorCAL II colorimeter, and the spectral energy outputs of the CRT phosphors were measured by a Cambridge Research Systems SpectroCAL MKII spectral radiometer. Background luminance was confirmed using the ColorCAL. Viewing distance was maintained by a chin rest positioned 175 cm from the display.

Chromaticity and spatial configuration

In order to keep the adaptation level of all cones equal, targets were presented against a background chromaticity metameric to Equal Energy White (EEW) at 41.7 cd/m^2. In two of the three chromatic paradigms, the targets isolated individual cone photoreceptor classes, L or M, by varying the signal presented to a given cone class while holding the signals presented to the other classes constant. The cone contrasts were defined relative to the background chromaticity in terms of the Stockman and Sharpe 2-deg cone fundamentals (Stockman & Sharpe, 2000; scaled so that L(λ)+M(λ) is equal

[1]The MATLAB and PC versions could not be verified due to COVID-19 building closures.

to the luminosity function V(λ), Sharpe et al., 2005, Sharpe, Stockman, Jagla & Jägle, 2011, and that S(λ) peaks at unity). In the third chromatic paradigm, the contrast presented to the L and M cones varied equally.

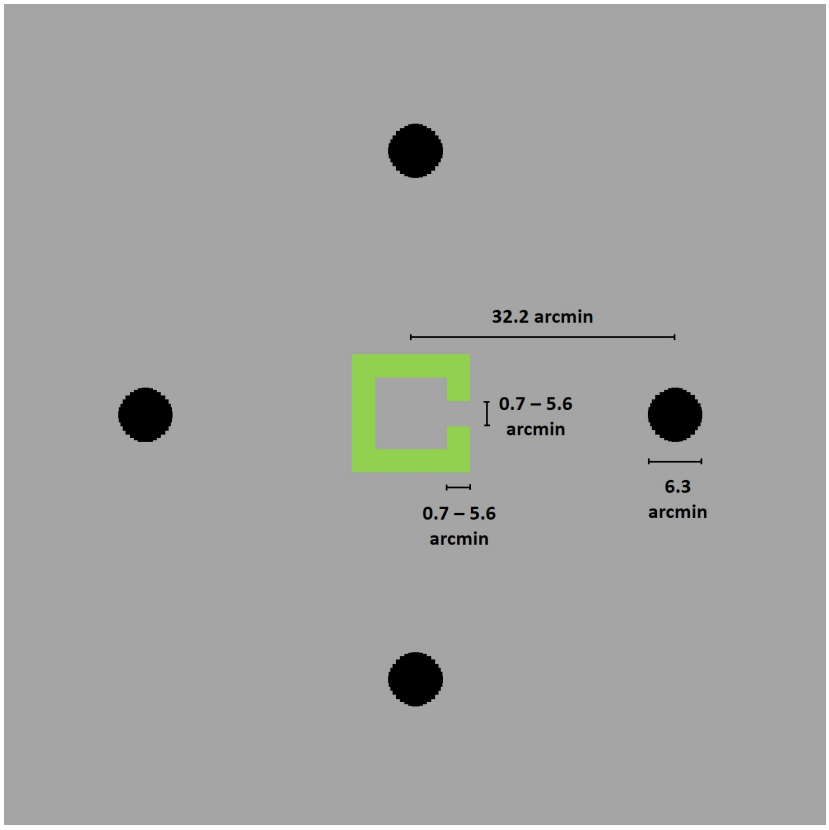


Figure 3.1: Spatial configuration of the square C optotype and the four dots indicating the optotype location. During the size threshold task, the gap and stroke width of the target ranged from 0.7-5.6 arcmin (1-8 pixels) and the 6.3-arcmin (9-pixel) dots were 32.2 arcmin from the centre of the stimuli. The target shown here has a gap and stroke width of 2.8 arcmin (4 pixels), the size used for the contrast threshold task. The background chromaticity was metameric to Equal Energy White (EEW) at 41.7 cd/m^2 and the optotype either isolated the L cone class, isolated the M cone class, or presented equal contrast to the L and M cone classes, all while holding the signals presented to the remaining cone classes constant. The level of contrast in the optotype was determined by a staircase.

The spatial configuration of the trial stimuli are shown in Fig 3.1. The stimuli were square-shaped C optotypes with a stroke width equal to one-fifth of the height and width, and a gap width equal to that of the stroke width. The smallest target was five by five pixels with a stroke and gap width of one pixel, corresponding to 0.715 arcmin, approximately 1.4 times the width of a foveal cone. The gap appeared on either the top, bottom, left or right of the stimulus. The targets were presented in the middle of four 'fixation' dots presented 32.2 arcmin above, below, and to the right and left of the centre of the target.

Staircasing

Size or contrast on each trial was determined by one of two randomly-interleaved staircases, each terminating after seven reversals. Staircases were controlled by the Accelerated Stochastic

Approximation method for up-down staircases (Kesten, 1958). The threshold was taken to be the mean of the final two reversals of the staircase (that is, the sixth and the seventh). Anecdotal reports from other members of the lab, backed up by simulations by SER, indicate that threshold estimates from such staircases are more accurate when the first response to the first target presented to the staircase is correct. Therefore, the first target contrast was above the expected threshold and, in the rare event that the observer's first response was incorrect (which only ever occurred for contrast thresholding), the staircase did not begin and the starting contrast was repeated until the observer had given a correct response. Once the staircase had begun, if the algorithm's suggested contrast for the next trial exceeded the maximum available contrast, the maximum available was presented instead. If the maximum was exceeded more than once, the staircase was terminated early and a threshold was not recorded. Because perceptual space is logarithmic (Fechner, 1966), contrast amplitude varied on a log scale. However, the constraints of a pixelated display mean that the available gap sizes can only be whole, rather than fractional, pixels; therefore, the size staircases were run on a linear scale rounded to the nearest pixel.

The maximum cone contrast available relative to the background was 0.21 for the L cone and .25 for the M cone. Therefore, size thresholds were gathered using the highest contrast available to both cones (0.21) and contrast threshold staircases also started at this contrast. Size threshold staircases started at six pixels as this was higher than all L- and M-isolating size thresholds produced by observers in pilot testing. To obtain contrast thresholds, a gap size of four pixels was used because this was the smallest gap size for which a viable contrast staircase was reliably produced by the majority of observers in pilot testing.

Trial

The observer was instructed that the targets would appear in the centre of the fixation dots, but that they did not need to fixate or hold their eyes still on any one part of the screen, and that they should simply look at the screen in a way that allowed them to best perform the task.

There was a warning tone 650 ms prior to the onset of the stimulus. The stimulus was presented for 500 ms. It was desirable to record a participant's assessment of the gap location after any eye movements had had their full impact, rather than a participant's early assessment. To prevent the response from being initiated too soon, participants were able to respond 500 ms after stimulus offset whether they believed the gap was on the top, bottom, left or right of the stimulus by pressing buttons on a keypad. There was aural feedback as to whether the response was correct or incorrect, and the next stimulus was not presented until after the response to the previous stimulus had been

recorded.

Task

Every experimental session began with a two-minute practice while the observers adapted to the display background. The first 10 observers made threshold settings for both increments and decrements. Once it had been shown that there was no significant difference in the L:M ratios measured by increments and decrements, the remaining participants made threshold settings for increments only. Where possible, eight repeats were collected for each condition.

First, observers made settings for L- and M-isolating size thresholds. This served the dual purpose of obtaining L−M log size threshold differences and indicating the sizes for which it would be possible for a given observer to make contrast threshold settings.

Second, observers made settings for L- and M-isolating and (L+M) contrast thresholds (the latter of which were used in Chapter 5). Piloting indicated that for an observer with size threshold n pixels, it was possible to run a contrast staircase for a target gap size of $n+1$ pixels without the staircase exceeding the maximum available contrast. Therefore, for the first 10 participants, contrast thresholds were collected for target gap sizes of $n+1$ pixels up to six pixels. For the remaining participants, contrast thresholds were collected for gaps of four pixels. The order of conditions was pseudorandomised and counterbalanced.

Optical limitations

The degree to which observers' optical acuity and/or class-specific cone sampling density would determine thresholds for L- and M-isolating optotypes was unknown. If optical acuity alone determined thresholds, L and M thresholds should be the same for each individual participant. If optical acuity partially determined thresholds, some or all L and M thresholds should be somewhat elevated relative to the thresholds that would have been obtained by a perfect optical system. Larger threshold elevations would be expected for (i) those observers whose acuity was close to the cut-off for inclusion in the study, and (ii) those instances where class-specific cone density was high and the resulting threshold for that cone class, with a perfect optical system, would have been low, making it more vulnerable to elevation by the eye's imperfect optics. The effect of both of these factors would be to reduce the differences between the L and M thresholds of individual observers.

3.3 Data preparation

3.3.1 Colour vision deficiency tests

Three lab members not directly involved in the study inspected the results of the two CVD measures, the Anomaloscope settings and HRR plate responses, and participants were retrospectively excluded if two of the three lab members judged them to have a colour vision deficiency. 10 participants were excluded on this basis.

3.3.2 HFP score and L:M ratio

A given HFP setting represents the amplitude of the test light at the estimated flicker null as a proportion of the maximum available test amplitude. The HFP score of an individual participant is taken to be the mean of 10 such HFP settings.

To estimate the L:M ratio of an individual, the Stockman and Sharpe (2000) L and M cone fundamentals were combined with the spectra of the red and green lights at the spectral radiances of the mean flicker null to give the impact of each of the lights on each of the pigments at a point where the apparent luminances of both lights were equal. Assuming that each L and M cone contributes equally to the luminance channel, the impact of the L fundamental relative to the M fundamental on the apparent luminance of one of the lights will be a constant multiplier, n, equal to the number of L divided by the number of M cones in the retina. Knowledge of the intensity of two different lights at a point of equal apparent luminance makes it possible to solve for n and recover the relative numbers of L and M cones.

A specific property of L:M ratio is that, because it is the number of L cones expressed always in relation to one M cone (e.g. 2:1, 0.5:1), a compression can be seen below ratios of 1:1. Therefore, when L:M ratio was used in further analyses, a measure was taken to counteract this compression. For L:M ratios below 1:1, the ratio was transformed such that ratios of 2:1 and 1:2 (that is, 0.5:1) were equidistant from 1. Where ratios of 3:1, 2:1 and 1:1 would be recorded and analysed as 3, 2 and 1, ratios of 1:2 and 1:3 would be analysed as 0 and -1.

3.3.3 Square C optotype threshold difference

Because perceptual space is logarithmic (Fechner, 1966), both size and contrast thresholds were analysed as log thresholds. L:M ratio and its correlate, HFP score, can only indicate the ratio of the number of L and M cones, not the absolute densities of the two classes. Therefore, it is more appropriate to relate them to the *ratios* of the size and contrast thresholds for L and M cones, rather

than the absolute thresholds. Threshold ratios were calculated as the log of the ratio of the L and M thresholds, which can be expressed as

$$TR = log\left(\frac{T_L}{T_M}\right) = log(T_L) - log(T_M) \tag{3.1}$$

where TR is the threshold ratio, T_L, the L threshold and T_M, the M threshold. This will henceforth be referred to as the L−M log threshold difference (or, specifically, the L−M log size threshold difference and the L−M log contrast threshold difference).

3.3.4 Observer exclusion

Of the 100 observers, one was excluded for not having normal (20/20) monocular acuity; 10 were excluded due to colour vision deficiency; two were excluded for being unable to correctly follow instructions for the heterochromatic flicker photometry (HFP) task; two for making inconsistent settings on the HFP task; one made implausible HFP settings; eight for having square C optotype thresholds that were too high to carry out the task; one for taking too long on the early parts of the experiment to complete the final task; and three had missing data. Therefore, the data of 71 observers were analysed.

3.4 Results

3.4.1 Heterochromatic flicker photometry and estimated L:M ratios

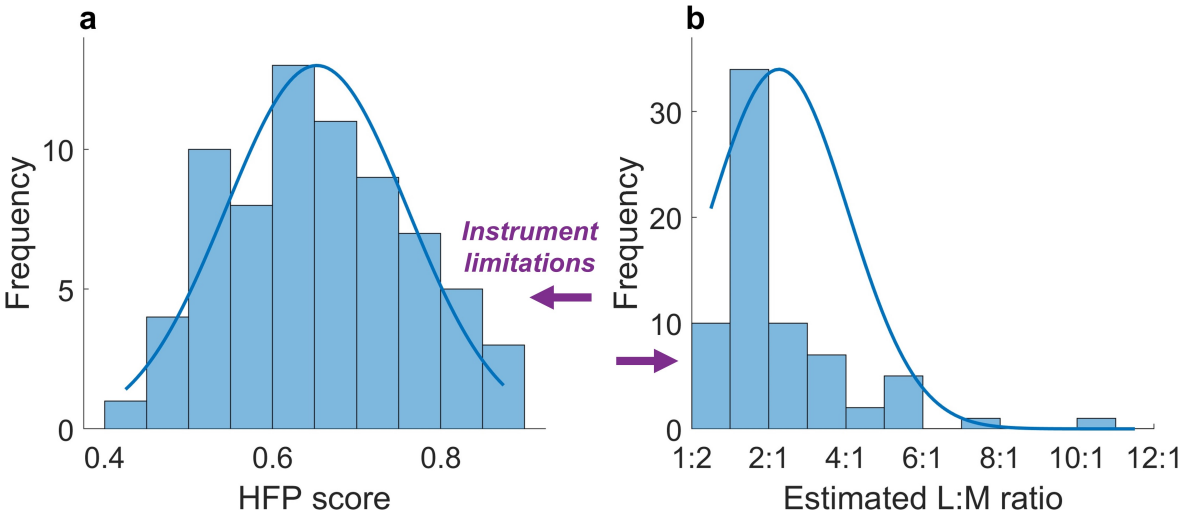


Figure 3.2: Histograms with fitted normal distributions for (a) HFP scores and (b) HFP-derived estimates of L:M ratio. Known instrumental limitations (purple arrows) compress the data at the upper end of the range of HFP scores and the lower end of the L:M estimates due to the negative relationship between HFP score and L:M ratio.

As seen from the histograms in Fig 3.2.a and b, the frequency distribution of HFP scores (a)

was normal, but that of estimated L:M ratio (b) was not (Shapiro-Wilk test, p = .388 and .000 respectively), the latter having a positive skew as expected from the transformation of HFP scores to ratio estimates.

3.4.2 Incremental vs decremental C optotypes

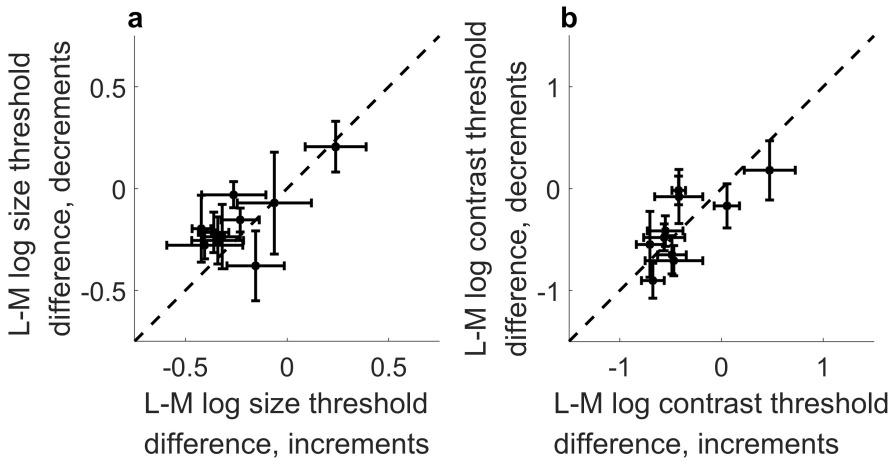


Figure 3.3: L−M log threshold differences for decremental vs incremental square C optotypes for size (a) and contrast (b) thresholds with error bars denoting the sum of the standard errors of the L−M log threshold differences. The dashed line shows the line of unity; deviation from this line would indicate that threshold differences are affected by a factor other than L:M ratio, perhaps chromatic aberration.

To control for any effects of chromatic aberration (Danilova et al., 2013), we measured L and M thresholds for both increments and decrements of cone signal relative to the background in 10 subjects, nine of whom went on to form part of the larger dataset. For both size (t(9) = -1.658, p = .132) and contrast (t(9) = 0.035, p = .973), any difference between increment and decrement L−M log threshold differences was not significant (see Fig 3.3.a and b respectively). The sign of the stimulus deviation from the background therefore did not affect the resulting ratio, and the main experiment proceeded with increments only.

3.4.3 Spatiochromatic thresholds

Fig 3.4.a shows the L-isolating and M-isolating log size thresholds (red and green crosses, respectively) for a subset of 12 of the final 67 observers, drawn equally spaced from the dataset ranked by L−M log size threshold difference from the highest (left) to the lowest (right) differences in order to illustrate examples the data without overcrowding the plots. Raw size as stroke and gap width in pixels is shown on the secondary y axis. One observer, AEK, has an L-isolating standard error (see error bar) of zero. This is not an error, but rather reflects the relatively large step sizes that had to be used for this task: gap sizes from one pixel upwards in one-pixel steps. This made it more likely than usual that an observer would obtain exactly the same threshold in multiple repeats. Fig 3.4.b

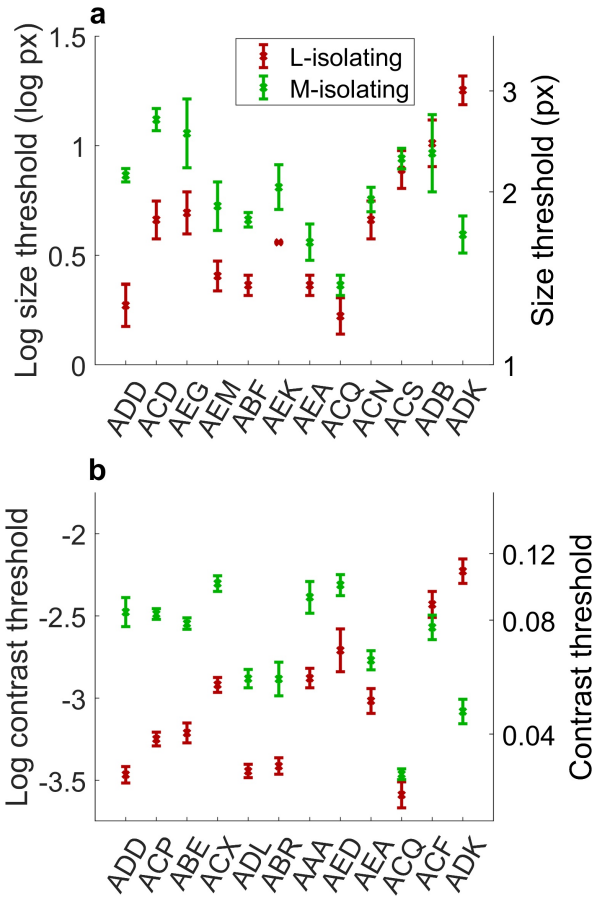


Figure 3.4: Examples of log thresholds for L- (red) and M-isolating (green) square C optotypes for size (a) and contrast (b) thresholds. Due to the large number of participants, for clarity, a subset of participants are shown, drawn from equally spaced points in the dataset when sorted by the L−M log threshold differences. Error bars show the standard error. The right-hand y axis shows the corresponding linear threshold values.

presents log contrast thresholds in the same way. Note that for both size and contrast thresholds, most observers' standard error bars for L and M stimuli do not overlap, suggesting a clear difference between L and M thresholds.

3.4.4 Spatiochromatic threshold differences

The frequency distributions of L−M log threshold differences for size (see Fig 3.5.a) and contrast (b) were normal (Shapiro-Wilk test, p = .003 and .000 respectively). Optical factors such as an observer's acuity (ability to read letters from a standard wall chart) are expected to elevate particularly low thresholds in a way that may limit the deviation from zero of an observer's L−M log threshold difference (yellow arrows).

3.4.5 Correlations

Scatter plots for relationships between HFP score, estimated L:M ratio, and L−M log threshold differences for size and contrast are presented in Fig 3.6. Yellow arrows indicate the most likely direction of compression of L-M log threshold differences by optical factors, and purple arrows

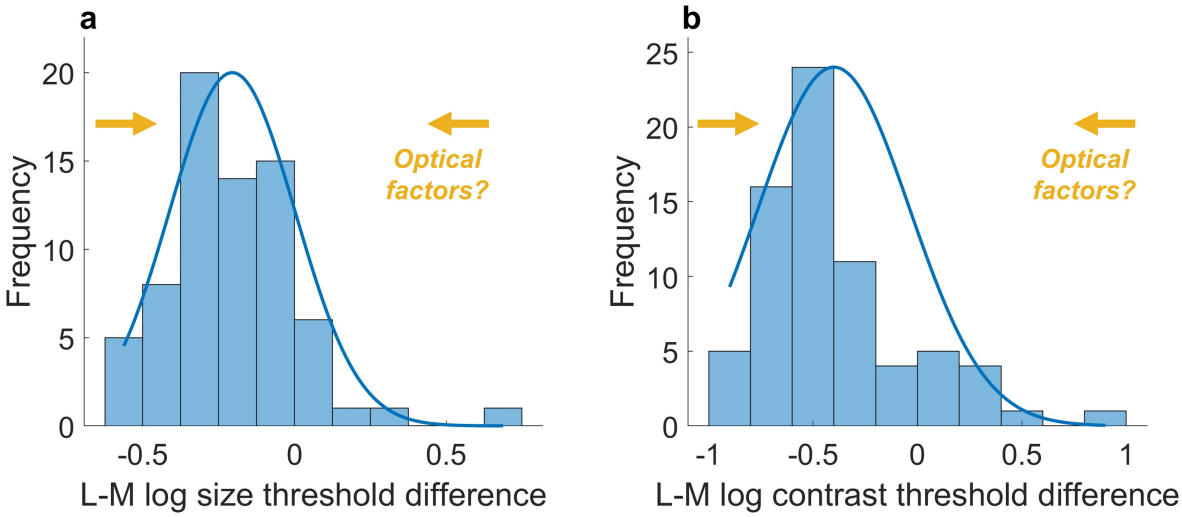


Figure 3.5: Histograms with fitted normal distributions for (a) L−M log size threshold differences and (b) L−M log contrast threshold differences. Both metrics are likely to be limited by optical factors (yellow arrows) for individuals with particularly asymmetrical L:M cone ratios.

show the direction of compression of HFP score and estimated L:M ratio by the choice of primaries for the HFP apparatus.

Fig 3.6.a shows L−M log contrast threshold difference, the intended primary measure of spatiochromatic threshold difference, plotted against HFP score. A linear regression line (red dashes) has been plotted for illustrative purposes only; because L−M log contrast threshold difference is not normally distributed, the non-parametric Spearman's rank correlation was carried out. The correlation was positive, which was expected as high values in both metrics indicate low L:M ratios, weak and significant (ρ = 0.364, p = .0019, n = 71). L−M log contrast threshold difference is also displayed against the estimated L:M ratios obtained from the HFP scores in Fig 3.6.b. As high HFP scores correspond to low L:M ratios, the direction of the relationship between contrast and L:M ratio is reversed compared to L−M log contrast threshold difference and HFP score. As neither metric was normally distributed, the non-parametric Spearman's rank correlation was carried out. The correlation was weak and significant (ρ = -0.362, p = .0022, n = 70). There was one fewer observer than in the previous correlation because the lowest HFP score was so low that an L:M ratio could not be estimated with the instrument as calibrated.

Size threshold difference was moderately correlated with both mean HFP setting (Spearman's ρ = .406, p = .0005, n = 71; see Fig 3.6.c) and contrast threshold difference (Spearman's ρ = .503, p = .0005, n = 71; see Fig 3.6.d)

3.4.6 Gender and family history of red-green colour vision deficiency

Fig 3.7 shows a pair of histograms of the frequency distributions of HFP score (panel a) and L−M log contrast threshold (panel b) for female (yellow bars) and male (purple bars) observers. The

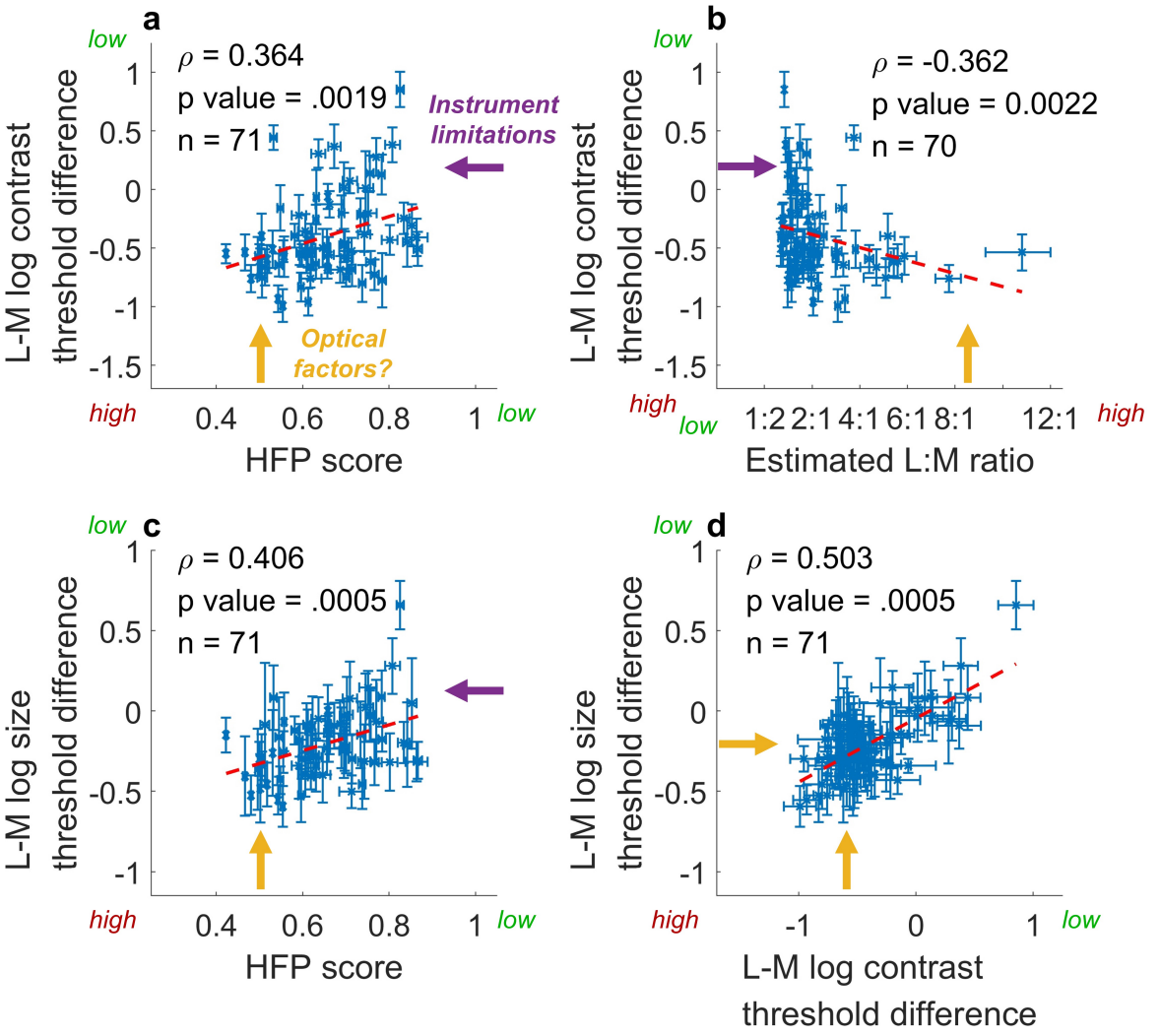


Figure 3.6: Scatter plots showing the relationships between HFP score, HFP-derived L:M ratio estimates, and L−M log threshold differences for size and contrast thresholds. As in Figs 3.2 and 3.5, purple arrows denote the HFP instrumental limitations and yellow arrows indicate the assumed optical limitations, both of which may depress Spearman's ρ. Red dashed lines show linear regressions for illustration purposes only. L−M log contrast threshold difference is plotted against HFP score (a) and estimated L:M ratio (b), and L−M log size threshold difference is plotted against HFP score (c); limitations from the instrument and optical factors may pressure the data into a more triangular shape, which is particularly visible in the first two plots. L−M log size threshold difference is also plotted against L−M log contrast threshold difference (d). Error bars for HFP score and estimated L:M ratio represent the standard error and error bars for L−M log threshold differences represent the sum of the standard error.

vertical dashed lines show the HFP settings of those reporting a first- or second-degree relative with a CVD, all of whom in practice reported fathers with a red-green deficiency. One of the 49 female participants reported a CVD relative. She had an HFP score in the 2nd percentile of the whole sample population regardless of gender, the lowest of all female participants, and her L−M log contrast threshold difference was in the 44th percentile. Two of the 22 male participants made the same report, one with an HFP score in the 46th percentile and an L−M log contrast threshold difference in the 71st percentile, and the other with an HFP score in the 84th percentile and an L−M log contrast threshold difference in the 15th percentile. (Not marked is the female participant who reported a first cousin with a colour vision deficiency causing the confusion of greens and browns.)

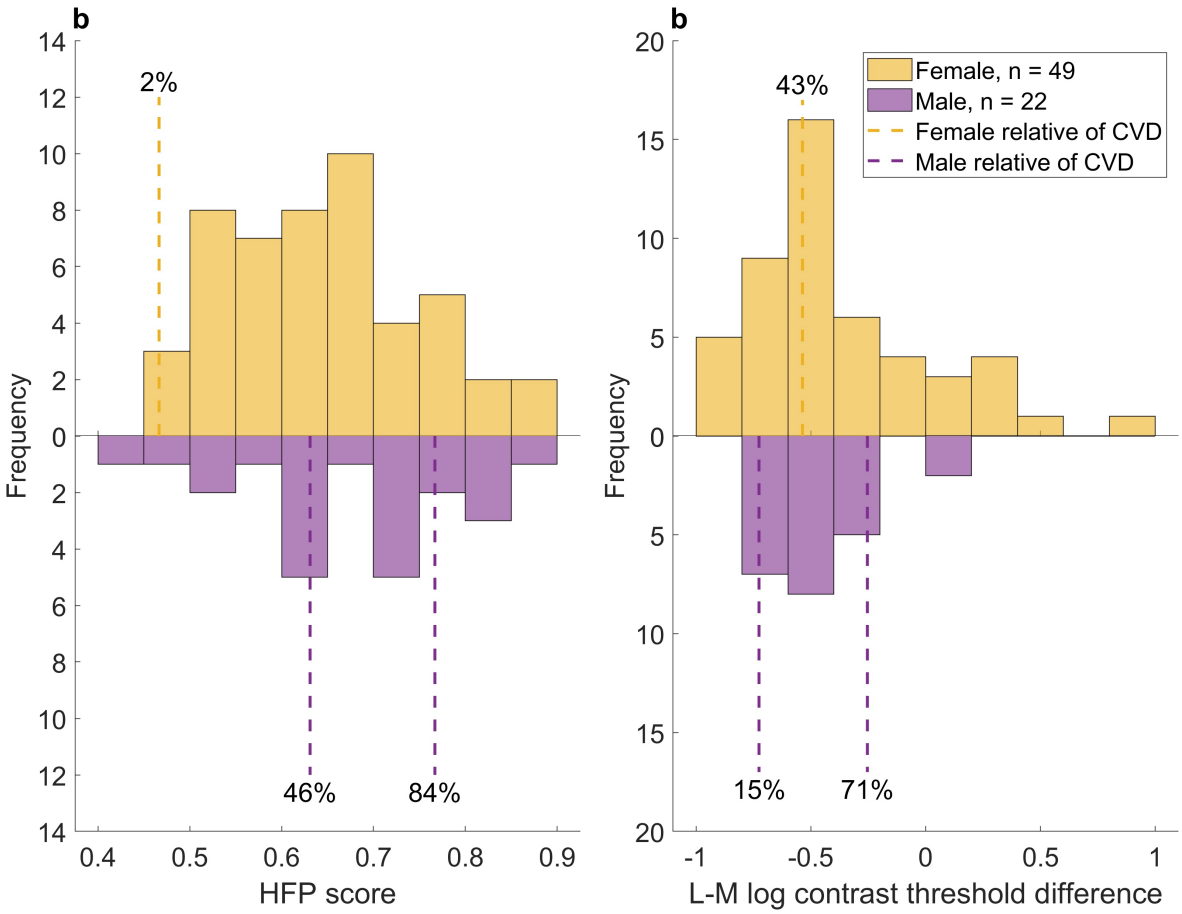


Figure 3.7: Histograms for HFP score (a) and L−M log contrast threshold difference (b) for female (yellow bars) and male (purple bars) participants. Participants reporting a first-degree relative with red-green CVD are marked with dashed lines and their percentile within the ordered dataset.

Sample sizes were too low to confidently calculate kurtosis.

3.4.7 Ethnicity

To test whether observers of Black and/or Asian descent have higher L:M ratios than white observers (Kuchenbecker et al., 2014; McMahon et al., 2008; Yamauchi et al., 2013), ethnicity was included on the study questionnaire. Of the 67 observers, 57 were white, four were from mixed/multiple ethnic groups, nine were Asian/Asian British, none were Black/Black British and one was from an ethnic group not listed on the questionnaire; as such, only the Asian participants were compared with white observers.

Fig 3.8 shows a pair of histograms of the frequency distributions of HFP scores (panel a) and log contrast threshold difference (panel b) for Asian/Asian British (blue bars) and white (red bars) observers. The vertical dashed lines show the median settings. For HFP scores, the Asian/Asian British sample showed a higher median HFP score (0.700) than the white sample (0.645), indicating a lower median L:M ratio. However, a two-sample one-way t-test indicated that the difference between Asian/Asian British and white observers was not significant ($t(64) = 1.381$, $p = .086$).

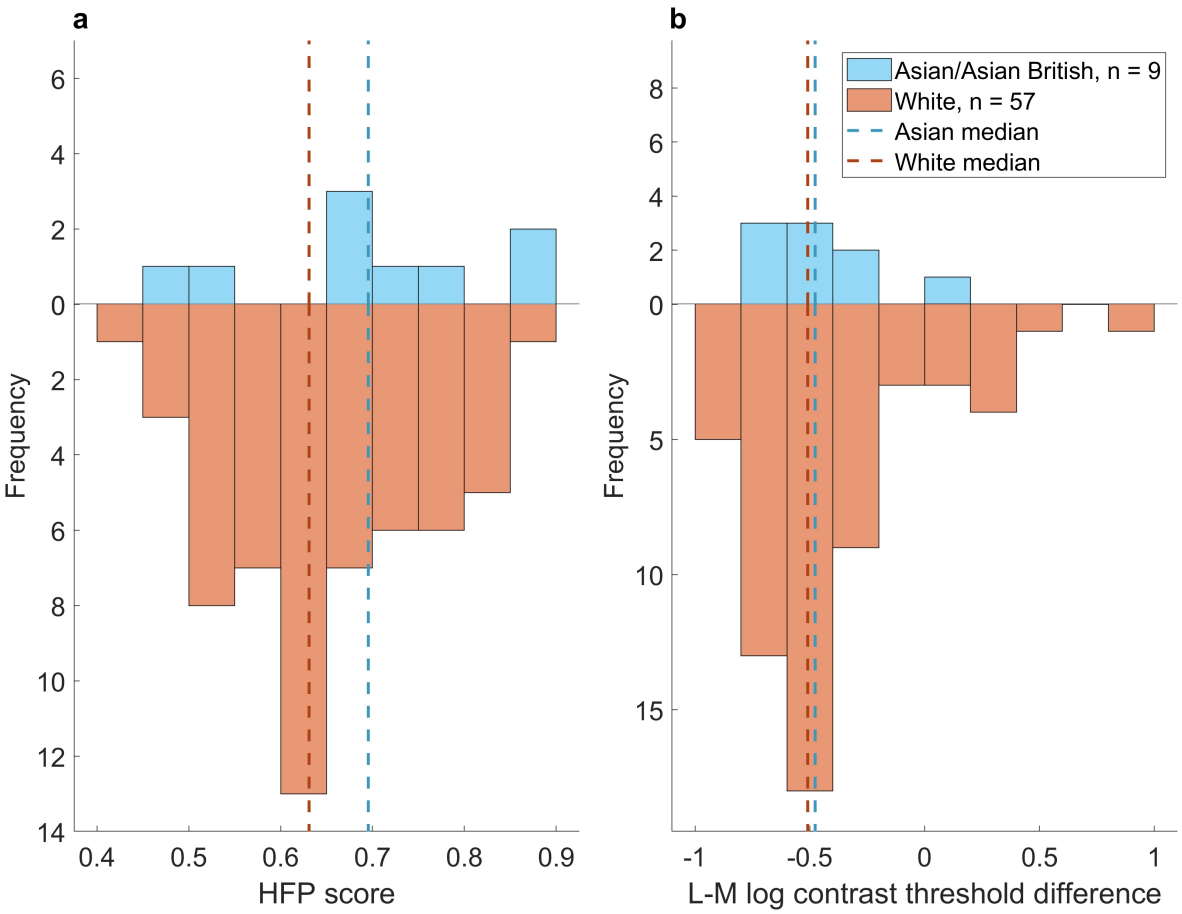


Figure 3.8: Histograms for HFP score (a) and L−M log contrast threshold difference (b) for Asian/Asian British (blue bars) and White (red bars) participants. Dashed lines show the population median.

For log contrast threshold difference, the white sample was not normally distributed (Shapiro-Wilk test, p = .0004), so a one-way Mann-Whitney U-test was carried out. Similarly to the HFP scores, the Asian/Asian British contrast threshold difference sample showed a higher median L−M log contrast threshold difference (-0.477) than the white sample (-0.501), corresponding to a lower L:M ratio. However, the difference was not significant (U = 250, p = .455).

3.5 Discussion

3.5.1 Does L:M ratio affect foveal spatiochromatic vision?

To determine whether L:M ratio affects foveal spatiochromatic vision, we measured L:M ratio using heterochromatic flicker photometry (HFP) and assessed its effect on foveal spatiochromatic vision using a modified version of the L:M Landolt C test (Danilova et al., 2013). For the latter task, log thresholds for L- and M-isolating, local optotypes were measured and the difference between the two calculated; the difference of the log thresholds can also be expressed as the log of the ratio of the linear thresholds (see Eq 3.1). As can be seen from Fig 3.6.a, L−M log contrast threshold difference and HFP score were highly significantly, weakly, positively correlated; therefore, L:M ratio

can indeed weakly affect foveal spatiochromatic vision. The sign of the correlation was as expected, indicating that when L:M ratios are high (low HFP scores) and there are many more L cones than M cones, lower L contrast is needed to see L-isolating local stimuli than M contrast for M stimuli. The same pattern can be seen for L−M log size threshold difference and HFP score (although the correlation was slightly stronger; see Fig 3.6.c), showing that individuals with more – and therefore more closely packed – L cones can see smaller L than M stimuli when contrast for both is fixed to the same value.

3.5.2 Are thresholds for the square C optotype test primarily determined by opsin peak rather than L:M ratio?

One polymorphism in the opsin genes is particularly common: at codon 180 of the L opsin gene, 62% of the human population have the amino acid serine, and 38%, alanine (Winderickx, Lindsey et al., 1992), producing a shift in the L-cone peak spectral sensitivity of around 5 nm (Merbs & Nathans, 1992). This impacts the luminous efficiency function, affecting HFP scores and their resulting L:M ratio estimates (Bieber et al., 1998), and is also likely to be correlated with the L and M resolution differences derived from optotype tests (Danilova et al., 2013). As genetic data were not collected in the present study, L opsin peak could not be controlled for in either the HFP or the spatiochromatic threshold measurements. Is it possible that the largest factor in the difference between L and M thresholds is L opsin peak? Danilova et al. (2013) argued that their parafoveal spatiochromatic thresholds were primarily determined by L:M cone ratio rather than opsin peak because L thresholds were lower than M thresholds. In the present study, L thresholds for both size and contrast were also on average lower than for M thresholds, which would be expected if some of the difference were caused by L:M ratio.

If the L opsin peak determined spatiochromatic L and M thresholds, it would be expected that L opsin deviations from the standard observer in either direction would elevate L thresholds by effectively reducing the L contrast compared to that calculated, and either have little effect on M thresholds or, in the case of an L opsin shift towards shorter wavelengths, lower M thresholds. This would lead to an elevation in the L−M threshold difference. If L:M ratio had no impact on L−M resolution difference but L opsin did, an L−M difference of zero or greater than zero would be expected. That L thresholds are lower than M thresholds points to an origin in L:M ratio.

3.5.3 Foveal and parafoveal ratios

In the previous parafoveal study, the mean non-log M:L contrast ratio for 'small targets' (presumably meaning the smallest used in that part of the study, 1 deg) for a colour-normal group of 10 observers is 1.6 (Danilova et al., 2013). In the present foveal study, excluding the subject for whom an L:M estimate could not be made, the equivalent mean metric for a 2.8-arcmin square C optotype is 1.60. For the size ratio, the parafoveal M:L spatial resolution ratio from Danilova et al., gathered at a fixed contrast of 0.25, was 2.08 for six observers. The foveal fixed contrast of 0.21 was as close as possible to this parafoveal contrast, and the foveal M:L size ratio was 1.26. M:L threshold ratio is expected to be positively correlated with L:M cone ratio, meaning that values from the two metrics can be directly compared with one another. The mean M:L threshold ratios given here, and summarised in Table 3.1, are in the range of the previously published mean L:M cone ratio of 1.86:1 (Carroll et al., 2002).

Importantly, if cone-isolating spatial resolution is determined by a spatial factor such as the mean distance to the nearest cone of like type, then the more relevant metric may be class-specific, centre-to-centre mean foveal cone spacing. This can be calculated for an observer by taking a published estimate of foveal cone density (168,162mm cones/mm^2; Zhang, Godara, Blanco, Griffin, Wang, Curcio & Zhang, 2015), multiplying it by the fraction of cones expected to be either L or M based on the observer's L:M cone ratio, and converting it to the mean centre-to-centre cone spacing for each class using the density-to-spacing equation given by Zhang et al. (2015). The equivalent M:L cone spacing ratio for the mean L:M cone ratio obtained by Carroll et al. (2002) is 1.36:1, which is also in the range of the above spatial values. Additionally, the mean estimated L:M ratio in the present study (excluding the subject for whom it could not be estimated) was 2.28:1, and the associated mean M:L class-specific, centre-to-centre foveal cone spacing was 1.51:1 (see Table 3.2).

Table 3.1: M:L spatiochromatic linear threshold ratios for C-shaped optotypes.

Study	**Eccentricity**	**Size thresholds**		**Contrast thresholds**	
		Contrast	**M:L ratio**	**Gap size**	**M:L ratio**
Danilova et al. (2013)	5 deg	0.25	2.08:1	1-deg	1.6:1
Present study	Foveal	0.21	1.26:1	2.8-arcmin	1.60:1

Here, four possible relationships between L:M cone ratio and size threshold differences will be described, considering those variables in terms of the equivalent linear M:L mean foveal cone-

Table 3.2: L:M cone ratio and M:L mean cone-to-cone spacing ratio estimates derived from flicker photometry.

Study	Eccentricity	Method	L:M ratio	M:L spacing ratio
Carroll et al. (2002)	Whole retina	ERG and genetics	1.86:1	1.36:1
Present study	Foveal	HFP	2.28:1	1.51:1

to-cone spacing ratio and the linear M:L size threshold ratios. If cone spacing ratio alone linearly determines any given psychophysical measure with a one-to-one mapping, the underlying relationship between the two variables should be representable by a line of unity (grey dashed line, Fig 3.9.a and b). If cone spacing ratio has no effect on that measure, the relationship should be represented by a horizontal line (black dashed line).

Figure 3.9: Illustrations of four possible relationships between linear M:L mean cone-to-cone spacing ratio and a given psychophysical measure. 1. Zero relationship (black dashed line). 2. Unity relationship (grey dashed line). 3. Compensatory mechanisms overcome any effect of cone spacing ratio at most levels except extremely asymmetric ones (blue dashed line and crosses, panel a). 4. Linear relationship with a shallower slope than unity, perhaps due to compensatory mechanisms (green arrows) that act equally at all cone spacing ratios (blue dashed line and crosses, panel b). The final blue cross of each panel, marked with a blue box, represents the first point at which the psychophysical measures at this level could be distinguished from those of observers with more even ratios.

In fact, prior work has found an effect on red-green contrast detection thresholds only in the case of considerably skewed L:M ratios (Hood et al., 2006), and an effect of L:M ratio on chromatic grating acuity only at the most asymmetric of ratios (Miyahara et al., 1998). This suggests that some additional factors impact spatiochromatic thresholds, and that in the present study we may instead expect to find that only observers with particularly asymmetrical cone spacing ratios show size or contrast threshold ratios that differ from those of participants with more balanced cone spacing ratios. In practice, those with very asymmetrical ratios will probably be those with high M:L cone spacing ratios (high L:M ratios). The final blue cross in each panel of Fig 3.9, marked out with a blue box, represents such observers.

It may be that these additional processes (e.g. gain control or eye movements, marked with green arrows in Fig 3.9.a) are wholly effective in counteracting the effect of cone spacing ratio on spatiochromatic threshold ratio up to a certain degree of cone spacing ratio asymmetry, at which they break down (final blue cross, in blue box) and a difference can be seen between those psychophysical values and those seen in observers with more balanced ratios. This would produce a relationship between cone ratio and spatiochromatic threshold ratio that was flat up to a certain point, and then rapidly became positive (see Fig 3.9.a). Alternatively, their effectiveness may gradually decrease as the observer's cone ratio deviates further from unity (blue dashed line, Fig 3.9.b), up to a point at which psychophysical measurements show a difference between these

asymmetric-ratio individuals and those with more balanced ratios (final blue cross, in blue box).

However, both of these predictions are rather different from the shape of the size ratio-cone spacing ratio relationship shown by the present dataset (Fig 3.10). Many of the central values do in fact cluster along the line of unity (grey dashed line), and the systematic deviations from the line are in directions consistent with prior predictions of pressures from instrument limitations (purple dashed line, purple arrow) and optical factors in the eye (yellow arrow). The foveal data from the present study are therefore consistent with the hypothesis that sampling density is linearly related to L and M size thresholds up to an observer's optical acuity.

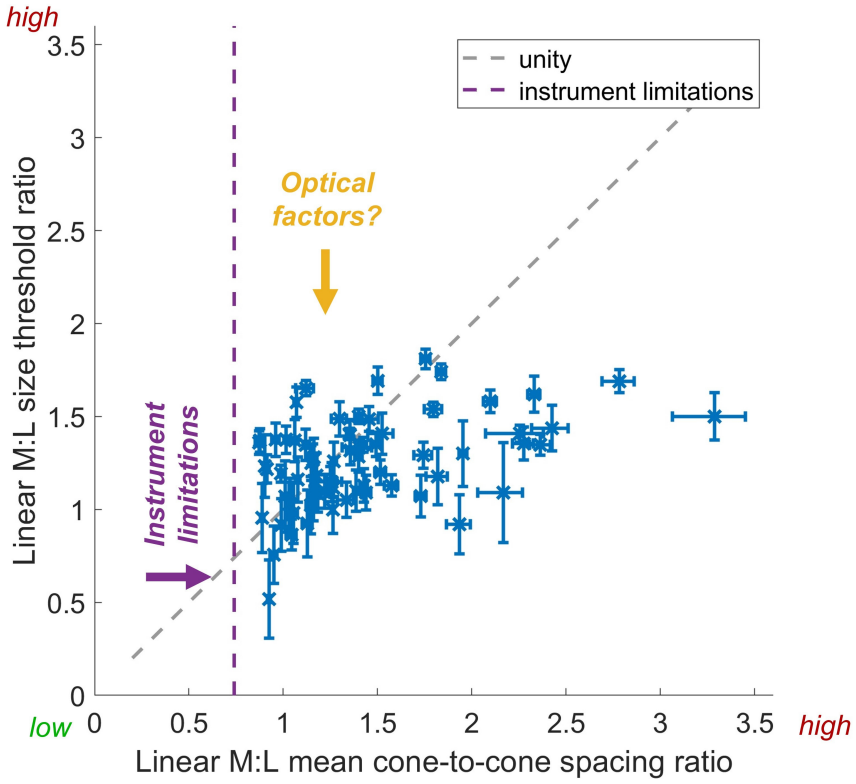


Figure 3.10: Panel (a): Linear M:L size threshold ratio plotted against estimated linear M:L mean foveal centre-to-centre cone spacing ratio. The grey dashed line represents the line of unity. As in Figs 3.2, 3.5 and 3.6, purple dashed lines represents the lowest setting available on the HFP apparatus, purple arrows denote the HFP instrumental limitations and yellow arrows indicate the assumed optical limitations.

3.5.4 Correlation strength

Why were the foveal L−M log threshold differences for size and contrast only moderately and weakly, respectively, correlated with HFP score? Firstly, there was a known instrumental limitation of the HFP apparatus: during its design and calibration, a choice had to be made as to whether to prioritise separating those with high ratios or with low ratios. High ratios were chosen, and as a result many of those with low ratios reached the high end of the range during the method of adjustment, compressing their HFP scores downwards and the resulting L:M ratio estimates

upwards (shown by purple arrows in Fig 3.2.a and b respectively). No adjustment was made to these participants' HFP scores and L:M ratio estimates as it was not possible to estimate how much these individuals' scores might have been compressed.

Gain control, believed to reduce some of the impact of L:M ratio on chromatic thresholds (Kremers et al., 2000), would likely reduce the difference between thresholds for L compared to M, thereby compressing the L−M log size and contrast threshold differences towards zero (shown by yellow arrows in the histograms in Fig 3.5.a and b). If these measurements were entirely without noise then gain ought not to influence the coefficient of the slope of the relationship – that is, the correlation strength – as the non-parametric Spearman's rank correlation was used and should not be affected by the shape of the relationship. However, in practice, the compression combined with noise would result in some reordering of the data, reducing the coefficients.

Additionally, there is likely to be some upward pressure of optical acuity on both size and contrast thresholds that would have the effect of compressing an individual's L−M log threshold difference towards zero: if thresholds for a target were affected by optical factors in addition to the sampling density of the relevant cone class, thresholds would be elevated in a way that would likely be more marked for the cone with the lowest thresholds – that is, the higher sampling density. This would reduce any difference between the thresholds obtained for the two cone classes, represented by the same yellow arrows.

If cone class density alone determined size and contrast thresholds, the L−M log threshold difference histograms would be expected to have a long negative tail; however, the tail is positive. This can be explained by the fact that optical factors (and perhaps other factors which act to overcome L:M cone ratio) would act most strongly at the most imbalanced – that is, in practice, the highest – cone ratios. First, observers who have cone ratios greater than 1:1 are expected to be greater in number than those expected to have lower cone ratios, and also to have ratios that deviate more from 1:1. Second, cone ratios (when expressed in terms of the dominant cone) equally different from 1:1 should be equally affected by acuity. Therefore, the observers with higher-than-unity L:M ratios will have an average L:M ratio that deviates more from 1:1 than observers with a lower-than-unity L:M ratio, so their ratios are more likely to be compressed, and compressed further, than those with lower-than-unity L:M ratios. Together, these features will turn the expected distribution of threshold differences from one with a long negative tail to one that may even have a positive tail, with the large number of individuals with an above-1:1 L:M ratio and a below-zero L−M threshold difference compressed up into a peak below zero, and the smaller number of the below-1:1 L:M ratio individuals either uncompressed or compressed into the same or a similar

range above zero as the range allowed by the compression factors below zero.

The three pressures identified above would explain the unusual shape of the scatter plots in Fig 3.6.a-c. Taking the plot of contrast threshold difference vs HFP score (a) as an example, with instrumental limitations pressing higher HFP scores leftwards (purple arrow) and gain/optical factors putatively shifting lower L−M log contrast threshold differences upwards (yellow arrow), a somewhat triangular shape is formed.

Why was the correlation coefficient for L−M log size threshold difference and HFP score stronger than that of L−M log contrast threshold difference and HFP score? Owing to display resolution limitations, size thresholds were gathered with relatively large, fixed step sizes: the gap sizes generally ranged from 1 to 6 pixels in 1-pixel steps. Contrast staircases, meanwhile, could vary on a scale much finer relative to perceivable changes. For this reason, size thresholds were gathered merely to confirm that the participants had gap size thresholds lower than around 3.5 pixels to enable them to perform contrast staircases at a gap size of four pixels without hitting the ceiling of contrasts that could be displayed.

A 4-pixel gap size equates to 2.8 arcmin, or 11 cycles per degree (cpd). At this point, the grating-measured L- and M-isolating contrast sensitivity functions are either just starting to descend from their peak (Green, 1968) or are already fairly steeply descending (Cavonius & Estévez, 1975). More relevantly, if we can extrapolate from spatially local parafoveal contrast sensitivity functions (Danilova et al., 2013) to 11 cpd, the curves should be fairly steep. Assuming a steep contrast sensitivity function, when contrast is fixed and size is varied, a change in size should make a big difference to the apparent contrast. This should make the error bars on the size thresholds relatively small. Conversely, when size is fixed and contrast is varied, change in contrast should make a small difference to the apparent spatial frequency of the target (whether it looks like a blob or whether more detail structure appears), resulting in relatively large error bars on the contrast thresholds. This comparatively greater uncertainty around the contrast thresholds compared to size thresholds should reduce the strength of its correlation with further variables.

3.5.5 Gender and family history of red-green colour vision deficiency

The experiment in Chapter 5 required observers with extreme L:M ratios; therefore, the HFP scores demarcating the lowest and highest quartiles from the central 50% of the normative HFP dataset from the present study were determined so that those with HFP scores below and above those limits could be invited to return for the subsequent experiment. In addition to recruiting from the current pool of participants, it was intended to selectively recruit observers believed likely to have

extreme L:M ratios and to compare their HFP scores with those limits to confirm their eligibility. Those observers to be targeted were female relatives of those with red-green colour blindness. As described in § 3.1.1, the 15% of women believed to be genetic carriers of red-green colour blindness (Hood et al., 2006) tend to have more extreme L:M ratios due to half of their non-S cones containing only M or M', or only L or L', opsin (Jordan & Mollon, 1997). To verify this approach, we gathered participants' gender, a correlate of sex, and whether they knew of any relatives with CVD. Unfortunately, only one of the 44 female participants who correctly completed the study reported a first- or second- relative with a CVD, so the intention could not be verified. The 2% of female participants who claimed a close relative with CVD was a far cry from the percentage expected to carry a gene for red-green CVD. There may simply be a lower number of female relatives of those with CVD in the sample, or the low number may be due to the fact that of those with a milder form, many are likely to be unaware of it, and their relatives in turn are also less likely to be aware of it or to remember when filling out the questionnaire. It was expected that the distribution of female HFP settings would be broader than the male HFP distribution as female carriers of a red-green CVD would be more likely to have HFP settings in the tails of the distribution; however, sample numbers were too low to test this by calculating kurtosis scores for each of the distributions. Finally, the large disparity between the percentiles for HFP and contrast-based L:M spatial resolution ratio, which might not be predicted given that HFP and L:M spatial resolution ratio were correlated, is nonetheless explainable by the noise that is apparent in both datasets in the errorbars seen in Fig 3.6.a and the weakness of the correlation between the two.

3.5.6 Ethnicity

Reports have pointed to lower L:M ratios in Black and Asian ethnic groups (McMahon et al., 2008; Yamauchi et al., 2013). To that end, ethnicity was included on the study questionnaire. There were no Black/Black British participants in the final dataset, perhaps due to local demographics. The Asian/Asian British sample showed higher HFP scores and L−M log contrast threshold differences, both corresponding to lower L:M ratios as expected; however, neither difference reached statistical significance. The small size of the Asian sample (n = 9) suggests that the test was underpowered for this analysis; if a larger sample had been obtained, the difference might have proven significant. It is also possible that a clearer picture of the two samples under discussion would have emerged if HFP scores had not been limited by the apparatus at the low-L:M end of the range of recording. The fact that this was a sample of opportunity calls for further targeted recruitment of Asian/Asian British observers up to a fixed sample size. As the L−M log threshold difference reflects a greater number

of factors than HFP, any difference between the two ethnic populations was not expected to come through as strongly. Due to the time-consuming nature of the spatial resolution measurements, the new Asian/Asian British sample should be confined to HFP.

3.5.7 Conclusion

The present study indicates that L:M ratio influences spatiochromatic thresholds in the fovea. While other work on contrast sensitivity functions (Miyahara et al., 1998) and red-green contrast detection thresholds (Hood et al., 2006) suggests that there should be an effect of L:M ratio on spatiochromatic threshold differences only at particularly imbalanced ratios (1:33.33 and 5:1 respectively), with no difference between more balanced ratios of 2:1 and 1:2, we found that the influence of L:M ratio was present at more balanced ratios but tailed off at higher, more imbalanced ratios, perhaps due to the influences of optical acuity or some mechanism that compensates for extreme ratios. Chapter 5 will explore whether eye movements help overcome imbalances of cone ratio in a high-acuity, foveal spatiochromatic task.

Chapter 4

Capturing retinal images and eye movements using the Oxford adaptive optics scanning laser ophthalmoscope

Declaration: The work contained in this chapter is entirely that of the author (SER). The adaptive optics scanning laser ophthalmoscope (AOSLO) and associated software were built by Young et al. (2018). 50% of the montages were made by ACH, RW and Samuel Ponting.

4.1 Introduction

The investigation of the relationship between fixational eye movements (FEM) and high-acuity colour vision described in Chapter 5 required an eye movement tracking system of extremely high spatial and temporal resolution. It therefore used the Oxford adaptive optics scanning laser ophthalmoscope (AOSLO) introduced in § 1.4.3, which collects video streams of an observer's cone photoreceptor mosaic. Because the AOSLO captures images by raster-scanning a point of light over the retina, if the retina is translated during the acquisition of a frame, distortions are introduced. The eye motion that caused the distortions can be recovered from them by comparing the captured frame to a reference frame. Increasing the retinal area included in the reference frame increases the range of eye positions that can be successfully extracted from the data.

Prior to an eye movement experiment using the AOSLO, raw frames are obtained of various overlapping areas of the retina to be used as a reference for frames collected during experimental trials. The frames for each area are automatically corrected for any internal motion against their own reference frames and are registered and aligned with each other, then averaged together to

create an image with a lower level of noise compared to individual raw frames. These images of overlapping areas are subsequently manually aligned with one another and averaged, resulting in a final montage that is larger than the area spanned by a single frame, against which the trial images can be referenced. While the observer performs the experiment, further video streams are collected of the retina spanning a period from shortly before the presentation of each stimulus to shortly afterwards. The trial frames are compared to the montage in order to extract the retinal positions during the trial relative to the montage. The resulting eye position data are then cleaned by rejecting any frames with quality metric values below a given cut-off point. These steps are explained in detail in this chapter.

4.2 Image acquisition

An AOSLO images the retina by operating as a confocal microscope in which the eye's lens and cornea act as the objective lens of the instrument. As with all scanning laser ophthalmoscopes (SLOs), the AOSLO raster-scans a point of light across the sample and measures the intensity of the reflected light at each scan position. The resulting data consist of a stream of pixel intensities and scan positions and their corresponding time points. The Oxford AOSLO raster scans at a frame rate of 30 Hz. The raster scan is performed by two scanning mirrors: the horizontally scanning mirror oscillates sinusoidally at a rate of 533 cycles per frame and the vertical mirror slowly angles downwards and then flicks up in a sawtooth pattern. While the eye position extraction algorithm produces data at 16 kHz, with one position estimate per pixel, the samples are not independent. The final data are downsampled to 1 kHz.

Because the lens and cornea serve as the objective lens, a vital part of the setup of the AOSLO for image acquisition is to align the participant's eye to the instrument. To maintain head stability, a dental impression is mounted at the end of the AOSLO on an adjustable stage that can be translated in three dimensions to align the eye. During the alignment steps, the participant looks at a fixation cross on the display at the intended fixation position for the experiment. The aim is to adjust the horizontal and vertical position of the eye until the imaging light passes through the pupil as centrally as possible and the axial position until the eye's pupil is correctly conjugated to the wavefront sensor and deformable mirror. To guide the setup, the instrument provides a live video stream of partially processed images and the image from the wavefront sensor indicating the pupil position. The position of the stage upon which the dental impression is mounted is first translated horizontally and vertically so that the scan patch appears as bright as possible to the observer, and then in all three dimensions of space so that the frames in the live video stream are as clear

as possible. The photodiode gain, AO system gain and exposure setting for the wavefront sensor may be adjusted to improve image quality. Subtle day-to-day differences in the alignment of the AOSLO may degrade image quality necessitating much more careful optimisation of the setup, and may compound other factors, described below, that affect image quality. If the fixation cross or stimulus is moved around the screen after the initial alignment, it may be necessary to further adjust the horizontal and vertical translation of the dental impression in order to keep the imaging light centred on the pupil and therefore keep the image intensity high.

If the observer does not have normal acuity, where possible their own glasses or contact lenses should be used to correct the level and type of low-order aberrations in the eye (defocus and astigmatism); if their own correction is insufficient, it may be improved upon by mounting one or more trial lenses in front of the eye. However, glasses, contact lenses and trial lenses all attenuate the intensity of the imaging light and its reflection from the retina. Glasses and trial lenses can cause reflections that appear as artefacts in the image and may affect the AO correction, and they can also move over time during the imaging session, further affecting the clarity of the images. Difficulty correcting an observer's level and type of aberrations, particularly in the case of strong aberrations, can prevent the acquisition of good or sufficient images. An individual may additionally have a less reflective retina or a smaller pupil size than other participants, either making it impossible to obtain images or necessitating particularly frequent measures to elevate or maintain the amount of light entering and exiting the eye.

Image quality is likely to degrade throughout the imaging session, more quickly for some participants than others. One notable tension during the session is the level of ambient illumination. To obtain the brightest images possible, it is desirable for the participant's pupil to be large to allow the maximum amount of light in and out; since dilation drops would require the presence of a clinician and would negatively affect simultaneous behavioural experiments, the best way to achieve natural dilation is to image in a dimly lit or dark room. However, the longer the participant is in the dark, the more the opsin in their photoreceptors returns to an absorbent state and the less reflective the cells appear in the images. Turning on the lights for a few minutes or asking the observer to move to a more brightly lit space usually increases the brightness of the frames, likely because the opsin then absorbs light, reducing the number of non-activated photopigment molecules, and the cones become more reflective. Other techniques likely to improve lost quality include asking the participant to blink quickly (to restore the tear film) or to look away from the screen to an object at a different depth and back again (to refresh accommodation). A few minutes' break of any kind also tends to improve quality, but particularly if the observer is engaged in

conversation, perhaps because it alleviates boredom and thus restores concentration.

In order to image different locations on the retina, the paradigm makes use of the way the eye moves to bring stimuli to bear on the usual retinal location at which a target of interest is centred, or the preferred retinal locus (PRL). The fixation cross is translated across the screen and when the eye moves to relocate the cross onto the PRL, the imaging light falls on a different patch of the retina. The effect of asking the participant to look at different locations relative to the scan patch is shown in Fig 4.1. In the top half of the figure, the participant fixates on a cross centred on the scan patch. Consequently, the light source scans an area centred on the observer's PRL. Moving the fixation cross 1 deg above the centre of the scan patch would lead to the acquisition of an image centred 1 deg above the PRL when looking at the retina through the pupil, while presenting the cross 3 deg to the left of the scan patch would obtain an image 3 deg to the right of the PRL. The lower half of the figure depicts what happens if the cross is translated by (-3, 1) deg: the resulting image is centred on a point (3, 1) deg from the PRL.

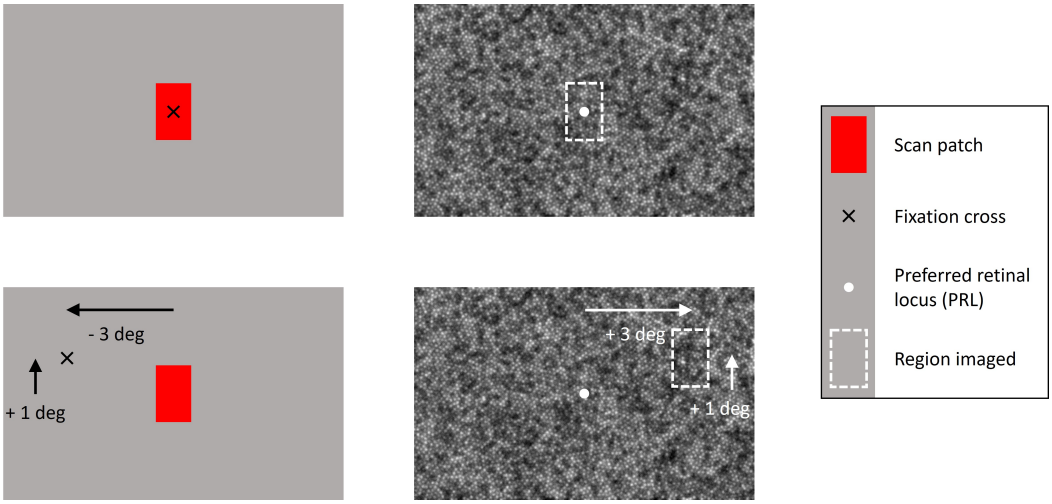


Figure 4.1: Effect of fixation relative to the AOSLO scan patch on the area of retina imaged. The left panels represent the display (grey box) with a fixation cross (black cross). When the observer is properly aligned, the scan patch will be visible against the screen (red rectangle). The right panels show the area of retina captured (white box) relative to the preferred retinal locus (PRL; white dot). In the top pair of panels, the fixation cross is displayed in the centre of the scan patch, resulting in an image centred on the PRL. In the lower pair of panels, the fixation cross is displayed 3 deg to the left of the centre of the scan patch, resulting in an image whose centre is 3 deg to the right of the PRL when looking at the retina through the pupil.

4.3 Eye motion estimation

As described in § 4.2, the AOSLO raster-scans a point of light across the retina to yield a set of pixel intensities, scan positions and time points. These are processed into raw image frames using custom-written Python software. It is important to stress that the pixels in a raw frame are not gathered simultaneously, but one at a time while the imaging light is scanned across the sample. If

the sample moves halfway through a scan, the upper and lower halves of the image frame will not be perfectly aligned. Because the eye is constantly in motion even during fixation, eye motion distorts the raw frames and must therefore be estimated and removed. There are many ways to achieve this, but the general approach is common across labs: a reference frame is identified from a set of frames and small patches from each frame are registered and aligned to that reference; if required, the signal-to-noise level in a final image can be increased by averaging together multiple frames. To begin the specific pipeline for the Oxford AOSLO, the 'best' frame of a given video stream to be used as the reference frame is automatically identified using cross correlation to determine the frame that is most similar to all the other frames in the set. All other frames are registered to the reference frame by applying the Farneback optical flow algorithm (Farneback, 2003), thus removing motion distortions in a given frame and gross motion between frames. To improve the signal-to-noise ratio of the final image, frames whose motion falls below a fixed threshold are averaged together to produce a higher quality image.

When frames are motion-corrected in this way by registering them to the best reference frame, the image processing algorithm also returns the estimated eye position relative to the reference frame for each time point. Eye position is expressed in terms of visual angle; it was not possible to convert these to units of distance such as mm on the retinal surface because this would have required a measurement of the axial length of the eye, which requires access to instrumentation not currently routinely available to the lab.

4.4 Reference montage generation

If the eye moves such that little or none of the image overlaps with the reference frame, the algorithm will not be able to return an accurate relative eye position. In order to measure large excursions accurately, eye position is instead estimated by comparing frames to a larger reference image covering 3 x 3 deg (see Figure 4.2, white box) rather than the 1 x 1.6 deg of an AOSLO frame (see Figure 4.2, red and green boxes). To generate this larger reference frame, multiple images from adjacent locations must be assembled into a montage. Significant overlap between images is required in order to compare large enough overlapping parts of the images to register them to one another correctly. Therefore, a set of images was collected by collecting 60-frame (2-sec) streams while the fixation cross was presented at intervals of 0.25 deg from -4.7 to -2.7 deg horizontally from the centre of the scan patch and at intervals of 0.35 deg from -0.7 to 0.7 deg vertically. This provided horizontal and vertical overlaps with the immediately adjacent images of 75% and 78% respectively. It resulted in retinal images centred on positions from (2.7, -0.7) to (4.7, 0.7) deg covering a 3 x 3 deg

area centred on (3.7, 0.0) deg from the PRL. Each image contributing to the montage was produced by averaging the frames following registration to the best reference frame, as described in § 4.3.

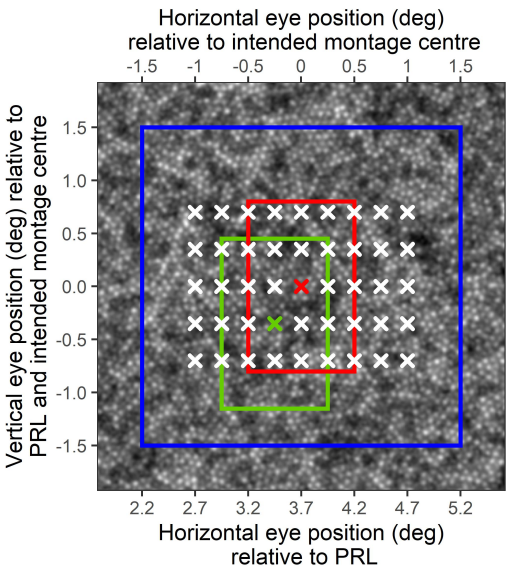


Figure 4.2: Spacing of images for the retinal image montage. The red cross shows the position of the intended centre of the montage, (3.7, 0.0) deg relative to the preferred retinal locus (PRL), and the red box denotes the edges of the corresponding 1 x 1.6 deg image. The green cross and box indicate the centre and borders of an image obtained (-0.75, -0.35) deg from intended centre of the montage, or (2.95, -0.35) deg from the PRL. The white crosses show the centres of the 45 images obtained for the montage, spaced horizontally relative to the PRL from 2.7 to 4.7 deg at intervals of 0.25 deg, and -0.7 to 0.7 deg vertically at intervals of 0.35 deg. The blue box shows the intended 3.0 x 3.0-deg boundaries of the resulting montage.

As shown in Fig 4.3, the images collected for the montage were laid out in GNU Image Manipulation Program (GIMP) (The GIMP Team, 2018). They were compared by varying the transparency of overlapping images to enable multiple images to be seen simultaneously at a given location. Their correct positions were determined by beginning with the central frame and working outwards, adjusting the translation and rotation of each added frame until the cones appeared as clear as possible in the overlapped image. Images were cropped or discarded when the quality was poor, for example due to excessive motion, residual blur or noise. Once the arrangement had been determined, the frames were exported to MATLAB and averaged to form a montage. The resulting averaged montage was used to inform further alterations to frame positions in GIMP, and the process repeated iteratively until a satisfactory montage was obtained.

Once the montage was completed, eye positions during trials could be analysed using the same program described in § 4.3, but with the montage replacing the best frame as the reference image.

4.5 Axes of eye position space

The eye position space as returned by the post-processing algorithm has two dimensions, the x axis representing horizontal eye position in terms of degrees of visual angle and the y axis, the

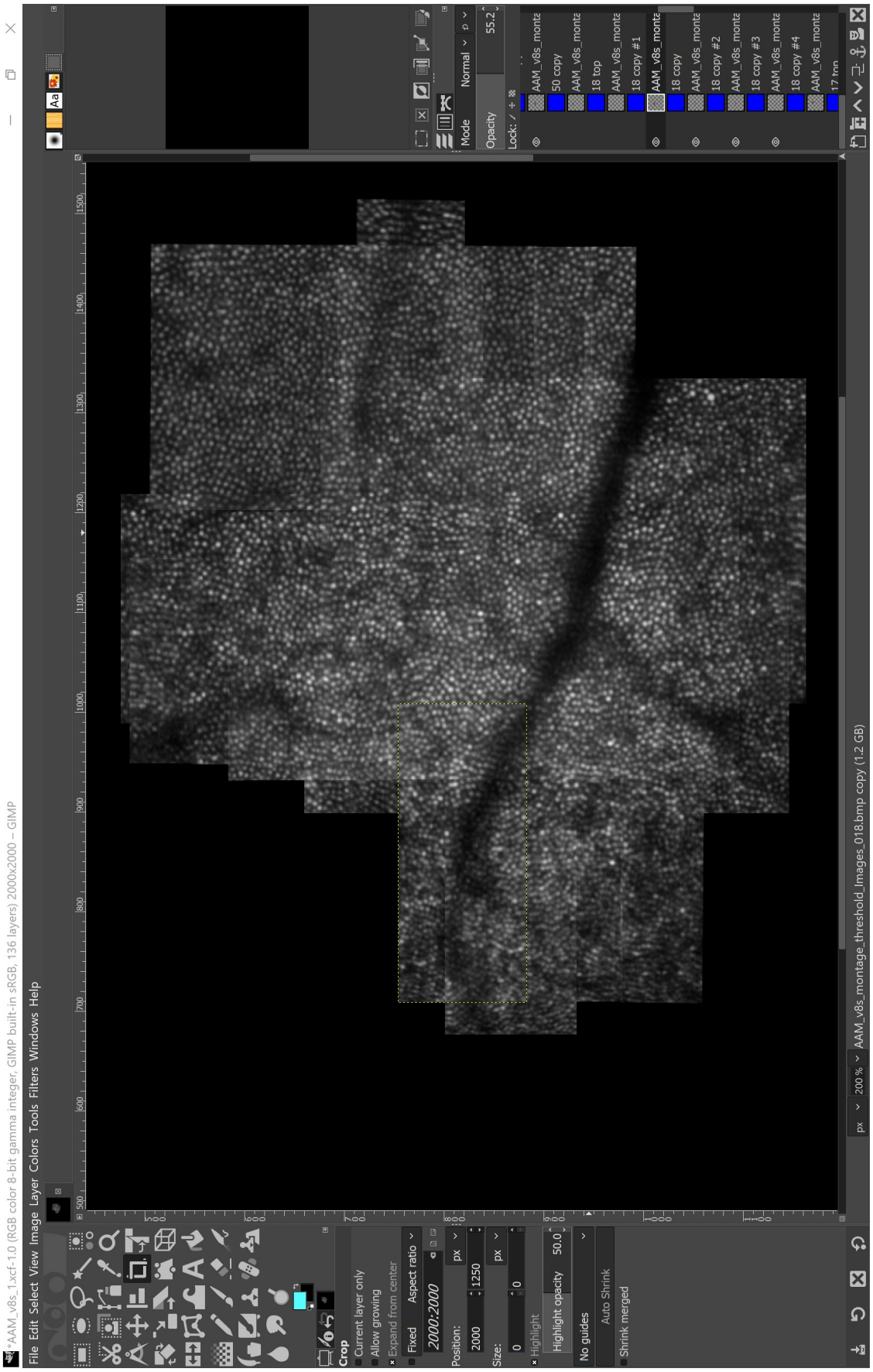


Figure 4.3: Set of central montage images for participant AAM laid out with partial transparency in GNU Image Manipulation Program (GIMP).

vertical position. However, the origin of the space, (0, 0) deg, does not refer to a physiologically or behaviourally relevant location: it merely denotes the location of the imaged area of the observer's retina relative to the centre of the montaged retinal image. Because the fixation cross is presented (-3.7, 0.0) deg from the centre of the scan patch, this will be nearby the location (3.7, 0.0) deg from the PRL. However, it will not exactly coincide with the location primarily because fixation is imperfect and individual montage frames will therefore not be centred on their intended locations; by extension, when the final montage is cropped so as to remove any blank space on any of the four edges, but so as to maintain as much area as possible, the centre of the montage is unlikely to be exactly where intended (in this case, (3.7, 0.0) deg). This is illustrated in Fig 4.4 which shows an example montage. The red cross denotes the centre of the image obtained by presenting the fixation cross (-3.7, 0.0) deg from the centre of the scan patch, and the blue box shows the expected boundaries of a perfect 3 × 3 deg montage. However, due to variability in fixation, some of the images at the edges of the montage do not reach the boundary; others stray beyond it. Each final montage is cropped to a rectangle, as is required by the eye position extraction algorithm, so that as much of the montage as possible is preserved. The boundaries (grey box) may therefore be asymmetrical around the intended centre of the montage, and the actual centre of the montage (grey cross) may be displaced away from the intended centre.

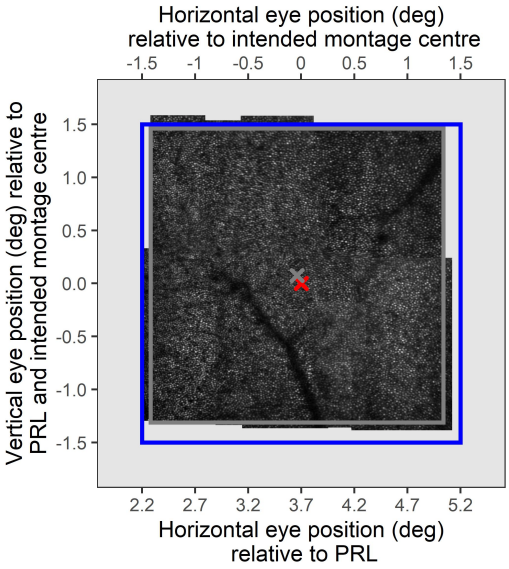


Figure 4.4: Intended versus actual location of a retinal image montage. The red cross denotes the centre of the image taken (3.7, 0.0) deg from the preferred retinal locus (PRL), which we have taken to be the intended montage centre and the actual position (3.7, 0.0) deg from the PRL; the blue box shows the expected boundaries of a 3.0 x 3.0-deg montage. The grey box shows the boundaries for the obtained montage when as little is cropped away as possible to obtain a rectangular image; the grey cross shows the centre of the cropped montage, (-0.04, 0.07) deg from the intended centre.

It may be desirable to set the origin of the eye position space to a behaviourally relevant location, such as the mean/median location on the retina when the participant is fixating a particular

stimulus – that is, the PRL. Note that this does not have a precise physiological meaning that is consistent between observers as the PRL does not necessarily coincide with the location of peak cone density (PCD), nor is it offset from the PCD in any systematic direction (Putnam et al., 2005; Wilk et al., 2017). For the experiment in Chapter 5, two methods were trialled for setting the origin to represent the position of the eye when centring the PRL on the experimental stimuli. The first method set the origin to the eye's position during recording of the central montage image, when the fixation cross appeared at the location on the display where the experimental stimuli were to be positioned. The second method set the origin relative to the eye position associated with fixation on the expected location of the stimulus prior to the experimental task.

The first method referenced the data stream for the central montage image to the final montage using the usual optic flow method. The median eye position during central image capture (Fig 4.5 (a), red cross, (x_{io}, y_{io}) provided the offset of the intended montage origin from the actual montage origin (grey cross). This offset could then be subtracted from all trial eye positions (Fig 4.5 (b)) so that all eye positions would be given relative to the median fixation position when looking at a cross centred on the centre of the trial stimuli. Anatomically, this origin would correspond to the location on the retina (3.7, 0.0) deg from the PRL, where the scan patch would appear if perfectly fixating a cross (-3.7, 0.0) deg relative to the centre of the scan patch on the display.

To reduce the level of noise in the origin recovery, it would possible to use more data from other frames in the montage. Multiple component montage frames of sufficient quality may be selected and referenced to the montage; the mean offsets between their actual and expected positions relative to the centre of the montage could then be applied to all trial eye positions. Fig 4.6.a shows an expanded view of Fig 4.5.a. Again, the red cross denotes the median eye position during the capture of the central image and the grey cross (hidden behind the red circle) denotes the actual montage origin. Additionally, the green cross represents the median eye position during the capture of a different location in the montage, and the red and green circles represent the expected locations of the eye during data capture had the actual and intended montage centre been the same. Fig 4.6.b shows a zoomed-out view of Fig 4.5.b after adjusting for the mean offsets between the two images' actual and expected positions.

Ultimately, there proved to be too much variability in the positions of the montage images to use this method, even when integrating the positions of the non-central montage images into the offset calculation. This may have been because the eye drifted away from the fixation cross during data capture. Instead, for the experiment in Chapter 5, the origin of the eye position space was calculated such that it represented the eye position prior to the onset of the experimental stimulus

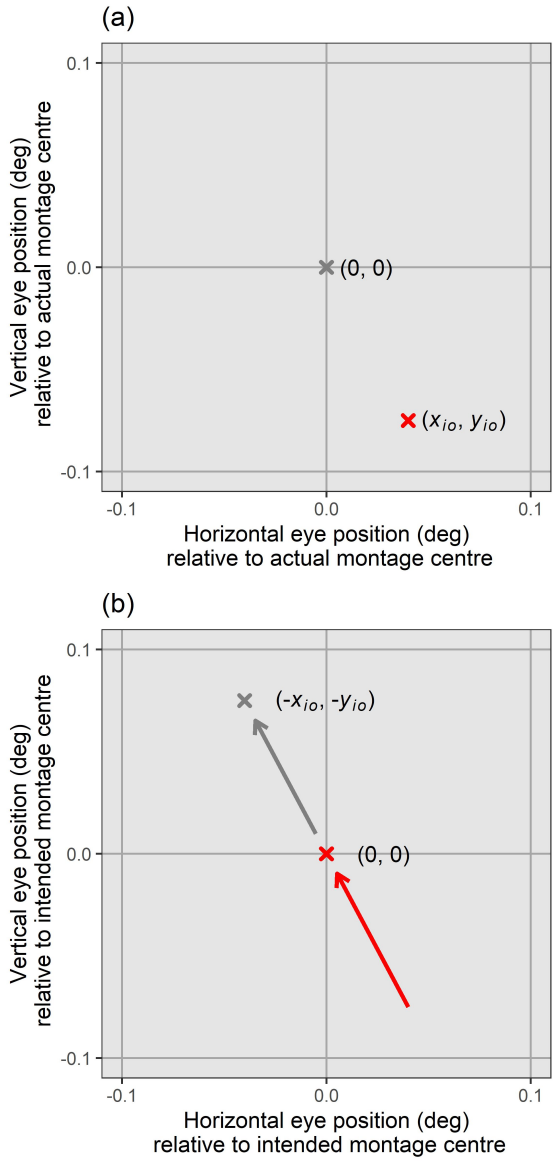


Figure 4.5: Adjustment of origin made to trial eye position output coordinates for a given observer. (a) Raw eye position, with the eye position corresponding to the centre of the montage given by a grey cross. The median position (x_{io}, y_{io}) of the central image's first 500 ms of data capture when referenced to the montage (red cross) gives the offset of the trial eye position coordinates relative to the intended montage origin, or (3.7, 0.0) deg from the preferred retinal locus (PRL). (b) Adjusted eye position. When the offset is subtracted from the trial eye movement coordinates, the intended origin (red cross) and raw origin (grey cross) are translated (-x_{io}, -y_{io}) so that the eye position is given relative to the intended montage centre.

when the observer is likely to be attempting to centre the PRL on the location where they expect the stimulus to appear. To achieve this, for each participant, the average eye position across all trials during the 100 ms immediately prior to stimulus onset was obtained and subtracted from the eye positions returned by the eye position extraction program. The use of all trials in the calculation of the offset incorporated eye positions from many more images than the first method, which used only the montage images.

Other methods of experimentally defining the PRL include identifying where the stimulus *lands* on the retina at the moment of stimulus onset, referred to as the 'functional PRL' (Gautier, Bowers,

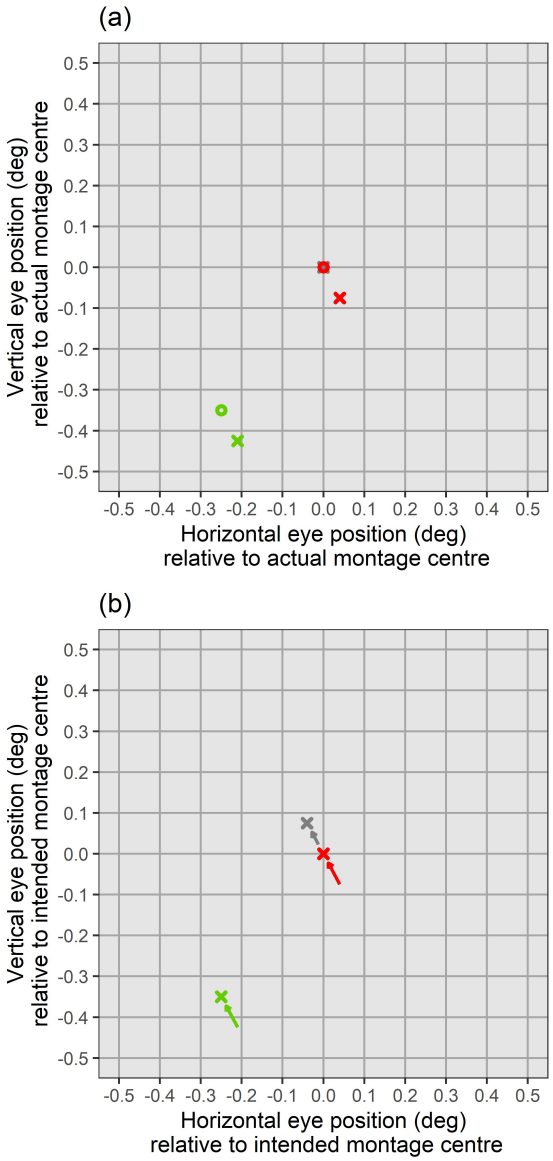


Figure 4.6: As Fig 4.5, except (1) the circles denote the *expected* image centre position for the cross of the corresponding colour, based on where the fixation cross was presented at the time of image capture; and (2) the green symbols represent the image centre position of a montage component image captured with a fixation cross presented (-3.95, -0.35) deg relative to the centre of the scan patch, with a corresponding retinal position of (3.45, 0.35) deg and an intended position relative to the centre of the montage of (-0.25, -0.35) deg. In (b), the mean difference of the actual vs expected positions of both the red and green image centres have been taken as the offset and used to adjust the eye positions.

Banks & Roorda, 2020); however, such an approach incorporates fewer data points and is therefore more affected by noise.

4.6 Eye position quality metric

4.6.1 Creation of an eye position quality gold standard

To reduce the amount of noise in the eye-tracking data, poor-quality estimates of eye position were removed. The eye movement extraction program estimates tracking quality by dividing each

motion-corrected frame into patches, and for each patch assesses its similarity to the corresponding patch in the reference montage, quantified by the Structural Similarity Index which has values between -1 and 1. This produces a single quality metric for each time point from the Farneback optical flow algorithm (Farneback, 2003), whereby poor estimates of eye position with metric values below a given cut-off can be excluded from further analysis. The appropriate cut-off had not been previously determined; a subset of trials were therefore selected from the present study and manually assessed by the author.

For each of the 18 participants who completed the eye movement study in Chapter 5, the lists of trials were broken down into ten blocks of length (total trials) / 10 and two trials were selected at random from each block. This ensured that eye movements recorded in the earlier and later parts of each session, and from different sessions, were all evaluated. Traces were presented in a random order as plots in MATLAB 2019b and sections of the trace with serious 'jitter', where vertical and horizontal positions jumped implausibly from frame to frame, were marked as tracking failures using the function *ginput*. Assessment took place for the whole length of each trace, but quality metric analysis was limited to the period during which the stimulus was present.

4.6.2 Analysis of the gold standard

Close inspection of a subset of the data revealed that the eye position trace within a given frame often varied in a similar pattern between frames, reflecting consistent patterns of image quality for a given participant or session, for example the top left corner being consistently dimmer and more blurry across frames and trials, either due to imperfection in the original alignment of the participant or reflecting gradual drift of the alignment during the session. Because the metric is based on image similarity, this would mean that even if the algorithm correctly extracted the eye position, the quality metric would be poorer for the time points when, for example, the light source was raster scanning over the top left, lower-quality part of the image. For a correctly positioned frame whose mean eye position quality metric was close to the quality cut-off, the individual quality data points could stray up and down across the cut-off in spite of the motion-corrected frame appearing to align well with the montage when inspected by the author. As such, rather than attempting to impose a quality cut-off on individual data points, the cut-off was applied to the mean quality metric for each frame, and as such whole frames were accepted or rejected.

While poor eye position quality metrics sometimes derived from changes in image quality during the trials, they could also derive from poor-quality montages. Inspection of the montages of participants with notably low quality metrics of frames that had been manually rated as acceptable

revealed two participants, ACI and AEC, whose montages were judged to be of particularly low quality. Their data were excluded from all further analysis.

Analysis of the frame-averaged quality metric (see Fig 4.7) for the remaining 16 participants showed that the mean values for the manually rejected and accepted frames were 0.031 (SE = 0.085; red dashed line) and 0.147 (SE = 0.029; blue dashed line) arbitrary units, respectively. The quality cut-off was adjusted to include as many manually accepted frames as possible while including no more than 5% of manually rejected frames, resulting in the inclusion of 74.7% of accepted frames from the human rater's gold standard. The designated quality value was 0.079 arbitrary units (yellow dotted line).

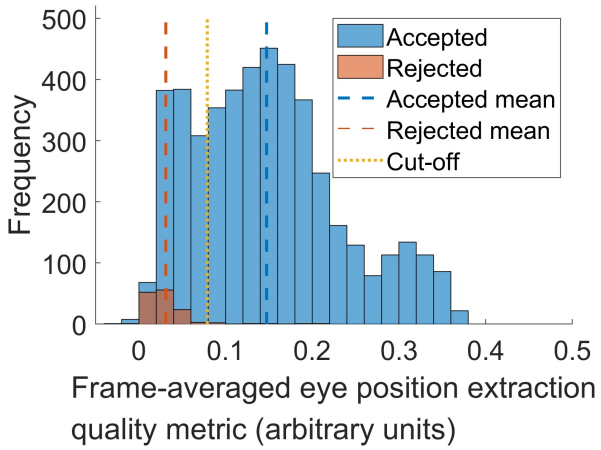


Figure 4.7: Histogram showing the distribution of the frame-averaged eye position extraction quality metric for the frames of which more than 50% of the eye position trace was marked as accepted (blue bars) and more than 50% was marked as rejected (red bars). The accepted frames had a mean quality metric of 0.147 arbitrary units (blue dashed line) and the rejected frames had a mean of 0.031 arbitrary units (red dashed line). The yellow dotted line at x = 0.079 represents the cut-off point below which frames in the experiment in Chapter 5 were rejected. The quality metric values are on a scale between -1 and 1.

4.6.3 When does the quality metric fail to capture the effectiveness of eye position extraction?

Direct comparison of the frame-averaged quality metric traces with the eye movement-corrected frames revealed that when the metric was high but a visual inspection of the placement of the frame against the montage indicated that eye movement tracking had failed, the frame in question had a large amount of uncorrected internal motion. Meanwhile, many frames with low quality metrics but good placement against the montage were noisy or somewhat blurred, but dimly visible blood vessels made it clear that placement was correct. Noise lowers the similarity between a frame and a montage because it is random, so it comprises structure in one image that is not present in another. The effect of blur is attributable to an over-reliance of the raw eye position quality metric on good mapping between the images of structures on the same spatial scale: because blurred cones appear

broader, a blurred and a sharp image of the same cones can be well-aligned yet dissimilar in the size of their structures. It would be useful to develop a metric less susceptible to noise and blur.

To conclude, the Oxford AOSLO can provide highly accurate eye position estimates, referenced to the photoreceptor mosaic. This chapter summarises the steps required to produce such estimates, including the collection of montage images and the manual generation of a montage; the collection of retinal images during trials and the referencing of these images to the montage; the adjustment of the origin to a behaviourally relevant point on the retina; and the rejection of poorly extracted eye positions, which necessitated the creation of a dataset of eye traces manually labelled for quality, and the calculation of an appropriate cut-off value for the eye position extraction quality metric based on the manually labelled traces.

Chapter 5

Eye movements with L- and M-cone isolating stimuli near the acuity limit

Declaration: The work contained in this chapter is entirely that of the author (SER). The adaptive optics scanning laser ophthalmoscope (AOSLO) and associated software were built by Young et al. (2018). 50% of the montages were made by ACH, RW and Samuel Ponting.

5.1 Introduction

Chapter 3 of this thesis described previous studies (Danilova et al., 2013; Miyahara et al., 1998) and presented fresh evidence that shows that the relative numbers of L and M cones – and therefore, the relative sampling densities of the two cone classes – impact performance on high-acuity chromatic tasks. However, in the field of colour vision there has, to date, been little consideration of the dynamic nature of vision: the eye does not remain still even during fixation, but is constantly in motion. It is known that performance on achromatic acuity tasks can be aided by the small eye movements that occur during fixation (FEM; Rucci et al., 2007) which permit resampling of the stimulus (Patrick et al., 2017) and may even be tuned to the requirements of the task (Intoy & Rucci, 2020). That sampling densities of L and M cones vary within and between individuals provides the opportunity for a natural experiment on the interactions between FEM, stimulus resampling, class-specific photoreceptor sampling density and chromatic high-acuity task performance.

5.1.1 Sampling- and optics-limited regimes

As long ago as 1857, Bergmann (translated by D'Zmura, 1996) argued that the discrimination of two points required that between the two nerve fibres carrying information about the intensity of the points there must be another nerve fibre communicating the different intensity of the gap. The

same is true for gratings, for each cycle of which there must be at least two neurons; if there are fewer, *aliasing* becomes possible, a phenomenon in which the grating resembles a lower-frequency pattern (Thibos, 1998).

However, before a stimulus can fall on the retina, it must pass through the eye's optics. Imperfections in optical transmission can affect perceptual sensitivity to gratings. To separate the limitations imposed by neural sampling and the eye's optics, Campbell and Green (1965) measured foveal contrast sensitivity functions (CSFs) for sinusoidal gratings presented to the eye as a whole, and to the retina directly, bypassing the optics of the eye as interference fringes. Regardless of spatial frequency, the contrast sensitivities measured for the whole eye were lower than those measured with interference fringes, indicating that foveal spatial resolution is primarily limited by optical factors rather than neural sampling. This explains why aliasing does not occur for foveally presented stimuli that are below the resolution limit: instead, such stimuli appear blurred (Thibos, 1998).

Conversely, with increasing eccentricity, the separation increases between the spatial frequency at which a grating's presence can be detected compared with the frequency at which it can be resolved (Thibos, Still & Bradley, 1996). This indicates that aliasing is occurring, which points to neural sampling density becoming increasingly relevant as cone and ganglion cell densities decline with increasing eccentricity (Curcio & Allen, 1990; Curcio, Sloan, Kalina & Hendrickson, 1990).

5.1.2 Resampling

Park et al. (2003) describe the super-resolution image reconstruction that can be achieved by resampling a scene following a subpixel shift in the detection array. Consider a 10 x 10 array of evenly spaced detectors (see Fig 5.1.a) that samples a stimulus (e) at time t_1 (b), moves 0.5 pixels to the right and downwards, and resamples the stimulus at t_2 (c). If the information gathered at t_1 (f) and t_2 (g) is combined (h), the stimulus has now been sampled in twice the number of locations with smaller spacing between each detector (d), thus improving the effective resolution of the detector array. The shifts themselves need not be subpixel: for an evenly-spaced array, provided that the shift in either the x or y dimension is not an integer shift, the stimulus will be sampled at locations intermediate to the original locations of data capture. For the human photoreceptor array, whose hexagonal arrangement is not perfectly regular (Curcio et al., 1990), any shift should permit the resampling of the stimulus at intermediate positions for much of the retina.

Patrick et al. (2017) have demonstrated that resampling itself is a key part of the benefit conferred by eye movements on peripheral vision by detangling the effects of (a) increased information

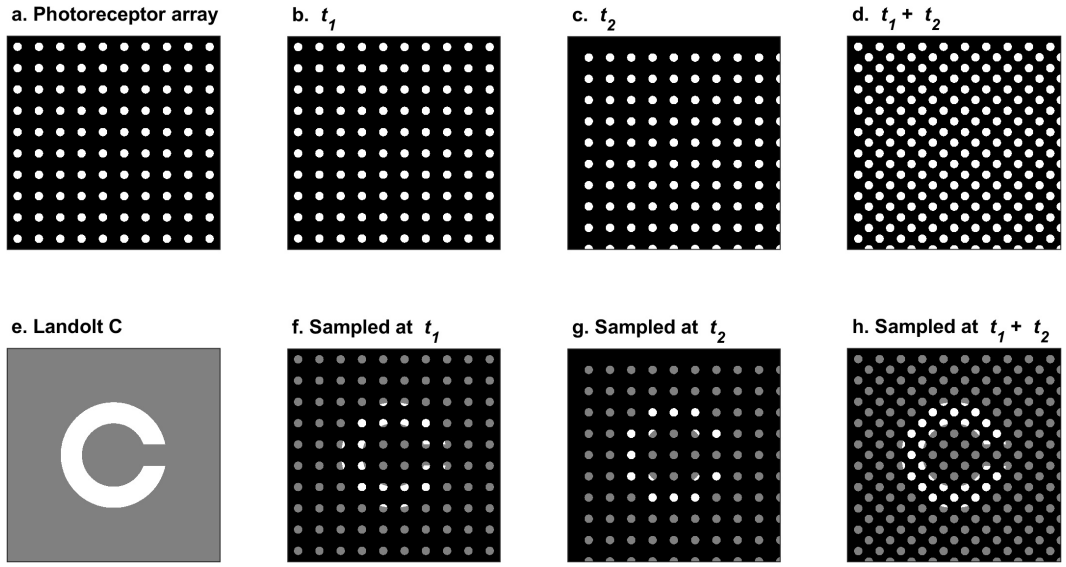


Figure 5.1: Effect of resampling on resolution. The first row of panels shows a 10 x 10 array of evenly-spaced detectors (a) in one position at time 1 (t_1; b) and at t_2 after repositioning to the right and downwards by half the spacing distance of the detectors (c). When the sampling locations at t_1 and t_2 are combined, a stimulus can be sampled with double the sampling density (d). The second row shows a Landolt C stimulus (e) as sampled by the array at t_1 (f), t_2 (g), and if the two sets of samples were combined, thus improving the effective resolution of the sampling array (h).

available from resampling and (b) specifically motion-related improvement. In Experiment 1, they overlaid static and moving Landolt C stimuli with static, random masks and required observers to report the location of the gap in the C. Performance thresholds were higher for the static target than for the moving target; however, samples of more positions on the stimulus were available on moving-target trials, meaning that it could not be determined whether the performance enhancement arose from resampling or from a motion-specific effect. In Experiment 2, they overlaid static and moving stimuli with masks that randomly updated on every frame, meaning that the number of independent samples of the targets were matched between conditions. The benefit conferred by stimulus motion was much reduced; however, a significant degree remained. Therefore, both increased sampling conferred by motion, and motion itself, support gap location discrimination.

It has proven difficult to find an intervention that closely mimics the experience of FEM. If the eye and mask remain static and the stimulus moves, a given photoreceptor will remain masked or unmasked throughout the trial; if unmasked, it will 'see' various parts of the stimulus. If the mask is moved and the eye and target remain steady, a given photoreceptor may or may not be exposed at some point during presentation, but if it is, it will only ever 'see' the same part of the stimulus. The first condition, then, is a better representation than the second of what happens when eye movements translate a stimulus across the retina. However, first, the fact that the stimulus will move relative to any other visible part of the room, for example the display boundaries, will lead to incongruous cues as to the simulated eye motion; second, a participant's own FEM will cause

additional motion of the image across the photoreceptor array. The second of these issues can be addressed by stabilising the image relative to the retina.

A number of methods have been employed to stabilise an image on the retina. One is to attach an image to a contact lens: as the eye moves, the lens, and therefore the image, move with it (Pritchard, 1961). However, the technique is uncomfortable for the participant, is limited in terms of stimuli that can be presented, and can only stabilise images rather than carefully manipulate their position over the course of a trial. Another is to stabilise the head and record the eye position with an eye tracker, rapidly updating a display to maintain or manipulate the position of the stimulus relative to the retina. This requires an extremely high-resolution eye tracker, such as a Dual Purkinje Image (DPI) eye tracker or an adaptive optics scanning laser ophthalmoscope (AOSLO) in combination with eye-tracking software, and a short latency between the eye movement, the extraction of the eye's new position and the update to the display. Stabilisation error achieved with DPI eye tracking and the EyeRIS eye-movement-contingent display (EMCD) system (Santini, Redner, Iovin & Rucci, 2007) has been estimated at less than one arcmin (the width of two foveal cones; see Supplementary Information, Rucci et al., 2007). Meanwhile, Yang, Arathorn, Tiruveedhula, Vogel and Roorda (2010) have achieved stabilisation with an estimated error of 0.15 arcmin, displaying stimuli by modulating the pattern of a scanning laser and adjusting the pattern using eye positions obtained from AOSLO imaging.

Stabilisation has allowed the study of the impact of different degrees of FEM on acuity. By presenting observers with unstabilised and stabilised stimuli in the fovea, Rucci et al. (2007) found that acuity for orientation discrimination of high-spatial frequency gratings was better when the stimuli moved across the retina with an observer's natural FEM compared to no movement. The results of this study defied previous modelling that had shown that optical factors would prevent FEM from improving foveal vision (Hennig & Wörgötter, 2004). Agaoglu, Sheehy, Tiruveedhula, Roorda and Chung (2018) later measured high-spatial frequency grating orientation discrimination performance with a range of manipulations including stabilisation, partial stabilisation, natural FEM, exaggeration of the observer's own FEM and motion opposite to the observer's own FEM. The best performance was obtained with partial stabilisation, indicating that the amount of natural FEM is not tuned for optimal performance, but in fact exceeds the ideal amount of FEM.

By stabilising a stimulus on the retina, superimposing FEM recorded on a previous trial and excluding any trials containing microsaccades, Ratnam, Domdei, Harmening and Roorda (2017) were able to investigate whether any of the beneficial effect of drift and tremor on acuity was mediated by efferent information. Performance on a high-acuity E optotype orientation discrimination

task was, as in earlier studies, worse for the stabilised condition compared to the natural FEM condition, while it was the same or better for trials using previously recorded FEM compared to trials under natural FEM. This suggests that efferent information is unimportant to the performance enhancement that occurs with eye movements.

While these studies are informative, there may be confounding effects from interfering with an observer's natural FEM. Yet in more naturalistic studies, it has not always been clear what the control condition should be. Microsaccade rate had been reported to be lower during fine spatial judgements such as threading a needle than during sustained fixation (Winterson & Collewijn, 1976), yet when Ko et al. (2010) also compared needle-threading with free viewing of natural images, the needle-threading microsaccade rate was instead five times higher. Nor have quantities of motion been directly related to task performance. By examining the effect of different amounts of naturally occurring eye movements on task performance, the experiment in this chapter circumvents the challenge of selecting an appropriate control condition.

5.1.3 L and M cones and resampling

Prior to the work presented in this thesis, the improvement in acuity conferred by FEM had only been considered in terms of the photoreceptors in general rather than any specific class of photoreceptor. Yet the arrangement of the cone classes demands this consideration, both in terms of FEM in general and resampling in particular. As described in Chapters 1 and 3, the cone photoreceptor cells comprise three subtypes of cell: the long-, middle- and short-wavelength sensitive (L, M and S) cones. The sampling density of each of these cone classes is necessarily lower than the overall sampling density of the cones in general, meaning they may benefit more from the resampling made possible by FEM in the case of undersampling (Patrick et al., 2017). Furthermore, the classes are not equal in number: only 5% of cones are S cones (Curcio et al., 1991), while L cones outnumber S cones by a factor of 2-2.5 (Carroll et al., 2000, revised upwards by Hofer et al., 2005). The potential resampling benefit may be higher for the less numerous classes.

Consider Fig 5.2.a, which shows a simplified, simulated foveal photoreceptor array. It has a square packaging arrangement, unlike the retina which approximates a triangular arrangement (that is, a hexagonal mosaic; Curcio et al., 1990). Like the average retina, it has a 2:1 ratio between the numbers of L and M cones (Carroll et al., 2000); as in the retina, the L and M cones are randomly arranged (Marc & Sperling, 1977); and it is absent of S cones, as is the fovea (Castańo & Sperling, 1982). Panels b and c show the array at time 1 (t_1) and, after it has moved, at time 2 (t_2). Panels g and h show the information gathered if the L-isolating Landolt C stimulus (f) were to be sampled

at t_1 and t_2. When the sampling arrays at both time points are combined (d), a clearer picture of the stimulus emerges (i). The addition of the information gathered at two additional time points affords a clearer image still (j).

While benefit conferred by the integration of information sampled at multiple time points, and therefore multiple positions, is striking when comparing panels g and h to i and j, the location of the gap is quite clear from the combination of just two array locations. However, this would not be the case if the Landolt C sampled had instead isolated the M cones (k): the gap location does not jump out at either of the individual sampling times (l and m). Indeed, it is still challenging to locate the gap with two time points (n), with the addition of the third and fourth samples (o) much more important to the discovery of the gap than when inspecting an L-isolating stimulus of equal size and contrast. These examples emphasise that stimuli that target a less numerous cone class may have more to gain from FEM than other types of stimuli.

Indeed, the ratio of the number of L and M cones provides a unique natural experiment on FEM and acuity. In the study by Patrick et al. (2017) described in § 5.1.2, the researchers also varied the density of the masks used by blocking out zero, 50% or 75% of the mask pixels. No masking is necessary when using stimuli that isolate either the L or M cones: the arrangement of the cones themselves provides a mask that is not only non-artificial, but also one that the observer has been accustomed to from birth. To further manipulate the mask densities, one can simply compare observers with different L:M cone ratios which, varying as they do from 0.37:1 to 16.5:1 (Hofer et al., 2005; or occasionally as low as 0.03:1, Miyahara et al., 1998), yield an ample potential range for the independent variable.

5.1.4 Fixational eye movements and colour

If FEM are more beneficial for stimuli or aspects thereof that are particularly visible to a minority cone class, we may expect FEM to have different degrees of effect on the perception of different colours. Until the work presented in this thesis, the study of colour and FEM has been confined primarily to early studies on fading. McCree (1960) studied the effect of prolonged fixation on chromatic discrimination in bipartite fields. Two colours were presented immediately adjacent to one another without a gap: one, a reference field fixed in chromaticity and luminance, and the other, a test field. Beginning with chromaticities close to that of the reference field, observers were presented with a test field of a given chromaticity and required to adjust its luminance until it matched that of the reference field. They then fixated on the stimulus until the border disappeared and the two fields appeared to be the same colour. This process was then repeated with a gradually

Figure 5.2: Effect of resampling on resolution for a simulated array of L and M cones. The first row of panels shows a simulated, evenly-spaced array of photoreceptors in a square packing arrangement (a), two thirds of which are randomly assigned as L photoreceptors (red dots) and one third as M photoreceptors (green dots). The array is shown in one position at time 1 (t_1; b) and at t_2 after repositioning (c). When the sampling locations at t_1 and t_2 are combined (d), and when sampling at other array positions at a further two time points are added (e), a stimulus can be sampled with greater and greater resolution. The second and third rows show an L-isolating (f) and an M-isolating (k) Landolt C stimulus as sampled by the array at t_1 (g and l), t_2 (h and m), if the two sets of samples were combined (i and n), and if a further two time points are added (j and o). The relative sparseness of the M cone array means that less information is gathered about the M-isolating stimulus than the L-isolating stimulus at each time point; it may therefore have more to gain from resampling.

increasing chromaticity difference between the reference and test fields until fixation could no longer merge the two fields. Thus, the maximum mergeable wavelength was determined.

McCree found that for some observers, fixation could blend very large differences in wavelength (for example, 400 and 600 nm fields were mergeable at lower luminances in a well-practised participant). Large variations in boundary fading susceptibility between participants were observed, and the ten observers included four who rarely saw the fixation effect. That the observer who produced the biggest fixation effects was the well-trained author suggests that those who saw little effect perhaps had poorer eye stability during fixation.

In general, maximum mergeable wavelength differences were much larger at shorter wavelengths (below around 570 nm) when the stimulus did not appear red, and low when the stimulus did appear red. This was echoed in further experiments using a bipartite field by Clowes (1962), who stabilised the image on the retina and found, similarly, that blue and green fields tended to fuse together; that green desaturated more than other colours; and that fade-out of the overall stimulus was stronger for blue and green than for other colours.

Because the above studies were not analysed in relation to the signals in particular cone classes or confusion lines in colour space, it is not possible to link them to more recent work on colour contrast thresholds for L- and M-isolating stimuli. However, that there is an uneven effect of fixation on hue discrimination for different wavelengths and an uneven number of the three classes of cones in the human retina is extremely interesting.

5.1.5 Quantifying eye movement

Two notable measures of fixation instability are the bivariate contour ellipse area (BCEA; Steinman, 1965) and the isoline area (Whittaker, Budd & Cummings, 1988). The BCEA, also referred to as the area of the bivariate normal ellipse (Timberlake et al., 2005), is an estimate of the area of an ellipse fitted to the distribution of a set of fixation position coordinates. The parameters of the ellipse are allowed to vary in the x and y dimensions to enclose, typically, 68% of the positions should those positions from a bivariate normal distribution. In the original description of the BCEA in normal observers (Steinman, 1965), finds that the data of every subject departs from bivariate normality; however, he argues that the deviations are so small that assumption of normality is generally acceptable in observers with normal vision.

The isoline area (ISOA), as it later came to be known, is a measure of fixation instability that originated with Whittaker et al. (1988). They calculated the most strongly preferred retinal locations in a vector by decomposing the area covered into a two-dimensional array of cells and calculating

the number of time points at which the eye was positioned within each cell. Starting with the most frequently occupied, cells were added in descending order until the selected cells comprised 68% of all time points. Greater instability is therefore expressed as a greater included area. As this technique does not assume normality and is unaffected by outliers, it is to be preferred to BCEA (Castet & Crossland, 2012).

Intoy and Rucci (2020) argue that the most appropriate model for drift is one of diffusion, a perfect random walk that does not attempt to make improvements to the positioning of the stimulus on the retina, and that drift is interrupted by positionally corrective microsaccades. However, this model is only valid if drift is indeed pure diffusion. The root mean square error (RMSE) of eye position (see Eq 5.2) calculates the deviation of eye position over time from a given location. It has a formal relationship to the diffusion constant (D) as the gradient of mean square error (MSE) over time from a starting position is equal to 4D. Unlike the diffusion model, it is free of any assumptions about the behaviour of drift; nor does it require the decomposition of an eye position trace into different types of eye movement, which itself can be problematic (Hauperich, Smithson & Young, 2020).

5.1.6 The present study

As described in Chapter 3, the clumping exhibited by a randomly mixed L and M cone mosaic means that measurements made with spatially extended targets do not offer an accurate estimate of a cone array's sampling resolution (Otake et al., 2000). As such, local targets such as a Landolt C should be employed (Danilova et al., 2013). This chapter will present an experiment in which observers were required to discriminate the location of the gap in L- and M-isolating square approximations of Landolt Cs. The targets were presented at fixed contrast yielding performance of approximately 66.7% correct, and eye movements were measured during the task using the AOSLO described in Chapter 4. The majority of FEM that occur during the task should be drift as microsaccades only occur on average once per second but, although it is a common measure of drift statistics, merely measuring overall eye displacement (e.g. West & Boyce, 1968, Schulz, 1984) is not a true reflection of the underlying eye movement, as drift changes direction frequently (Engbert, Mergenthaler, Sinn & Pikovsky, 2011). Instead, eye movement is better expressed by a metric, such as RMSE eye position, that takes into account the moment-to-moment eye position, and this metric will be used in this chapter to summarise the 'amount' of eye movement.

Hypotheses

Discrimination performance and eye movements were evaluated with respect to the following hypotheses:

1. **Regardless of scene content, individuals with different L:M ratios move their eyes differently from one another.**

   If individuals adapt their eye movements to their own L:M ratios over a lifetime, participants with high L:M ratios (scarce M-cone populations) should move their eyes a different amount to those with more even L:M ratios. They may move their eyes more in order to ensure that an M cone or enough M cones sample and resample the stimuli.

2. **Eye movements are tuned to scene content.**

   If the stimulus content affects eye movements, those with scarce M cones may move their eyes differently when presented with M- compared to L-isolating stimuli. Specifically, they may move their eyes more for M stimuli to improve the sampling of the stimulus with the relevant cone class.

3. **Each participant has an amount of eye movement associated with peak performance.**

   If eye movements aid chromatic vision, there should be a particular amount of eye movement – either across or within individual participants – that is associated with better chromatic discrimination performance. Because this study hypothesises that there may be eye movement differences between individuals and between which cone is isolated, performance versus eye movement is considered separately for each combination of observer and cone isolated.

5.2 Methods of data collection

5.2.1 Ethics

This study was approved by the Medical Sciences Interdivisional Research Ethics Committee at the University of Oxford.

5.2.2 Observers

To increase the chance of obtaining significant results if the hypotheses were true, it was desirable that the L:M ratios of the observers were as well-spaced as possible. Therefore, observers with the top and bottom 25% of HFP scores (a measure of L:M ratio) in Chapter 3 were invited to return for

the present study. Recruitment was intended to continue until 10 observers from each group had been recruited and had provided clear images, or all eligible observers had been invited back. In practice, all eligible observers were invited back, and 16 returned: eight in the lower quartile of L:M ratios and eight in the upper quartile L:M ratios. A further four observers from the central 50% of the distribution were also recruited to reach 20 observers.

Cones could not be imaged in two of these observers, who therefore did not complete the study. During the development of the eye movement quality metric described in § 4.6, two participants were found to have particularly poor-quality montages against which to reference their trial eye positions, so their data was excluded from further analysis. As detailed below in § 5.4.1, a particularly high proportion of the eye movement measurements of six of the remaining participants were of poor quality, and those observers were also excluded. Therefore, the final analyses included data from 10 observers.

5.2.3 Apparatus

As in Chapter 3, stimuli were presented on a CRS Display++ LCD monitor controlled by a CRS ViSaGe graphics system with 10 bits per channel RGB-encoded output. Imaging of the eye was carried out using the AOSLO described in § 1.4.3; the monitor was positioned 175 cm from the participant's eye, held steady with a bite bar to maintain the same viewing distance across experiments. Participants' location discrimination responses in the C optotype task were collected using a keypad.

Unlike in Chapter 3 the full spatial extent of the display was not used. The display was positioned so that the AOSLO scan patch appeared in the horizontal centre of the monitor; the display height was non-adjustable, and the scan patch naturally fell 2.8 deg below the vertical middle of the display. The scan patch had a height and width of 1.6 and 1 deg respectively. To increase the chance of obtaining clear features in the images while also maintaining as much background as possible, the stimuli were presented 3.7 deg to the left of the screen centre to produce the largest possible background that was symmetrical around the stimuli yet did not overlap with the scan patch.

A photodiode (OPT101 Monolithic Photodiode and Single-Supply Transimpedance Amplifier, Texas Instruments), in circuit with custom electronics developed in the Oxford Perception Lab and connected to the computer that ran the AOSLO, was positioned facing the far side of the display from the stimulus to record the screen intensity at that location to facilitate synchronisation of the computers that controlled the stimulus display and the AOSLO recording.

5.2.4 Trial stimuli

L- and M-isolating square C optotypes, as described in § 3.2, were presented for 500 ms at three contrasts each, the six conditions being randomly interleaved. The contrasts chosen were the contrast threshold for an (L+M) square C optotype obtained in Chapter 3 and the contrasts 0.5 log units above and below that value. If an observer required a long set-up time or many breaks, the contrasts that yielded performance closest to 66.7% correct were prioritised and the others were dropped.

5.2.5 Data collection

At the start of the first session, the observer was set up for the AOSLO and a set of montage images was collected as described in Chapter 4. For the trials, the observer looked at the screen in the middle of the four stimulus position dots until clear AOSLO images appeared in the live viewing window, at which point the experimenter triggered the display computer to begin the trial. The participant heard two warning beeps prior to stimulus onset, one occurring 1.65 s beforehand and the second 1 s before. Following stimulus presentation, the observer first indicated whether they thought the location of the gap was at the top, bottom, left or right of the stimulus, and then how confident they felt on a scale of 1 (Not confident at all) to 4 (Extremely confident), by pressing buttons on a keypad.

Image capture occurred in 70-frame (2.33-sec) bursts. The AOSLO was activated by presenting a white on-screen patch, hidden from the participant, to the photodiode 1.65 sec prior to stimulus onset; however, the actual time between triggering and recording onset varied. The length of image capture and the time between the appearance of the triggering patch until stimulus presentation were chosen because they captured the whole 500-ms presentation time, plus at least 200 ms before and after, regardless of the actual delay between triggering to AOSLO activation. The stimulus onset coincided with the presentation of a grey on-screen patch to the photodiode which enabled the time of stimulus onset to be identified in the AOSLO recording.

5.3 Methods of analysis

5.3.1 L:M ratio metric

In the first instance, all analyses of L:M ratios were undertaken using HFP scores because these were closer to a normal distribution. However, to demonstrate that the results would be obtained regardless of the transformation of the HFP scores, the same analyses were repeated using the L:M

ratios estimated from the HFP scores. As described in 3.3.2, L:M ratios below 1:1 were transformed in such a way that ratios of 2:1 and 1:2 were equidistant from 1:1. Because repetition of the analysis with L:M ratio was to demonstrate that any non-significant results did not occur simply because the independent variable had not been appropriately transformed, no corrections for multiple comparisons were applied.

The errors for L:M ratio are derived from the ends of the standard error bars for HFP, calculating $\bar{h} + SE_h$ and $\bar{h} - SE_h$ (where $\bar{h}$ is the individual's mean HFP score and SE_h is the standard error of the HFP score), then converting these values into L:M ratios to represent the ends of the error bars above and below the observer's L:M ratio.

Mean centre-to-centre cone spacing

For Hypothesis 2 (Eye movements are tuned to scene content), because the independent variables were HFP score and L:M ratio and the data were separated by which cone was isolated by the stimuli, it was possible to analyse eye movement in relation to class-specific cone spacing. As the shape of the relationship between L:M ratio and the spacing of a particular cone class is curved, class-specific mean cone spacing may prove to be a more useful regressor than L:M ratio. Statistical tests were therefore also carried out on the cone spacings. An equation to convert cone density in cells/mm^2 (d) to centre-to-centre cone spacing (s_{cc}) has been provided by Zhang et al. (2015):

$$s_{cc} = 1000[2/(\sqrt{3}d)]^{0.5}. \tag{5.1}$$

The mean foveal cone density, 168,162 cells/mm^2 (Zhang et al., 2015), can be multiplied by the proportion of L or M cones indicated by an estimated foveal L:M ratio. For example, to obtain an estimate of the foveal L-cone density for an individual with a ratio of 2:1, the general foveal cone density would be multiplied by 0.667. This value could then be converted to centre-to-centre L-cone spacing with Equation 5.1.

5.3.2 Eye movement metric

The process for obtaining eye position traces for each trial from the AOSLO retinal images is described in Chapter 4. In brief, the best quality montage frames that had been collected for each participant were selected and manually joined to create a larger, montaged image. The retinal images for each trial were referenced to the montage to determine the eye's position at each time point during data collection. Extraction of eye position from AOSLO retinal images depends critically on the quality of the images obtained during each trial and on that of each participant's

montage image against which the trial images are referenced. As discussed in § 4.2, there are many factors that affect image quality, such as the subtle day-to-day differences in the alignment of the AOSLO and differences between the eyes to be imaged, such as the level and types of aberrations and the ease of correcting them, and the reflectiveness of the retina. The resulting variability of image quality within and between participants meant that it was necessary to formalise a process for excluding data where poor tracking occurred and which would add noise to the within- and between-participant analyses. A significant step in this formalisation was to enable the application of an eye position extraction quality metric associated with each time-eye position unit of data, described in § 4.6. A subset of eye position traces for each participant were manually assessed by the author for apparent tracking failures that were then compared with the associated metric to determine a cut-off point that would remove most of the poorly-tracked parts of the eye movement traces while keeping as much acceptable data as possible. Additionally two participants whose quality metrics were particularly low were found to have poor-quality montages that could not be improved upon, and were excluded (see § 4.6.2); and, as will be described in § 5.4.1, some participants had notably high rate of frame exclusion following application of the quality cut-off, and were excluded.

Thus, eye movements were extracted from the AOSLO recordings and frames for which the average eye movement extraction quality metric was below 0.079 arbitrary units were removed and some participants were excluded based on this quality metric. Then, amount of eye movement in each trial was quantified as the eye position root mean square error (RMSE), defined here as

$$RMSE = \sqrt{\frac{\Sigma_{i=1}^{n}(d_i)^2}{n}} \tag{5.2}$$

where d_i is the Euclidean distance between each eye position in a given time window of length n time points and the median of all eye positions within that window. RMSE was computed for the 500-ms duration of stimulus presentation, corresponding to 15 frames. The eye position estimates had frequency of 1 kHz.

5.3.3 Trial selection

Of the stimulus contrasts presented during data collection, the contrast for each participant and cone isolated was selected that would retain the greatest number of trials following resampling to 66.7% correct (as described below). For the eye position RMSE values for each chosen contrast for each participant and cone isolated, outliers more than three scaled median absolute deviations (MAD) away from the median (as identified by the MATLAB function *isoutlier*) were discarded.

Following outlier discard, in order to keep difficulty constant for the L- and M-isolating trials, resampling was undertaken so that 66.7% of the included trials for each participant and each cone isolated were correct and 33.3% were incorrect.

5.3.4 Linear regression versus Spearman's rank correlation

For two of the hypotheses that follow, linear regressions are applied as the method of analysis. Where the independent variable was HFP score, the sample size was $n = 10$, and where it was L:M ratio or mean cone spacing (in which the value for one participant could not be estimated), $n = 9$, which were low sample sizes. As a result, the slight homoscedacity of residuals or some minor non-normality of residual distributions were difficult to evaluate. Spearman's rank correlation coefficient is a non-parametric alternative to linear regression that does not require the same assumptions; however, it was likely to be underpowered, so linear regression was used, which is known to be robust to minor violations of the non-parametric assumptions.

5.3.5 Hypothesis 1: Regardless of scene content, individuals with different L:M ratios move their eyes differently from one another

If eye movements affect performance differently for L- and M-targeting stimuli and observers therefore adapt their eye movements to suit their own L:M ratio over a lifetime, then individuals with different L:M ratios may move their eyes differently regardless of scene content. Specifically, those with few M cones who may benefit more from eye movements may typically move their eyes more when viewing any stimuli than those with a more balanced L:M ratio. This was assessed with a linear regression of RMSE eye position (collapsed over all trials) against HFP score and the estimates of L:M ratio obtained in the previous chapter. A negative slope for the regression against HFP, or a positive slope coefficient against L:M ratio, would confirm the hypothesis that those with high L:M ratios typically move their eyes more than those with lower ratios.

5.3.6 Hypothesis 2: Eye movements are tuned to scene content

If eye movements are tuned to the scene content, it may be expected that those with few M cones move their eyes more during the presentation of M- than L-isolating stimuli in order to resample and build up a better image of the stimuli. It is therefore hypothesised that when RMSE during presentation of M stimuli is subtracted from those during L stimuli, the regression slope of this difference against L:M ratio should be positive for regression against HFP score and negative for regression against L:M ratio estimates.

For the data underlying this analysis of difference, a linear regression of median RMSE against HFP score for M-isolating trials, which particularly challenge the sparseness of the M cones, should result in a negative slope coefficient, and against L:M ratio estimates it should result in a positive slope coefficient. Similarly, the slope coefficient for L-isolating trials should be positive for the regression against HFP and negative for the regression against L:M ratio, as those with high L:M ratios should have many more L cones than M cones and may therefore not benefit as much from eye movements as those with low L:M ratios. A t-test for difference of regression slopes should show that the M-isolating slope coefficient is less than that of the L-isolating slope for HFP score, and greater for L:M ratio. As the statistical tests of the data separated by cone were merely intended to illustrate the drivers of the analysis of difference of RMSE, no corrections for multiple comparisons were made.

Because individuals' median RMSE eye positions across stimulus type (Fig 5.4) and separated by which cone was isolated (Fig 5.5) were not normally distributed, the median RMSE values were analysed rather than the means.

5.3.7 Mean centre-to-centre cone spacing

As described in § 5.3.1, class-specific mean cone spacing is an additional useful regressor when separating trials by cone isolated. Statistical tests were therefore also carried out on the cone spacings, regressing median RMSE during L-isolating trials against estimated mean centre-to-centre L-cone spacing, and for M-isolating trials against estimated M-cone spacing. If eye movements increase as the spacing between cones increases (that is, as the number of that cone class decreases) and if eye movements are sensitive to which cone is being targeted, there should be a positive relationship between the centre-to-centre spacing of a cone class and the median RMSE eye position when isolating that cone.

5.3.8 Hypothesis 3: Each participant has an amount of RMSE eye position associated with peak performance

If eye movements aid chromatic vision, there may be a peak in chromatic task performance at a certain amount of RMSE eye movement. As the performance data per trial is binary, a logistic regression was performed for accuracy versus RMSE. If the performance peak is at the lowest or highest amount of eye movements that occurred during the trials, then a logistic regression of accuracy on RMSE would have a statistically significant slope coefficient.

If the peak is at an intermediate point along the RMSE axis, the resulting regression would

be close to flat. It may be that the amount of eye movement tends to be approximately correct for optimal performance. Therefore, it can alternatively be hypothesised that the amount of eye movements associated with the highest performance may be the median amount performed by the participant. Taking the absolute of RMSE − (median RMSE) should then provide the highest performance for values close to zero (that is, closest to the median) and the lowest performance for the values furthest from zero. The hypothesis would be confirmed in a significant logistic regression slope for accuracy versus abs(RMSE − (median RMSE)).

The regressions were performed separately for L- and M-isolating trials because it is hypothesised that the optimal amount of eye movement may be different between the two. They were also carried out separately for each participant because it is hypothesised that individuals will benefit differently for different amounts of eye movement depending on the cone isolated and the participant's cone ratio. They will be single rather than multiple regressions because the aim is not to obtain the best prediction of performance, but simply to discover (1) if a significant regression slope is obtained for either predictor, and therefore if there is a peak in performance at either the extreme amounts of eye movement or the median, and (2) if a pattern is to be found among participants, at which extreme (and for the absolute difference from median RMSE, at which value of the median) the peak occurs.

Follow-up: Can the median RMSE be predicted from estimated mean centre-to-centre cone spacing?

Were the linear regressions of median RMSE on estimated mean centre-to-centre L- and M-cone spacing carried out in § 5.3.7 to have significant coefficients, the logistic regressions in § 5.3.5 could be repeated with accuracy against abs(RMSE − (predicted median RMSE)). Predicted median RMSE may be preferable to measured median RMSE: compared to the eye position measurements, the HFP measurements were carried out in more controlled conditions, in the same lighting conditions and starting LED intensities, with scores comprising the average of 10 staircases; meanwhile, eye positions were obtained under variable conditions of lighting, stimulus visibility and between-trial gap length, and data clean-up may not have been equally effective for all participants. The predicted median RMSE may therefore smooth out any noise in the measured median RMSE.

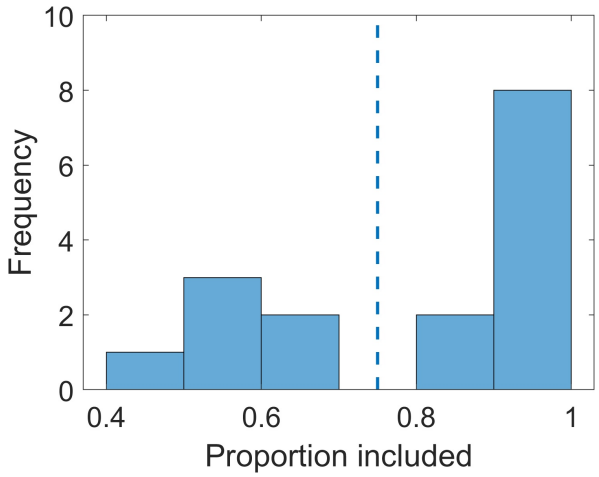


Figure 5.3: Histogram of the proportion of frames remaining for each participant following the exclusion of any frame with a mean quality score of less than 0.079. The dashed line indicates the gap in the bimodal distribution.

5.4 Results

5.4.1 Eye position data quality

Following removal of all frames with a mean eye movement extraction quality metric of less than 0.079, Fig 5.3 reveals a clear bimodal distribution of the percentage of frames remaining per participant. 10 out of 16 participants had 82% of frames or more remaining, while six had 66% or fewer remaining. Therefore, the data from those six participants were considered to be unreliable and excluded from further analysis.

5.4.2 Descriptives about RMSE spread

The amount of deviation in eye position during each 500-ms trial from the median eye position of that trial was computed as root mean square error (RMSE) eye position (see Eq 5.2) for the final 10 participants at the chosen contrasts and following outlier exclusion and resampling.

Boxplots for each participant's RMSE eye position, horizontally spaced by mean HFP score, are shown across stimulus type (Fig 5.4) and separated by which cone was isolated (Fig 5.5). It can be seen that there is considerable variation between participants in terms of both median and spread; for example, regardless of whether the data is separated by targeted cone or not, the tail of the distribution from observer AEL is particularly long. There is no immediately obvious difference between the spread of data for L- and M-isolating trials.

5.4.3 HFP to L:M ratio

One participant, ABS, returned such low a HFP score that his L:M ratio could not be modelled. All regressions on HFP score were performed using all 10 participants, while those on L:M ratio were

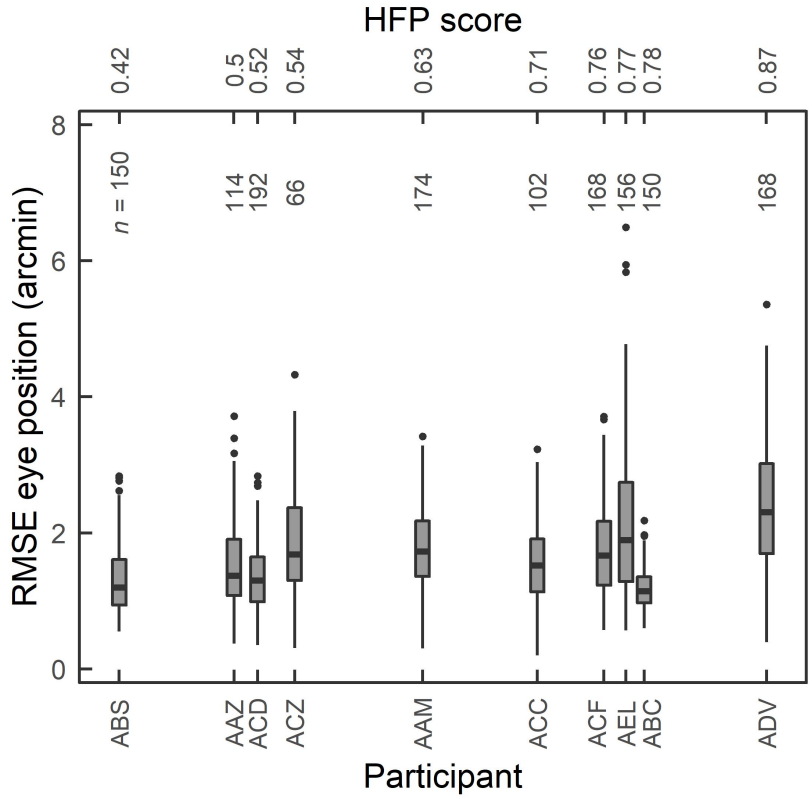


Figure 5.4: Box plots of the distributions of the amounts of eye movement (RMSE eye position, arcmin) included in the final analyses following (1) narrowing to a single contrast per cone isolated per participant, (2) the exclusion of poor-quality eye traces and outliers and (3) resampling. Boxes are spaced according to HFP score (shown on the upper *x* axis). Each box is labelled with the number of trials (*n*) in the distribution, and trials from both the L- and M-isolating conditions are included.

performed on the remaining nine participants. ACH, who designed and built the HFP apparatus and carried out the L:M conversions, estimates that the system is capable of measuring L:M ratios up to 12:1, suggesting that ABS has a ratio in excess of that figure (Hexley, personal communication).

5.4.4 Hypothesis 1: Regardless of scene content, individuals with different L:M ratios move their eyes differently from one another

The slope coefficient for the linear regression of individuals' median RMSE eye position across L- and M-isolating trials against mean HFP score (see Fig 5.6.a) was positive and non-significant (.44, p = .066). For the regression of median RMSE against estimated L:M ratio (converted as described in § 5.3.1, with L:M values below 1:1 decompressed), the slope coefficient was negative (-0.092) and non-significant (p = .214).

Fig 5.7 illustrates the regression assumptions for RMSE eye position versus HFP score (a and c) and estimated L:M ratio versus HFP score (b and d). The RMSE residuals for neither HFP nor L:M ratio are homoscedastic when plotted against the independent variable (a and b). The residuals are

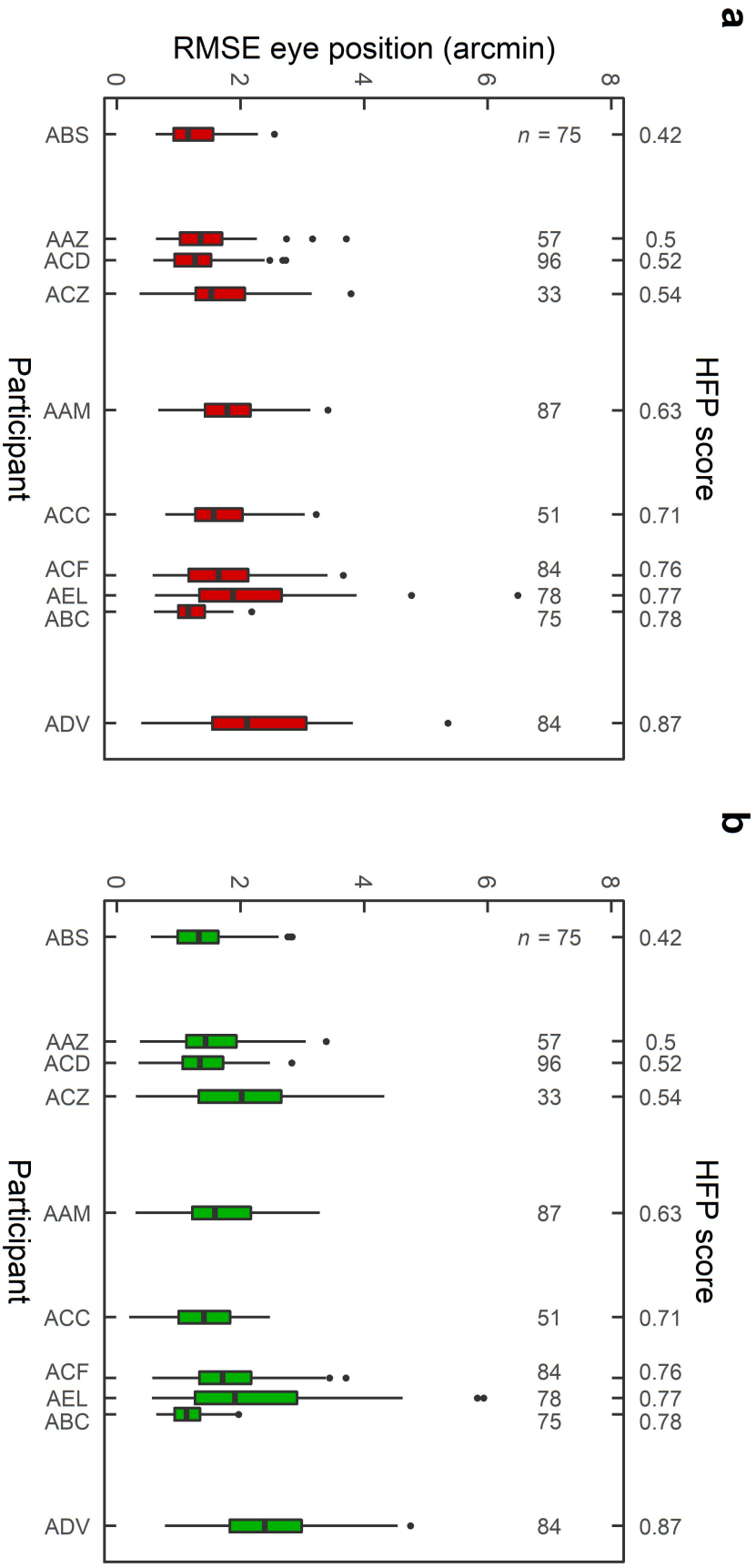


Figure 5.5: Box plots of the distributions of the amounts of eye movement (RMSE eye position, arcmin) for (a) L-isolating trials and (b) M-isolating trials included in the final analyses following (1) the exclusion of poor-quality eye traces and outliers and (2) resampling (as described in § 5.3.3). Boxes are spaced according to HFP score (shown on the upper x axis). Each box is labelled with the number of trials (n) in the distribution.

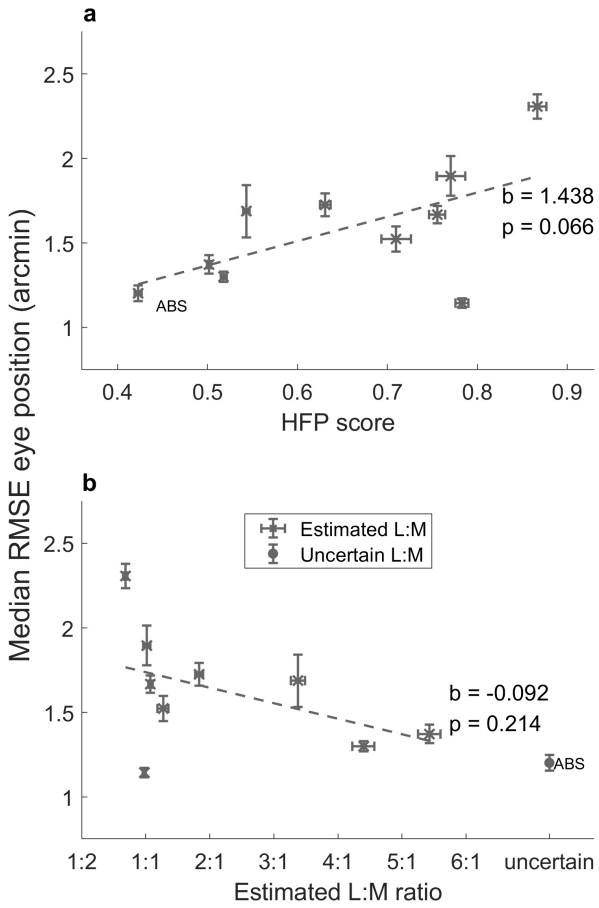


Figure 5.6: Amount of eye movement (median RMSE eye position, arcmin) plotted against (a) HFP score and (b) L:M ratio estimated from HFP score (crosses) with linear regression lines (dashed lines). In panel b, the median RMSE eye position for ABS, whose L:M ratio cannot be estimated, is plotted (filled circle) beyond the upper end of the L:M ratio scale. HFP score error bars represent the standard error of the mean, L:M ratio error bars represent the end points of the HFP error bars transformed into L:M ratios, and median RMSE eye position error bars represent the standard error of the median. The linear relationship between HFP score and median RMSE eye position has a slope coefficient, b, of 1.44 (p = .066), and that of the estimated L:M ratio and median RMSE eye position is -0.092 (p = .214).

not dissimilar from a normal distribution (c and d), although there are too few to attempt to fit such a distribution to them.

5.4.5 Hypothesis 2: Eye movements are tuned to scene content

The within-participants difference in median RMSE eye position between L-isolating compared to M-isolating trials was regressed against mean HFP score (Fig 5.8.a) and estimated L:M ratio (c). The HFP regression slope was positive but not significantly so (0.25, p = .615). The RMSE difference residuals for the HFP regression, shown in Fig 5.9.a, indicate that the data are slightly heteroscedastic as the residuals are somewhat closer to the zero line for extreme x values than for the middle x values. The histogram of the RMSE difference residuals (c) is slightly skewed, but there are too few data points to draw a firm conclusion. The residuals themselves are therefore only mildly contraindicative to a linear model.

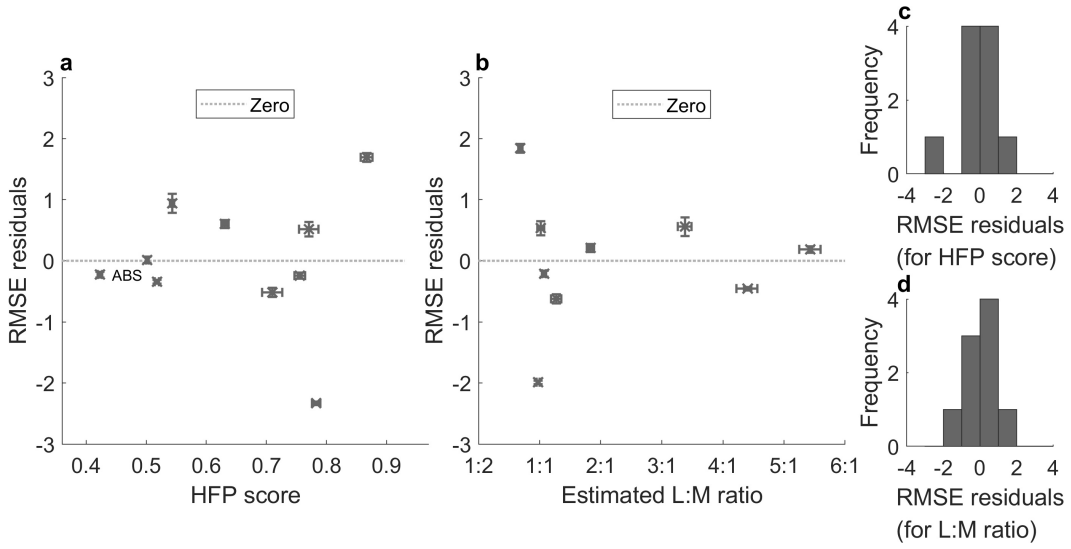


Figure 5.7: Regression assumptions for a linear regression of RMSE eye position versus HFP score (left-hand panels) and estimated L:M ratio (right-hand panels). RMSE residuals are plotted against the independent variables, HFP score and estimated L:M ratio (a and b, respectively). HFP score error bars represent the standard error of the mean, L:M ratio error bars represent the end points of the HFP error bars transformed into L:M ratios, and median RMSE eye position error bars represent the standard error of the median. Panels c and d show histograms of the RMSE residuals for the HFP and L:M ratio regressions, respectively.

For the L:M ratio regression, the residual plot in Fig 5.9.b is quite heteroscedastic, and the RMSE difference residuals are somewhat flatly distributed (Fig 5.9.d). The slope coefficient was non-significant and negative (-0.029, p = .546); see Fig 5.8.c).

To illustrate the data underlying the RMSE difference regressions, Fig 5.8 shows the linear regressions of median RMSE eye position on HFP score (b) and estimated L:M ratio (d) carried out for L- (red crosses) and M-isolating trials (green crosses) separately rather than as difference scores. The slope coefficients were non-significant for both HFP (L-isolating, 1.39, p = .045 and M-isolating, 1.136, p = .207, α = 0.025 after a Bonferroni correction for two comparisons) and L:M ratio (L-isolating, -0.093, p = .142 and M-isolating, -0.065, p = .455 , α = 0.025). Moreover, a t-test for difference of regression slopes found the difference between the L and M slope coefficients to be non-significant for both HFP (critical value = 1.75, t = 0.209, α = .05) and L:M ratio (critical value = 1.76, t = 0.264, α = .05).

Mean centre-to-centre cone spacing

Fig 5.11 presents the above data with estimated L:M ratio converted into L and M mean centre-to-centre class-specific cone spacing (red and green crosses). The linear regression of median RMSE eye position during L-isolating trials against L cone spacing (red dashed line) had a slope coefficient, b, of 0.404 (p = .132) and the regression of median RMSE during M-isolating trials against M cone spacing (green dashed line) had a slope coefficient of -0.093 (p = .465).

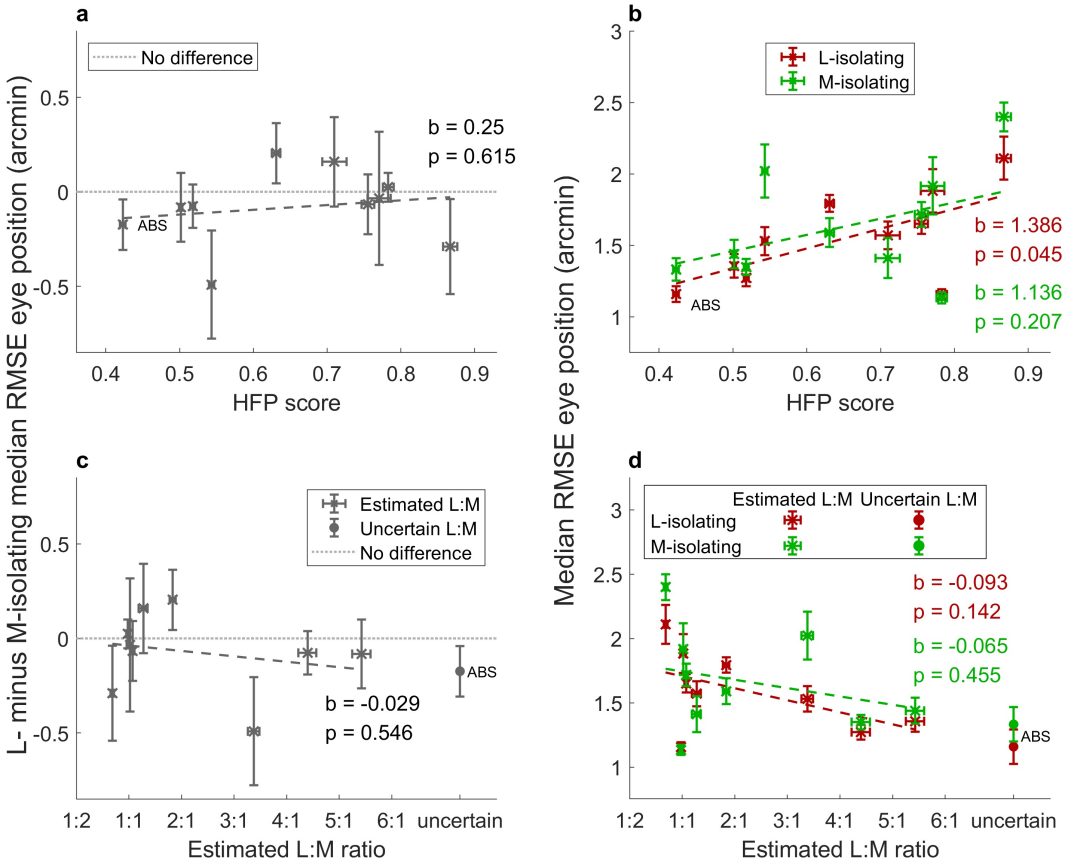


Figure 5.8: Linear regressions of the differences in amounts of eye movement for L- and M-isolating trials versus L:M ratio metrics, and underlying eye movement values. (a) The effect of HFP score on the difference between the amount of eye movement for L- and M-isolating trials is shown by plotting L- minus M-isolating median RMSE eye position (arcmin) against HFP score (grey crosses). The slope coefficient, b, of the linear regression (dashed line) is 0.25 (p = .615); the zero line (grey line) represents no difference between L- and M-isolating trials. (b) The median RMSE positions for L- and M-isolating trials (red and green crosses, respectively) that underlie the difference values in panel (a). The slope coefficients for the linear regressions for the L and M conditions (red and green text) are 1.386 (p = .045, α = 0.025 after correction for multiple comparisons) and 1.136 (p = .207, α = 0.025) respectively.

(c) As for (a), except the independent variable is estimated L:M ratio. The slope coefficient, b, of the linear regression (dashed line) is -0.029 (p = .546). For illustrative purposes, the RMSE difference value for ABS, whose L:M ratio cannot be estimated, is plotted beyond the high end of the L:M ratio scale (grey circle). (d) As for (b), except the independent variable is estimated L:M ratio and median RMSE eye positions underlie the difference values in panel (c). The slope coefficients for the linear regressions for the L and M conditions are -0.093 (p = .142, α = 0.025 after correction for multiple comparisons) and -0.065 (p = .455 , α = 0.025) respectively.

HFP score error bars represent the standard error of the mean, L:M ratio error bars represent the end points of the HFP error bars transformed into L:M ratios, and RMSE difference error bars represent the sum of the standard errors of the median.

5.4.6 Hypothesis 3: Each participant has an amount of RMSE eye position associated with peak performance

Logistic regressions were performed for accuracy versus median RMSE eye position (see Fig 5.12 in § 5.6: Appendix) and accuracy versus absolute difference from median RMSE eye position (see Fig 5.13). The panels are ordered from high to low L:M ratio from top left to bottom right. No slopes were significant.

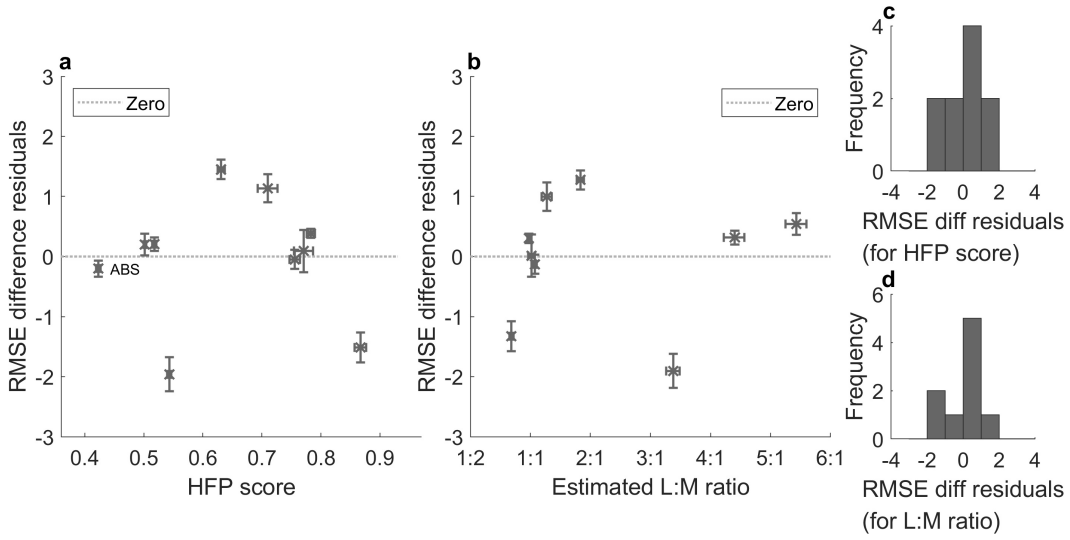


Figure 5.9: Regression assumptions for a linear regression of L- minus M-isolating median RMSE eye position versus HFP score and estimated L:M ratio. (a and b) RMSE difference residuals (arcmin) are plotted against the independent variables, HFP score and estimated L:M ratio, respectively. HFP score error bars represent the standard error of the mean, L:M ratio error bars represent the end points of the HFP error bars transformed into L:M ratios, and RMSE difference error bars represent the sum of the standard errors of the median. (c and d) Histograms of the RMSE difference residuals for the HFP and L:M ratio regressions, respectively.

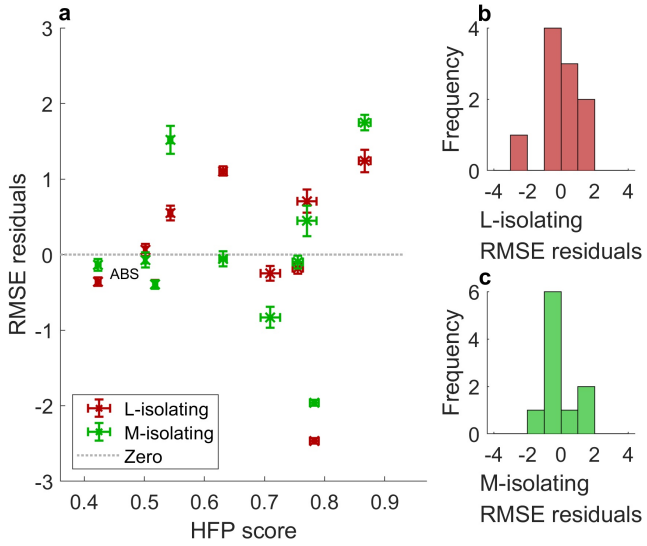


Figure 5.10: Regression assumptions for a linear regression of median RMSE eye position versus HFP score for L- and M-isolating trials (red and green crosses, respectively). (a) Eye position RMSE residuals for L- and M-isolating trials are plotted against the independent variable, HFP score. HFP score error bars represent the standard error of the mean and median RMSE eye position error bars represent the standard error of the median. (b and c) Histograms of the RMSE residuals for L- and M-isolating, respectively.

5.5 Discussion

Following studies indicating that fixational eye movements (FEM) are important in resolving high-spatial frequency stimuli (Martinez-Conde et al., 2006; Rucci et al., 2007), that particular amounts of FEM may be optimally suited to high-acuity task performance (Agaoglu et al., 2018), and specifically that FEM are adaptive to the task (Intoy & Rucci, 2020), it was hypothesised that the spatial scale of such movements may interact differently with spatiochromatic thresholds for stimuli that target

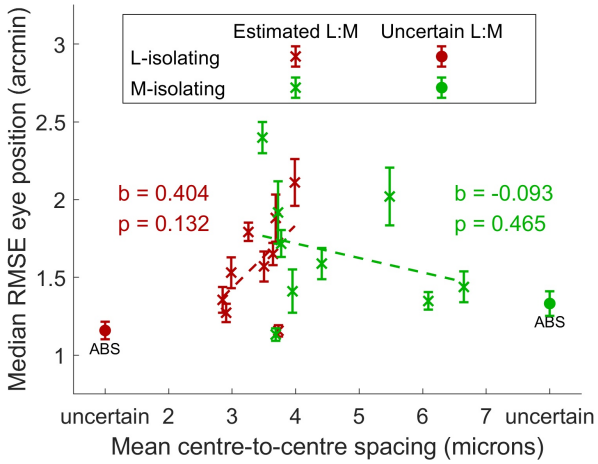


Figure 5.11: Median RMSE eye position (arcmin) for L-isolating trials plotted against mean estimated centre-to-centre L-cone spacing (red crosses) and median RMSE eye position for M-isolating trials against mean M-cone spacing (green crosses). The slope coefficient, b, of the linear regression for L trials is 0.404 (p = .132) and the coefficient for M trials is -0.093 (p = .465). For illustrative purposes, the mean RMSE for ABS, whose class-specific cone spacing cannot be estimated, is also shown (red and green circles). Median RMSE eye position error bars represent the standard error of the median. No error bars are used for estimated cone spacing to avoid crowding the plot.

cone classes with different mean spacings in individuals with different L:M cone ratios.

5.5.1 Hypothesis 1: Regardless of scene content, individuals with different L:M ratios move their eyes differently from one another

Where M cones form a small minority of the cone population (in a high-L:M retina with a low HFP score), observers may generally move their eyes more than those with a more balanced L:M ratio to ensure that the scene is well-sampled by both M and L cones. This would be borne out by a negative relationship between RMSE and HFP. In fact, the slope coefficient was positive, although non-significant (See Fig 5.6). The lower end of the HFP axis translates to the higher end of the L:M ratio axis, therefore although the slope coefficient for RMSE eye position versus L:M ratio should be positive according to the hypothesis, it was negative. There was therefore no evidence for the hypothesis that, in general, L:M ratio affects overall levels of eye movement, regardless of whether the data were analysed in terms of HFP settings or L:M ratio.

However, the p-value for the primary regression, median RMSE eye position versus HFP score, was close to the significance level (p = .066, α = .05). The low number of participants (n = 10) means that the test may have been underpowered. If the regression had reached significance, the positive direction of the relationship between amount of eye movement and HFP score may be explained in relation to a recent stabilisation study by Agaoglu et al. (2018), described in § 5.1. They found that peak performance occurred not with the natural amount of FEM, but instead when the stimulus was partially (43%) stabilised on the retina. It may therefore be the case that while FEM

are beneficial for high-acuity vision, individuals with sparse M cone populations have a greater need than those with more balanced L and M populations to keep their FEM to the most beneficial level, which is in fact a lower level than the average of the general population.

5.5.2 Hypothesis 2: Eye movements are tuned to scene content

The work in this chapter tests the strongest hypothesis to date regarding FEM and task performance: that not only do FEM adapt to suit the content of the stimuli, but those FEM adaptations are also affected by individual differences in retinal biology in individuals with normal vision. The hypothesis that eye movements are tuned to scene content differently based on an observer's L:M ratio was tested by regressing the difference between levels of eye movement on L- and M-isolating trials against, separately, HFP scores and L:M ratios. It was predicted that those with fewer M cones (lower HFP settings, higher L:M ratios) may have more to gain from eye movements when discriminating the shape of targets that are predominantly visible to M cones compared to those that are predominantly visible to L cones, and that more eye movements may therefore occur during presentation of M-isolating compared to L-isolating stimuli. By extension, those with lower HFP scores were expected to have a lower difference score than those with higher HFP scores, and the regression of L RMSE − M RMSE on HFP scores to have a positive slope. As those with low HFP scores have a high L:M ratio and vice versa, the regression of L RMSE − M RMSE on L:M ratio should have a negative slope. However, although the signs of the slope coefficients were as predicted, they were not significant (see Fig 5.8.a and c).

To understand the individual L-isolating and M-isolating data underlying this analysis of difference, Fig 5.8 also shows the linear regression (dashed lines) of individuals' median RMSE eye position against HFP scores (b) and estimated L:M ratio (d) for L-isolating (red) and M-isolating (green) trials. For HFP, although the direction of the (non-significant) L-isolating slope agreed with the prediction that L trials should elicit more eye movements with increasing HFP (decreasing L-cone density), the (again, non-significant) M-trial slope direction did not conform to the prediction that the M trials should see fewer movements with increasing HFP (increasing M-cone density). Furthermore, although the M slope was, as predicted, less positive than the L slope, the t-test for difference of regression slopes was once again non-significant. The same patterns were seen, reversed, in the regression of median RMSE against estimated L:M ratio. It should be emphasised that there are many participant-level factors that could affect the type and amount of FEM made by an observer, such as attention span, motivation and experience with eye tracking tasks. This made the within-subjects analyses of difference shown in Fig 5.8.a and c, rather than the separate

analyses shown in b and d, the most powerful test of stimulus-specific effects.

To further investigate whether different transformations of the data reveal relationships between the underlying L and M cone distributions and eye movements, the amount of eye movement was also analysed in terms of estimated mean centre-to-centre cone spacing (see Fig 5.11), which failed to yield a significant result. This method of plotting the data, showing that the L and M conditions have opposite slope coefficients, highlights the outcome of the analysis of the hypothesis under consideration: when targeting a particular cone class, there is no clear relationship between the amount of eye movement that occurs and the mean spacing of the cones targeted during such trials.

Therefore, although there is prior evidence that FEM may adapt to the spatial demands of a high-acuity task (Intoy & Rucci, 2020), no evidence is found that FEM adapt differently to stimuli that isolate different cone classes, even when the spatial statistics of those classes are different.

5.5.3 Hypothesis 3: Each participant has an amount of RMSE eye position associated with peak performance

To investigate whether participants' performances peaked at certain amounts of RMSE eye movement, logistic regressions were performed for accuracy versus median RMSE eye position (see Fig 5.12) and accuracy versus absolute difference from median RMSE eye position (see Fig 5.13). The panels are ordered from high to low L:M ratio from top left to bottom right. Potential patterns that might be expected to emerge include:

1. significant positive slopes among the regressions against median RMSE, indicating that more eye movements are associated with better performance;

2. significant negative slopes among the regressions against absolute difference from median RMSE for both L- and M-isolating trials, indicating that the median RMSE eye position achieved better performance than lower and higher RMSE; or

3. significant negative slopes among the regressions against absolute difference from median RMSE for M-isolating trials in high-L:M individuals, indicating that the median RMSE eye position achieved better performance than lower or higher RMSE where the relevant cone class was in the minority, but little or no impact of eye movement on performance for the dominant class or where cone ratio was more balanced.

However, of the 40 logistic regressions, none had significant slopes. There was therefore no evidence for peak performance at either the extremes of RMSE eye position seen during the trials, or

at median RMSE eye position. This is surprising because when stimuli are stabilised on the retina, simulating no FEM, task performance for grating orientation discrimination is reduced compared to trials without stabilisation which allow the stimuli to move across the retina with natural FEM (Agaoglu et al., 2018; Rucci et al., 2007). However, Agaoglu et al. (2018) also found that the best performance occurred when FEM were 43% stabilised, that is, when there was less movement than observers' eyes naturally made. If the findings of studies that use high-spatial frequency gratings can be extrapolated to studies such as the present one, which uses a local, high-acuity optotype, then in the present study it would be predicted that either (1) if enough trials occurred with such a reduced level of natural FEM, peak performance would be seen to occur when RMSE eye position was low; or (2) if insufficient trials occurred with such a reduced level of natural FEM, none of the regressions would be significant. The findings of the present study therefore do not directly support but are not inconsistent with those of Agaoglu et al. (2018).

5.5.4 Challenges and future directions

The most interesting result described in this chapter is that the relationship between median RMSE eye position and HFP score appeared, from visual inspection, to be positive (see Fig 5.6.a), and the linear regression returned a positive slope coefficient (1.11) with a p-value (.066) only slightly above the α level (.05). It is possible that the evidence for a relationship between this measure of L:M ratio and average amount of eye movement would have proven significant if more data had been available.

More data for the L- and M-isolating conditions could have been obtained had it been easier to target a fixed level of performance. The 66.7%-correct contrast thresholds obtained in Chapter 3 did not translate to the same performance when those contrasts were present in the fixed-contrast paradigm used in the present study, and the direction and magnitude of the change in performance was not consistent between cones or participants, meaning that it was necessary to present the stimuli at three contrasts in order to obtain one that yielded a performance close to 66.7% correct. There were several differences between this experiment and the experiment in Chapter 3 that may have made it difficult to use the data from that chapter to predict the appropriate fixed contrast for each cone and participant: (i) due to the ambient lighting requirements of AOSLO imaging (see § 4.2), the lighting conditions were slightly different between the two experiments; (ii) some participants may have found the pilot's single-contrast, fixed-threshold paradigm easier than the staircasing paradigm of Chapter 3 because in the former case, the contrast always remained the same; and (iii) the interleaving of the L- and M-isolating trials in the eye movement experiment may

have been more challenging compared to the first study, which presented trials isolating a given cone in blocks. Attempts to quickly determine the appropriate fixed contrast by testing different fixed contrasts at the start of each observer's eye movement study session were hampered by the slower rate of trial presentation when using the AOSLO, and noise in participant responses. These difficulties were unavoidable, but the experiment could nonetheless have benefited from a more rapid identification of the optimum contrast for each cone isolated for each participant in order to yield more trials per condition, perhaps selecting the best of three contrasts after approximately 300 trials (the usual number of trials aimed for being 600).

Resampling of the data to obtain performance of 66.7% correct for each cone isolated reduced the total amount of data available but was necessary so that the results were not confounded by variability in task difficulty. Confining analysis to a single contrast per participant, per cone isolated also reduced the amount of data, but was advisable due to the assumption of signal detection theory that the standard deviations of the distributions containing the noise and the signal+noise are both equal. The use of multiple contrasts varies the amount of signal in the signal+noise distribution, thus broadening it. This also occurs when, as in Chapters 2, 3 and this chapter, four-alternative forced choice (4AFC) is used rather than two-alternative forced choice (2AFC); however, 4AFC was justified in the former two chapters by the need for a low chance rate to increase the efficiency of the staircasing procedures and reduce the time taken to obtain thresholds, and in this chapter by the need (a) to use threshold levels obtained in Chapter 3, and (b) to make the task difficult enough that it required a high level of acuity while still retaining an analysable number of incorrect trials for the logistic regressions. The use of more than one contrast per observer, per cone isolated would also not often have added many trials to the total as performance for the additional contrasts was frequently either quite high or quite low and would have been subject to heavy resampling.

The amount of data available could also have been increased by increasing the number of participants. However, the number of observers whose data are included in the present experiment (n = 10) is comparable to or greater than those of the most similar studies. Research on eye movements and high-acuity tasks has been carried out with a maximum of six participants per individual experiment (Ko et al., 2010) and 13 participants per experiment (Intoy & Rucci, 2020), both using a DPI eye tracker which is easier to use than AOSLO eye tracking, while studies of individual differences in L:M ratios using AOSLOs have been performed with eight participants (Hofer et al., 2005) and five participants (Sabesan et al., 2015). A relatively high number of observers were excluded from the present study due to poor quality data (two before they could participate in the task, and eight based on a review of the eye position extraction quality metrics following the

experiment); however, reducing the number of exclusions risked increasing the amount of noise in the data.

The difficulty in recruiting participants with extreme L:M ratios to perform an additional five-hour task, and in obtaining usable images and eye movement traces with those who did participate, meant that the final number of observers whose data were analysed fell short of the initial target number of observers (10 with high and 10 with low L:M ratios). It would be desirable to increase the number of observers in the final analysis to reach the initial target, but this would involve recruiting another 100 participants to collect HFP settings in order to double the pool of observers with known ratios. However, the extremely time-consuming nature of data capture and analysis justifies the intention to selectively recruit observers with extreme L:M ratios, which increases the power in the regression.

Given that no peak in discrimination performance versus eye movement could be found, there was no evidence that a particular amount of eye movement supports vision when considered in terms of RMSE eye position. As described in § 5.5.3, it is possible that best performance in high-acuity tasks occurs with a lower average level of FEM than naturally occur (Agaoglu et al., 2018), and that this explains why no significant effects were found when looking for a peak in performance that related to three specific levels of eye movement. Alternatively, is is possible that (1) too many trials with large, deleterious eye movements were excluded due to having outlying RMSE values more than three scaled median absolute deviations from the median RMSE or (2) large eye movements caused tracking to fail resulting in the exclusion of many data points from the 500-ms trial, artificially lowering the RMSE value and adding noise to the data. These challenges are significant and improving data capture and clean-up is an ongoing effort not just in the Oxford Perception Lab but in labs internationally.

As discussed in § 5.1.5, RMSE better reflects the underlying eye movement than more traditional measures of assessing drift such as overall eye displacement, and has a formal relationship to the drift diffusion model favoured by Intoy and Rucci (2020) without making the same assumptions or requiring the decomposition of the eye position trace into different types of FEM. RMSE could be used further in the analysis of the present data by considering slices of only the first 100 or 200 ms rather than the entire 500-ms trial. There also are a number of additional ways the data could be cleaned prior to RMSE analysis, such as the identification and discard of microsaccades, the identification and separate analysis of drift and microsaccades, and the discard of any trial which itself has more than a threshold number of excluded timepoints. Alternatively, the eye position data could be quantified in terms of its isoline area (ISOA), another assumption-free metric of fixation

instability (Whittaker et al., 1988).

The paradigm this study has the most in common with is that of Intoy and Rucci (2020), who measured eye movements while discriminating locations of the legs of small tumbling E optotypes. They used a Dual Purkinje Image (DPI) eye tracker, a technology with one-arcmin resolution (two foveal cones; Crane & Steele, 1985). Since they found an effect of stimulus on eye movement and the present study did not, better comparison between our two studies could be facilitated by analysing the data above using their methods. Intoy and Rucci separated their data into microsaccades and drift, analysing drift in terms of speed and curvature (and, by extension, temporal power in the parvocellular ganglion cells), and microsaccades in terms of amplitude, direction and landing distance from the optotypes.

In the present study, confidence data were also collected on a rating scale of 1-4 from 'Not confident at all' to 'Extremely confident'. Although they were not analysed for this thesis, these data could be analysed to investigate the relationship between eye movements, confidence and cone isolation, and to estimate receiver operating characteristic (ROC) curves.

L:M cone ratio cannot perfectly predict the L- or M-cone spacing at the site of the stimulus, first because there is individual variability in general cone spacing (Curcio et al., 1990) and second because L and M cones are randomly arranged (Marc & Sperling, 1977), so a given location spanned by either the whole square C optotype or the part of the stimulus with the gap may not have an L:M ratio that is representative of the retina in general. Unfortunately, the premise of the experiment demanded that the stimulus be presented to the same retinal location throughout to encourage viewing with the PRL that might be more representative of the way stimuli are inspected when trying to discriminate small targets of fine spatial detail – however, the gap was presented in one of four different locations, hopefully narrowing the difference between the average micro-local L:M ratio and that measured on a larger scale by HFP. As there were only, on average, 18 trials per gap, per cone isolated, per participant, any analysis of the effect of gap location on performance for each cone and participant would have been underpowered.

5.5.5 Conclusion

The data and analyses presented here do not offer sufficient evidence to conclude that individuals' L:M ratios affect their eye movements, that eye movements are tuned according to which cone class the stimuli were more visible to, or that particular amounts of eye movement are more helpful for spatial chromatic discrimination near the acuity limit than any other degree of movement. It is not practicable to attempt to increase the amount of eye movement data due to the necessity

of either recruiting observers freely from the population, most of whose L:M ratios will be in the middle of the normal distribution and and whose data will therefore not be as informative as those with extreme ratios, or of doubling the number of observers for whom HFP scores are available to increase the pool of observers with extreme L:M ratios who can be recruited. Therefore, future work will be undertaken by the lab to improve data clean-up and implement alternative methods of analysis such as the ISOA and, if no significant results emerge, it will be concluded that there was no evidence for a strong relationship between L:M ratios, eye movements and chromatic target discrimination. It will be particularly interesting to see if a significant, positive relationship will be established between HFP score and median RMSE eye position if data clean-up can be improved and more data can be included in the analysis, as that would suggest that the more uneven an individual's L:M cone ratio, the less FEM they make, perhaps because low levels of FEM are better for performance than typical levels of FEM (Agaoglu et al., 2018).

5.6 Appendix

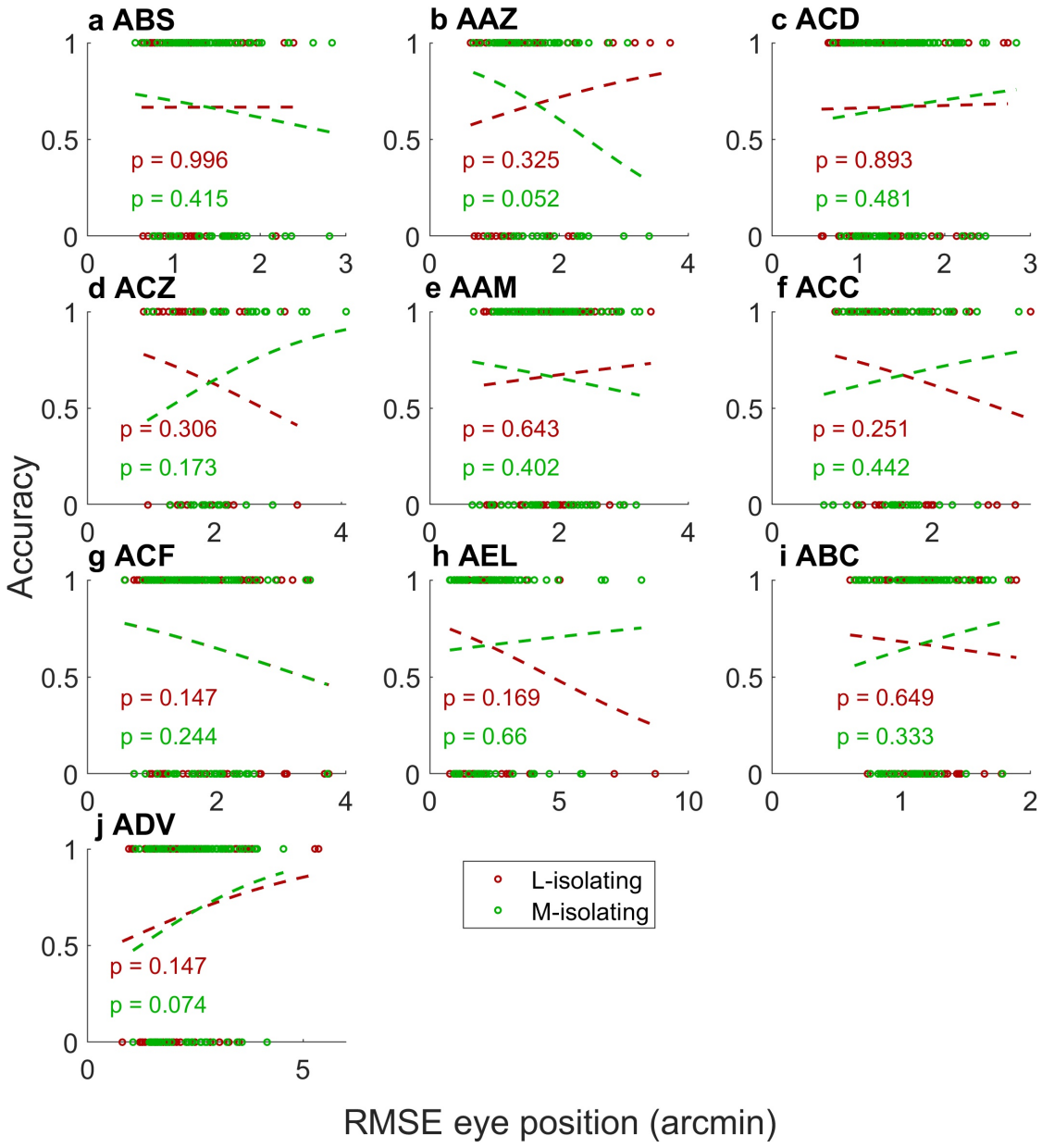


Figure 5.12: Accuracy versus RMSE eye position, single logistic regressions. Accuracy on L-isolating trials (red circles) and M-isolating trials (green circles) is plotted against RMSE eye position (arcmin). Logistic regressions for the L and M conditions (red and green dashed lines, respectively) are plotted together with the p value for each regression (red and green text). Participants are listed in order of L:M ratio from highest (a, ABS) to lowest (j, ADV). Note that the ranges of the x axes are not standardised to better illustrate the logistic regressions despite the considerable differences between the largest RMSE eye positions.

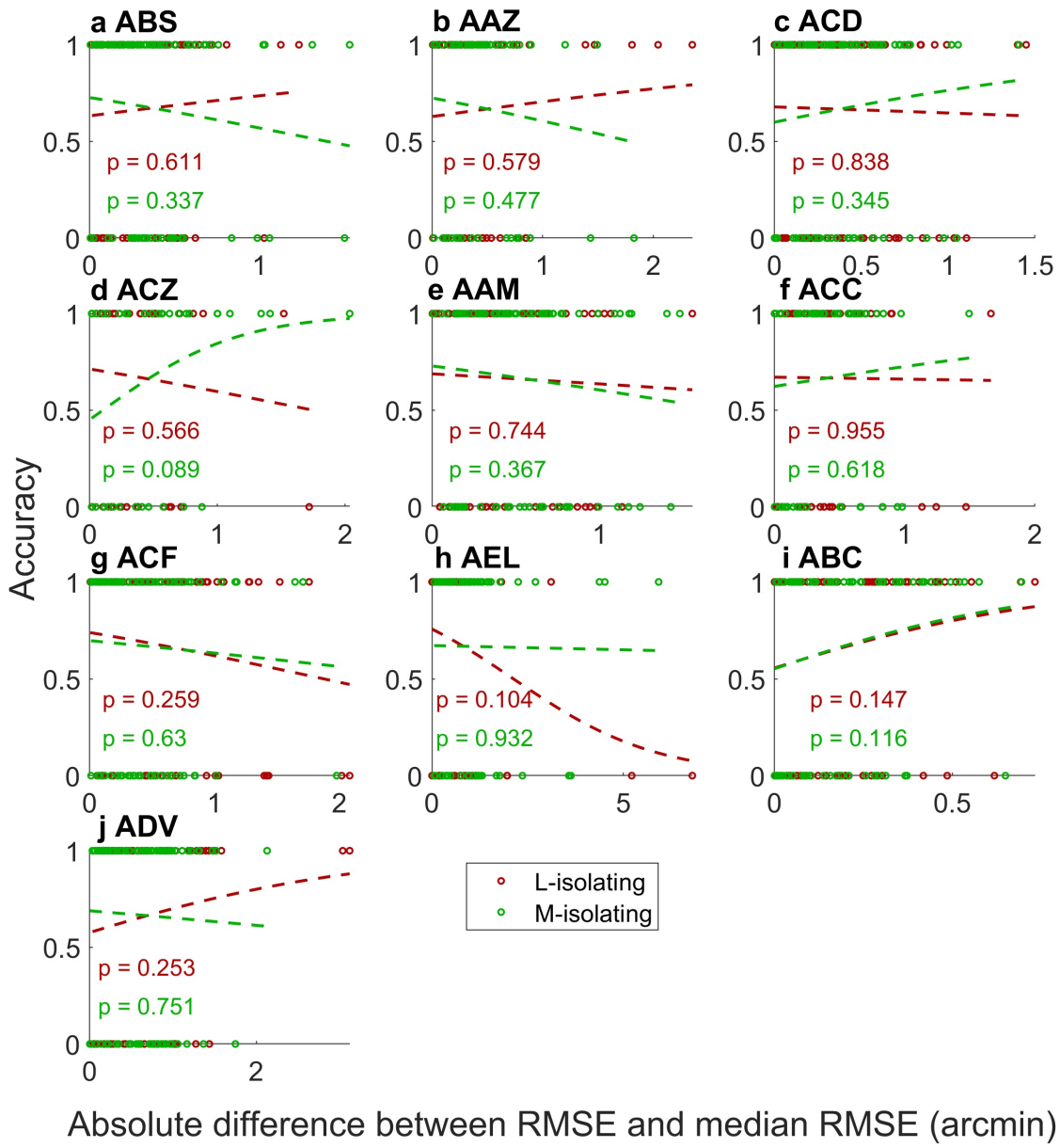


Figure 5.13: Accuracy versus absolute difference between RMSE and median RMSE eye position, single logistic regression. Accuracy on L-isolating trials (red circles) and M-isolating trials (green circles) is plotted against the absolute difference between RMSE eye position for that trial and the median RMSE eye position (arcmin). Logistic regressions for the L and M conditions (red and green dashed lines, respectively) are plotted together with the *p* value for each regression (red and green text). Participants are listed in order of L:M ratio from highest (a, ABS) to lowest (j, ADV). Note that the ranges of the *x* axes are not standardised to better illustrate the logistic regressions despite the considerable differences between the largest absolute differences.

Chapter 6

Cone photoreceptor localisation in adaptive optics retinal images

Declaration: The work contained in this thesis is entirely that of the author (SER), but for the exceptions described below. The retinal images were collected in collaboration with Mital Shah and Anna-Katharina Hauperich. The adaptive optics scanning laser ophthalmoscope (AOSLO) and associated software were built by Young et al. (2018). The automatic cone localisation algorithm was adapted from an algorithm written by Hannah E Smithson, modified from similar code by Mark D Fricker.

Useful units

Mean foveal cone spacing: 0.5 arcmin (Curcio et al., 1990)

Mean foveal cone spacing: 2.62 μm (Zhang et al., 2015)

Pixel width, Oxford AOSLO retinal image: 0.18 arcmin

6.1 Introduction

The introduction of adaptive optics technology to retinal imaging has revolutionised our ability to investigate living, seeing human eyes at high resolution. The Oxford Perception Lab's AOSLO is able to resolve single cones from the periphery to around 0.5 deg from the fovea (Young et al., 2018). Identification of cones in these images is necessary in order to extract relevant statistics such as cone density, cone spacing and anisotropy in packing arrangement. However, with hundreds of cones in any single image, manual cone identification is time consuming and depends, like all perceptual discriminations, on observer bias and sensitivity.

Individual groups have successfully implemented a variety of semi-automatic cone identification methods, but these methods tend to appear much less successful when applied to images collected by other labs (Mariotti & Devaney, 2015). This indicates that automatic cone identification methods are highly specific to the equipment whose images they were designed to assess, or to the lab-specific gold standards against which they are evaluated. Therefore (1) these methods may not work well for our own lab, (2) our lab may need its own method, and (3) the world of high-resolution retinal imaging needs more flexible cone identification programs and more robust evaluation of these programs. Indeed, Adam Dubis, Alf Dubra and Ann Elsner have all indicated that the field needs to see improvements in automated post-processing software (Marcos et al., 2017).

6.1.1 Existing methods

In AOSLO images focused at the plane of the inner photoreceptor layer, healthy cone photoreceptors appear as white blobs with soft edges on a darker background, the spacing between which increases with increasing eccentricity. The cones have a semi-regular hexagonal arrangement: in AOSLO images taken in the fovea, 55% had six neighbours, 23% had five, and 19% had seven, but the spacing became more inconsistent with increasing eccentricity as rods interrupted the cone mosaic (Wells-Gray, Choi, Bries & Doble, 2016). Rods can appear as much smaller white blobs, but may not be resolvable. As described in § 1.4.1, the intensity of the cones in the images varies both between cones and within a single cone over hours and days (Pallikaris et al., 2003) in a way that is, in a healthy retina, not linked to sensitivity (Bruce et al., 2015).

Most automated cone detection methods begin with a filter to remove high-frequency noise. Li and Roorda (2007) apply a low-pass filter to the image, take watersheds around local maxima, and finally merge cone markers that are improbably close together. Xue, Choi, Doble and Werner (2007) low-pass filter the image and subtract the filtered image from the original image, leaving cones clearly visible. The program then marks out cones in order from high to low intensities.

Both of the above methods are merely semi-automatic, in that they rely on the experimenter to make assumptions or interventions. Li and Roorda (2007) use the minimum foveal cone separation to determine the width of the gaussian used to initially filter the image, and the minimum separation for merging cones markers that are too close together. These minimum spacings may prove insufficient in peripheral images.

Xue et al. (2007) take this into account by filtering the image with a gaussian whose standard deviation is equal to one plus the retinal eccentricity, e.g. for images captured at 4 deg they use a standard deviation equal to 5 (units not given). However, as each cone is marked, it is also blocked

out with a square of width equal to the smallest possible foveal cone spacing to prevent cones from being multiply marked. The experimenter must also intervene by adjusting the threshold intensity for the dimmest cone eligible for inclusion until the image appears to be appropriately marked.

Further methods exist to segment the cone from the background, such as a least-cost path approach (Chiu et al., 2013); however, important cone statistics such as density, spacing and packing arrangement merely require the location of the centre of each cone.

Machine learning

A number of groups have applied machine learning to the problem of cone detection. The first to provide a fully automated approach, Cunefare et al. (2017) developed a convolutional neural network (CNN) that, for a dataset of confocal and split-detector (that is, splitting an annulus of non-confocal light into two halves and taking the difference between the two; Hamilton & Sheppard, 1984) AOSLO images. When validated against images from the same imaging system and retinal eccentricities as those in the training dataset, it provides comparable results to existing automated methods that required manual parameter adjustment, such as the graph theory and dynamic programming method of Chiu et al. (2013) which calls for a cone intensity threshold to be chosen by the experimenter. More recently, Hamwood, Alonso-Caneiro, Sampson, Collins and Chen (2019) employed a fully convolutional network (FCN), a type of CNN that runs more quickly and provides a per-pixel probability map for each image. Using images from the same confocal and split-detector datasets as Cunefare et al. for training and validation, for the confocal images they found the FCN to be statistically significantly better than the Cunefare CNN, but by only a small margin, and for the split-detector images there was no significant difference. Importantly, because both the CNN and FCN perform well for both types of imaging modalities without any kind of parameter adjustment, it is likely that it could be successfully used for other machines and modalities without adjustment.

However, this is provided that the experimenter generates and inputs a labelled training dataset for their own imaging system and relevant retinal eccentricities to train the network appropriately. The production of such a dataset is time-consuming and relies on the expertise and consistency of the labeller. Users of different confocal AOSLOs would benefit from a method that could adapt to the characteristics of their images without any training datasets or parameter intervention.

6.1.2 Method assessment

In 2015, Mariotti and Devaney used simulated AO fundus images to compare the performance of the Li and Roorda (2007), Xue et al. (2007) and Chiu et al. (2013) algorithms, plus the *daofind* algorithm

from the astronomy analysis environment DAOPHOT (Davis & Davis, 1994), which uses gaussian smoothing and local maxima extraction with the parameters of gaussian width and minimum height of the local maxima set by the experimenter. If the image is analysed at its original size, the detection rate of all of the algorithms is poor (85% or lower). Oversampling the image substantially improves performance, although using a factor greater than two rarely yields additional benefits.

6.1.3 The present study

The lab's existing automatic cone identification program uses a pipeline similar to that of Li and Roorda (2007) to label the middle of each cone. The image is first oversampled, which improves the performance of most if not all automatic cone identification programs (Mariotti & Devaney, 2015). It is then convolved with a laplacian of gaussians rather than a gaussian, which not only smooths out high-frequency noise but also enhances bright patches of an assumed size against a darker background, which has the benefit of helping to separate cones that are particularly bright or close together that would otherwise be interpreted by the program as a single cone. It then finds the watersheds around local maxima and computes the intensity-weighted centroids of each of the watersheds.

However, this pipeline requires the user to choose an appropriate standard deviation for the laplacian of gaussians; an inappropriately chosen standard deviation will systematically under- or overestimate the number of cones in a given frame. It would be preferable to automatically extract the most appropriate standard deviation from the image.

The present study collected AOSLO retinal images at a variety of horizontal and vertical retinal eccentricities and presented them to manual raters, randomly interleaved with simulated retinal images from a program that had been developed by another member of the lab (Young & Smithson, 2018). The raters, who were researchers experienced with AOSLO images, were asked to manually identify the locations of cones in each image in a carefully controlled setup. These cone coordinates were then analysed to:

1. assess the drawbacks of the use of manual cone localisation as the gold standard against which the performance of automated algorithms should be measured by comparing the differing cone localisations of the manual raters;

2. improve the level of automation in the Oxford Perception Lab's automatic cone localisation program using an improved manual gold standard; and

3. compare the performance of the classic Li and Roorda (2007) algorithm with the improved

Oxford Perception Lab algorithm, again using the improved manual gold standard as the benchmark for the programs.

The data collected will also support further analyses by the Oxford Perception Lab, allowing the group to:

1. evaluate manual raters' reliability when marking cones in the same image presented multiple times; and

2. assess manual raters' performance when marking cones in simulated AOSLO retinal images which have the advantage of a known ground truth.

6.2 Methods

6.2.1 Ethics

This study was approved by the Medical Sciences Interdivisional Research Ethics Committee at the University of Oxford.

6.2.2 Retinal image collection and selection

A total of nine subjects were imaged using the AOSLO. As shown in Fig 6.1, for each participant images were collected every 0.5 deg between 7 deg temporal and 7 deg nasal at 0 deg vertically, and between 4.5 deg inferior and 6.5 deg superior at 0 deg horizontally. The original intention had been to localise cones in patches from images gathered at 1, 3 and 6 deg; however, it was often not possible to obtain clear images at 1 deg as the cones are extremely small and closely packed at this eccentricity and are therefore more difficult to resolve. Instead images of patches of retina from 1.5, 3 and 6 deg were used. Of the original participants, seven provided usable images at sufficient locations. The best patch available in each case was chosen, based on (by eye) low signal-to-noise ratio, cone roundness, and lack of motion artefacts. Due to changeable image quality, sometimes an image gathered at an immediately neighbouring image location (0.5 deg away) was much better, and therefore a patch from the half of that image that overlapped with the image of the intended location was selected. The final set of human images are shown in Fig 6.2.

As the fall-off in cone density with increasing eccentricity occurs more rapidly vertically compared to horizontally (Curcio et al., 1990), images were selected along both a vertical and a horizontal axis (superior and temporal). The seven participants, three eccentricities and two axes therefore gave rise to 42 patches of real retinal images of 100 x 100 pixels (18.2 x 18.2 arcmin).

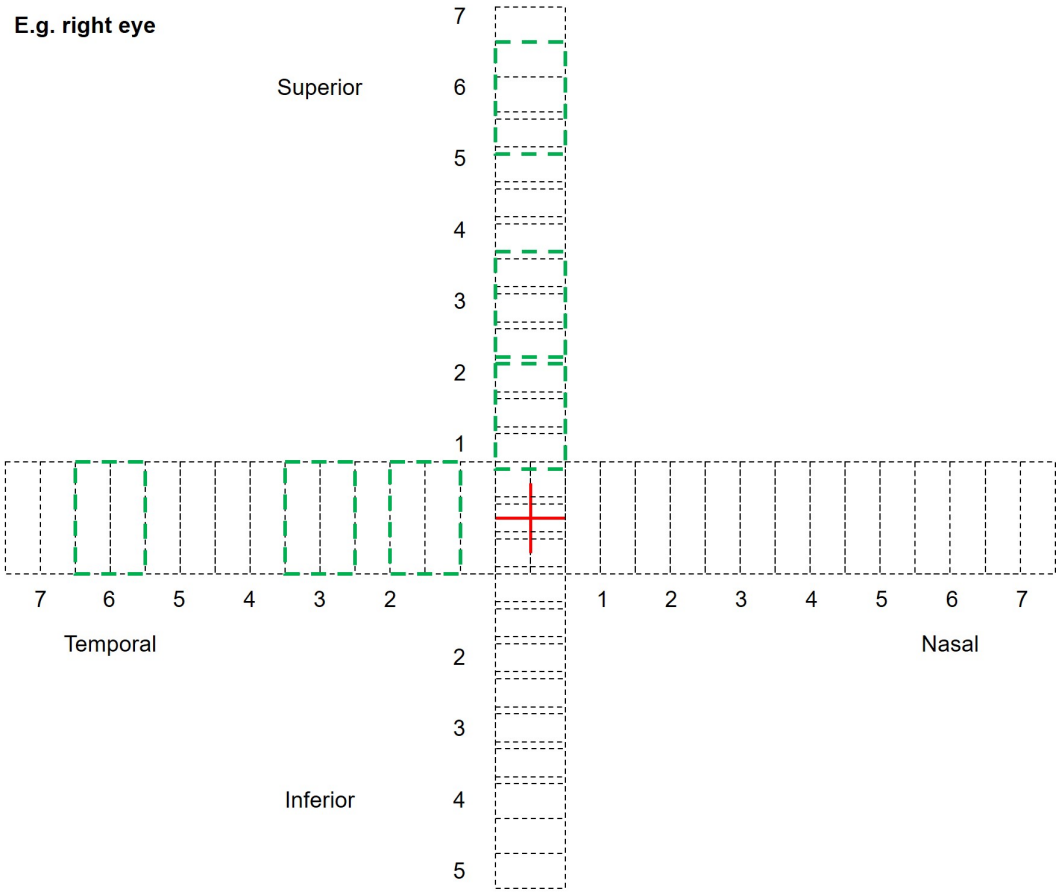


Figure 6.1: Location of image capture and use for a right eye. Images (black and green dashed boxes) were captured every 0.5 deg from (0,0) deg (red cross) between 7 deg temporal and 7 deg nasal at 0 deg horizontally, and between 4.5 deg inferior and 6.5 deg superior at 0 deg vertically. Images used for cone localisation (green dashed boxes) were taken from those collected 1.5, 3 and 6 deg temporally and superiorly.

The retinal image simulation program did not include direction from the fovea as a parameter, so four patches from 1.5, 3 and 6 degrees were chosen at random from a set of simulated images, providing 12 more 100 x 100-pixel images for rating.

A subset of patches were chosen for repeat rating: the central 50 x 50 pixels from each of the six larger patches from two real eyes (IDA and IDB, randomly selected from the seven available eyes), and from each of the three larger patches from two simulated eyes (simIDA and simIDB, the first two simulated eyes), giving a total of 18 smaller patches to be assessed five times by each manual cone localiser.

6.2.3 Manual cone localisation

Manual cone localisation collection

Six raters, all experienced in assessing AOSLO retinal images, carried out manual cone localisation on the images. Images were presented on an HP ENVY x360 15-bp103na Convertible Laptop with the keyboard folded away and propped up at a comfortable viewing angle on a desk; image

Figure 6.2: Images of human retina used for manual and automatic cone localisation. Six images from seven different eyes were used: one each at 1.5, 3 and 6 deg eccentricity in the temporal and superior retina.

presentation was achieved using MATLAB 2018b and Psychtoolbox-3.0.14, and observers marked the cones using a Microsoft Surface Pen Model 1776 which interfaced with Psychtoolbox using touchscreen-psychtoolbox (Borderies & Rautureau, 2018).

Localisation was carried out in a dark room, with a little light to reduce the sense of glare from the screen and with the screen and observer's head positioned so that there were no reflections on the screen. The laptop's ability to adjust its brightness based on the light levels in the room was turned off, and screen brightness was set to maximum.

The laptop display during manual cone localisation is shown in Fig 6.3. The images were oversampled with bicubic interpolation by a factor of seven, which was the amount required for the large images to nearly fill the screen. Observers were asked to use the stylus to mark the centre of every cone in the image; where cones were near the edge, they were asked to mark every cone whose centre they perceived to be visible. The cursor and cone markers were both represented by a 7 x 7-pixel (1.28 x 1.28-arcmin) blue square to represent a single pixel marker oversampled by the scale applied to the image.

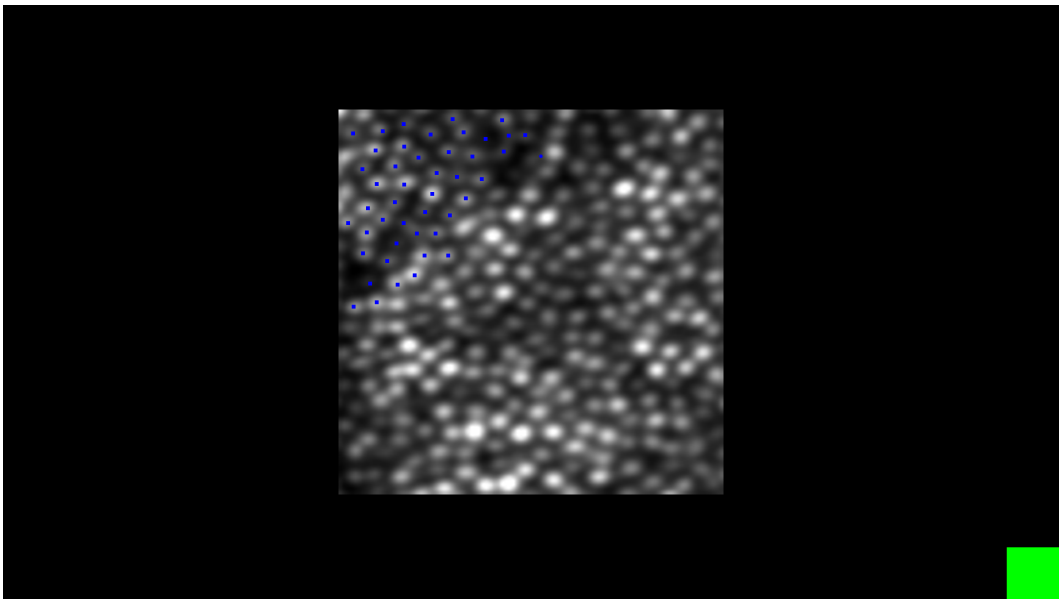


Figure 6.3: Manual cone localisation display. Manual raters are required to mark each cone by touching it with the tip of a stylus, leaving a blue mark on the screen to show the locations that have already been marked. Once the rater has decided that they have finished each image, they touch the green button to move onto the next image.

The first session began with stylus practice: observers were required to fully mark a simulated image for 1.5, 3 and 6 deg. Every further experimental session began with stylus practice, where observers were shown another set of images from each eccentricity and asked to mark at least 10 cones in each image, or more if they felt they needed more practice with the stylus.

The order of conditions (eccentricity and image size) was fully pseudorandomised and counter-balanced for two raters (Manual Raters Y and O; MRY and MRO), and partially pseudorandomised

and counterbalanced for the remaining four raters.

Manual cone localisation analysis

All steps of analysis of manual cone localisation took place with the cone coordinates oversampled with the same scaling factor used to present the images to observers for manual localisation.

First, accidental double-clicks were removed. A binary image was generated in which the coordinates were represented by a single pixel with a value of one against a background of zeroes. The image was then dilated with a disc-shaped flat morphological structuring element. The element's radius was equal to the radius of centre to the corner of the rating program's square cursor and coordinate marker, meaning that if the markers appeared to the observers to touch, or could have appeared to touch had the angle of image presentation been different, they would be merged. This means that any coordinates that have a single pixel-width or more gap between the dilated coordinate markers are taken to be separate cone localisations. A watershed segmentation was carried out to separate markers that did not touch. The centroids of the dilated markers within each watershed were calculated such that single or overlapping markers were represented by a single, averaged coordinate.

In order to create a group gold standard to which the automatic localisation algorithm could be compared, the manual localisations of different manual raters were merged. For each image, the raters were sorted in descending order of the number of locations they marked in that image, and in that order the raters were designated as Manual Raters 1-6 ($MR_{1\text{-}6}$) for the purposes of analysis of that image. The coordinates (c_1) provided by the observer who marked the most cones in that image (MR_1) were taken as a starting point. For each of the other observers $MR_{2\text{-}6}$, the Euclidean distance was calculated between that observer's coordinates (c_x) and MR_1's (c_1). Those pairs where the distance was greater than or equal to two times the oversample scale (that is, greater than 2 pixels in the original image) were discarded. The remaining distances were taken to represent connected coordinates. For any cone that was connected to more than one cone in the other set of coordinates, the shortest connection was preserved and the others discarded. This process was then repeated with any unconnected cones remaining in each of the sets c_{2-6} in turn.

For each of the coordinates in set c_1, and each of the cones in the other sets c_{2-6} that had not been connected to a cone in c_1, the mean was obtained of that coordinate and the coordinates of any linked cones in the other coordinate sets. The across-observer proportion certainty that a given mean coordinate was a cone was taken to be the total number of linked cones contributing to the given mean coordinate divided by the total number of observers (six). Fig 6.4 shows an AOSLO cone

image overlaid with the plotted coordinates of the group gold standard. The colours of the circular markers indicate the number of raters whose cones contributed to the marker's coordinate, from one (dark red) to six (blue-green).

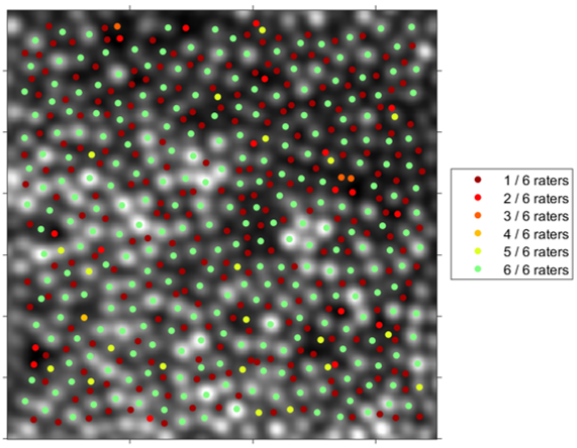


Figure 6.4: AOSLO cone image at an eccentricity of 1.5 deg overlaid with the cone locations identified by the manual raters. The colours of the symbols indicate the number of raters who marked a cone in that location. Manual Rater U (MRU) is responsible for the high number of cones marked by a single rater in this image (dark red dots).

Quantitative assessment of performance

An objective, quantitative method of evaluation was required in order to assess the performance of manual raters relative to each other, of the automatic cone localisation program relative to the manual gold standard, and (for simulated images) of manual raters and the automatic program relative to the simulation ground truth. From a linked test set (the cone coordinates under evaluation) and reference set (the coordinates against which they will be evaluated; that is, the ground truth obtained from manual cone localisation), the number of True Positives (cones present in both the test and reference sets; see Fig 6.5, green dots), False Positives (cones in the test set that do not appear in the reference set; see Fig 6.5, red cross) and False Negatives (cones in the reference set that are not identified in the test set; see Fig 6.5, blue circle) are extracted.

These numbers are translated into Precision, Recall and $F1$ scores for performance evaluation. Precision is defined as

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{6.1}$$

and expresses how many of the test set of cones were truly cones, while recall is given by

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{6.2}$$

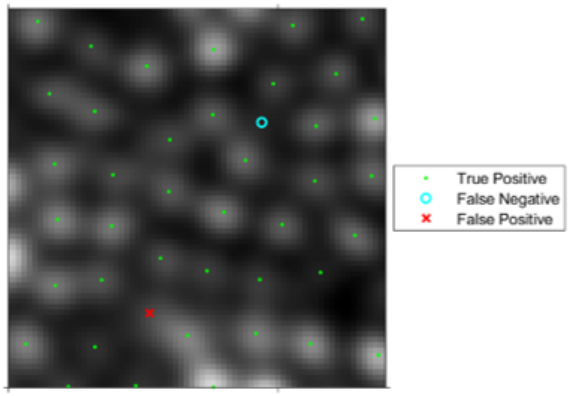


Figure 6.5: An AOSLO cone image at an eccentricity of 1.5 deg showing the comparison of a test and reference set of cone coordinates. The green dots represent cone coordinates that appeared in both the test and reference sets; the blue circle shows a coordinate that was in the reference set but not the test set; and a coordinate that was in the test set but not the reference set is indicated by a red cross.

and indicates how many of the true set of cones were correctly identified. Perfect performance would be given by $(Precision, Recall)$ coordinates of $(1,1)$. The deviation of a test set from the reference set can be expressed using the $F1$ score:

$$F1 = 2\frac{Precision * Recall}{Precision + Recall} \quad . \tag{6.3}$$

$F1$ varies from 0 to 1, where a value of 1 represents $(Precision, Recall)$ coordinates of $(1,1)$.

Gold standard generation

Where a true ground truth of cone locations is unavailable, the gold standard traditionally consists of manual localisation carried out by an expert in cone images (e.g. Chiu et al., 2013; Li & Roorda, 2007; Xue et al., 2007). To distil the assessments of the six manual raters into a single set of cone positions for each image, the judgements of each rater were compared to the performance of all of the raters. The temporary gold standard was taken to be the set of cones agreed upon by five out of the six raters (83.3%). In turn, each manual rater's cone positions were compared to the temporary standard to identify True Positives, False Positives and False Negatives which could then be converted into Precision and Recall scores, and then finally $F1$ scores using Equations 6.1, 6.2 and 6.3. Outliers were detected using the MATLAB function *isoutlier*, which identifies any value in a set that lies three scaled median absolute deviations (MAD) from the median.

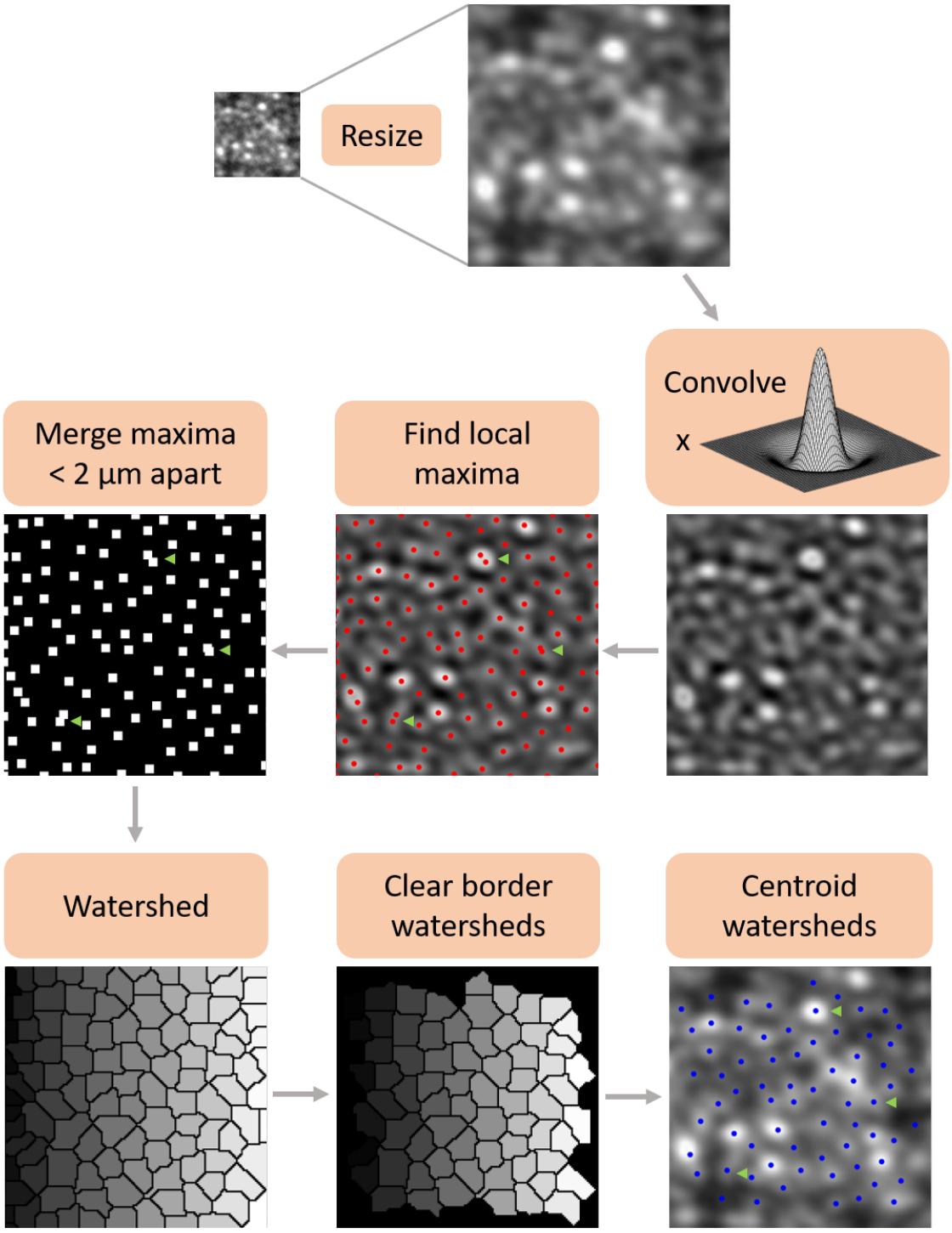


Figure 6.6: Processing pipeline for the automatic cone localisation algorithm. (1) Image may or may not be oversampled by a given factor. (2) Image may or may not be convolved with a filter; if a filter is used, it is either a gaussian (not pictured) or a laplacian of gaussians (pictured). (3) The local maxima are found (red circles). (4) If merging is carried out, the peak locations are dilated using a disk with the diameter of the minimum foveal cone spacing, 2 μm, which appears square as an artefact of how small the element is at the oversampling factor used in this example. Touching dilated peaks are marked with a green arrow. (5) A watershed is applied to either the local maxima of the convolved image or, if merging is used, the dilated maxima markers. (6) Watersheds touching the border are cleared. (7) The log centroid is calculated to find the centre of mass of each watershed, which is taken to be the centre of the cone (blue circles). By comparing the areas marked by the green arrows for the local maxima (red circles), the dilated peaks (binary image) and watershed centroids (blue circles), the effect is seen of merging cone coordinates that would be implausibly close together.

6.2.4 Automatic cone localisation program

Program

The automatic cone localisation program was written in MATLAB 2019b making particular use of the Image Processing Toolbox. The pipeline is represented in Fig 6.6. The image is oversampled at a given scale using bicubic interpolation and either convolved with gaussian or a laplacian of gaussians (whose standard deviation is also multiplied by the oversampling factor) using *imfilter* or left unfiltered. The local maxima in the convolved image are then identified by *imregionalmax*. The program then contains the option to merge all cones closer than the minimum possible cone separation by dilating a black and white representation of the centroids positions with *imdilate* and a disc-shaped structuring element of diameter equivalent to 2 μm, then applying *imregionalmax* to the dilated image to obtain new local maxima. The boundaries between the peaks and the areas within those boundaries are marked using *watershed*. All watersheds touching the edge of the image are removed using *imclearborder*, and *regionprops* generates the estimates of the cone locations by calculating the centre of each of the remaining watersheds weighted by pixel intensity.

Program parameters

Three types of filter were assessed: no filter, a gaussian and a laplacian of gaussians. The gaussian filter was chosen to test the efficacy of the cone localisation algorithm of Li and Roorda (2007), which used a gaussian filter, for this data set, while the laplacian of gaussians was chosen because it specifically enhances bright patches with a dim surround, which is how cone photoreceptors appear in AOSLO images. The algorithm for the present study was also run without a filter to determine whether the filters tested provided any benefit compared to not using a filter. The standard deviation of the gaussian filter was varied between 0.05 and one pixels in steps of 0.05 pixels (that is, 0.01 arcmin to 0.18 arcmin in steps of 0.01 arcmin), and that of the laplacian of gaussians filter was varied between 0.2 and two pixels in steps of 0.05 pixels (0.04 arcmin to 0.36 arcmin in steps of 0.01 arcmin), the extremes of which were determined with coarser steps during pilot analysis. The filter width was multiplied by the oversampling factor. The oversampling factors used were one (no oversampling) to seven in steps of one. Each filtering/oversampling parameter combination was carried out both with and without the cone-merging step.

Quantitative assessment of parameter performance

The cone positions output by the program were compared with the gold standard as described in § 6.2.3: Gold standard generation. The procedure for comparison was similar to that for comparing

cone positions between different manual raters described in § 6.2.3: Quantitative assessment of performance. The distances between all the cones in the test (automatic) and reference (gold standard) sets were calculated and any distances over a given tolerance were discarded. Where any reference cone was still linked to more than one test cone, the longest links were discarded and only the shortest was kept. This was then repeated for any test cones linked to more than one reference cone, such that no cone in either the test or reference set was multiply linked. Precision, Recall and $F1$ were calculated for each combination of eye, eccentricity and direction from the PRL using each combination of program parameters to determine the optimal scale, filter type, filter standard deviation and use of cone-merging within and across conditions.

Narrowing of parameter space

Each of the algorithm types requires the choice of an oversampling factor, and the latter two also require a filter standard deviation. It is possible to either choose them in combination or in sequence. The former approach would result in a parameter space with $F1$ scores for each parameter combination. If performance does not vary smoothly and consistently across the parameter space, the maximum $F1$ score may be an artefact of local noise, and another location in the space might have a larger area of high $F1$ scores. Smoothing the $F1$ values in parameter space would be necessary to help reduce noise and select a value in an area of the space more robust to variation in the optimal parameter; however, smoothing techniques either call for the discard of values at the extremes of the space, or affect them differently compared to those in the middle. As one extreme value (an oversampling factor of one) is particularly important and cannot be discarded, the parameters were chosen one at a time.

First an oversampling factor was chosen for each algorithm across filter standard deviations, as filter standard deviation is expected to vary with retinal location and it is desirable to vary no more parameters than necessary. For each oversampling factor, the mean $F1$ score across all other algorithm parameters, retinal locations and individual eyes was calculated. A high $F1$ score indicated the oversampling factor to be chosen for that algorithm.

There are two possible ways to choose the optimal width for a given filter and location: to identify the filter width with the highest $F1$ score for each individual eye, and then average those widths across eyes; or to select the width with the highest mean $F1$ score across eyes. The first has the disadvantage that noise may distort the eyes' width-$F1$ curves, requiring smoothing or curve-fitting. The second is more resistant to such noise as no width is selected until averaging across the eyes' curves has taken place. It is also more consistent with the motivation for choosing the width:

the topic under present consideration is not the ideal filter standard deviation for individual images, but the standard deviation that will work best across a range of images. As there is no 'default' standard deviation, no statistical test is necessary to demonstrate the comparative advantage of that with the maximum $F1$ score.

To choose whether the cone-merging step conferred an advantage that merited its inclusion for each filter type, $F1$ score for a filter type without merging was subtracted from its corresponding merged condition to see if the performance advantage exceeded zero. Paired statistical tests comparing performance with and without merging were run separately for each filter type.

Finally, when all of the parameters have been optimised, the overall performance of each of the three filter types were compared using a repeated-measures test.

6.3 Results

6.3.1 Cone localisation gold standard

The cone localisation performance of each manual rater for human images was assessed relative to a temporary gold standard of the set of cones agreed upon by five out the six raters (83.3%). The $F1$ score (0.770) of Manual Rater U (MRU) was an outlier. The mean score of the remaining five raters was 0.928 with a standard error of 0.0112. For all subsequent analyses MRU's localisations were discarded and the final gold standard for each image was taken to be the mean positions of the set of cones upon which four out of the remaining five manual raters (80%) agreed.

6.3.2 Narrowing of algorithm parameters

Fig 6.7 shows a set of heatmaps representing the effect of oversampling factor and filter standard deviation on mean $F1$ scores averaged across eyes when applying the laplacian of gaussians pipeline without cone merging to the human images. Based on the maximum value in each heatmap alone, there is some hint that smaller eccentricities may benefit from greater oversampling. There is also the suggestion of smaller filter standard deviations being necessary or acceptable at the lowest oversampling factors.

In order to choose an oversampling factor suitable across filter standard deviations, eccentricities and quadrants, the mean $F1$ score across eyes and filter standard deviations for a given oversampling factor was calculated. Fig 6.8 shows these values plotted against oversampling factor for each algorithm. For the no-filter and gaussian filter approaches, the effect of oversampling factor on mean $F1$ is similar for each eccentricity/quadrant combination: oversampling by a factor of two improves algorithm performance relative to no oversampling, but further increases in the

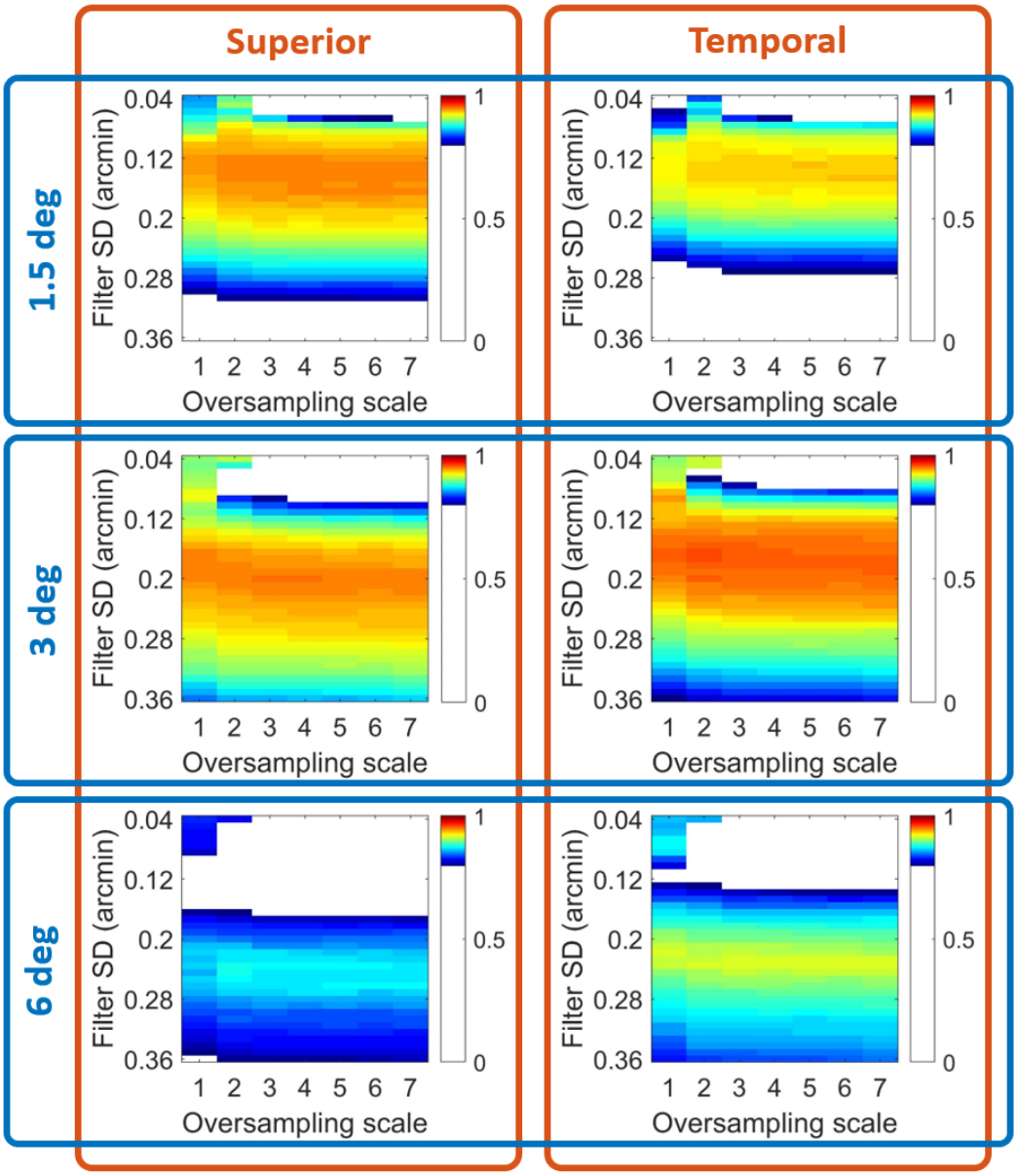


Figure 6.7: Heatmaps showing the mean $F1$ scores across eyes for the laplacian of gaussians convolution algorithm applied to images of human retina with final cone positions unmerged. Heatmaps are presented for combinations of retinal quadrant (columns: superior and temporal) and eccentricity (rows: 1.5, 3 and 6 deg). Individual heatmaps show the variation in $F1$ score by oversampling factor (x axis) and standard deviation of the convolution matrix (y axis). The colour scale is compressed to reveal small differences in the range of $F1$ scores from 0.8 to 1. A small interaction between oversampling factor and standard deviation is observed in that the lowest factors can tolerate lower standard deviations.

oversampling factor do not lead to additional gains in performance. Further analysis with these algorithms was conducted solely using a factor of two. For the laplacian of gaussians algorithm, oversampling has, on average, little or an adverse effect on performance. However, the underlying curves for the three eccentricities tested show that for 1.5-deg images (blue lines), an oversampling factor of two confers a performance benefit compared to the other factors tested. For further analysis the laplacian of gaussians algorithm used no oversampling.

To choose a filter standard deviation for the gaussian and laplacian of gaussians algorithms, the mean $F1$ score across eyes was plotted against filter standard deviation for each algorithm,

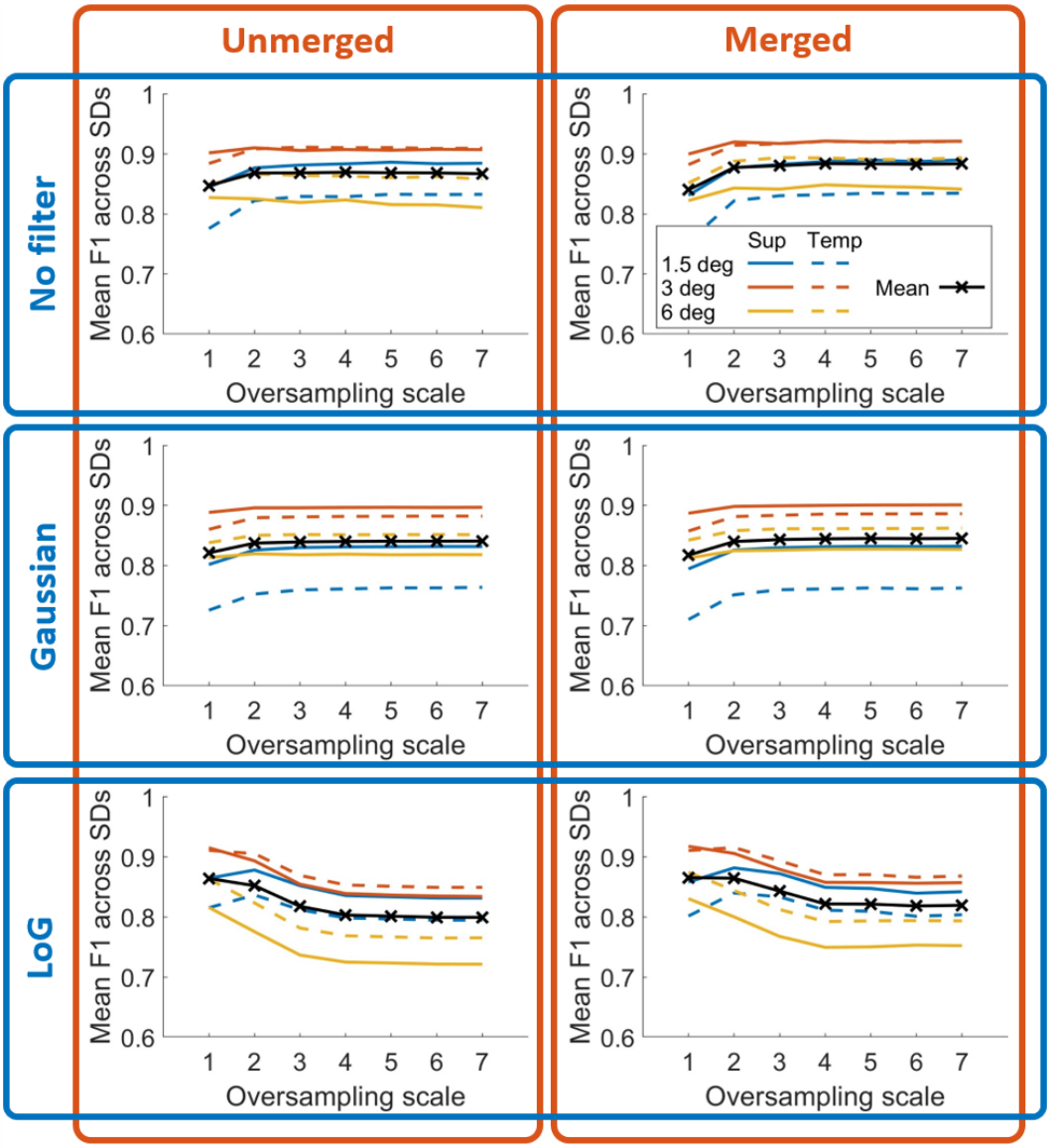


Figure 6.8: Effect of oversampling factor on algorithm performance. Plots are presented for combinations of cone merging status (columns: unmerged and merged) and filter type (rows: no filter, gaussian and laplacian of gaussians (LoG)). Mean $F1$ score across eyes and, if applicable, filter standard deviation is plotted against oversampling scale, broken down by eccentricity and quadrant (coloured lines) and averaged across these two factors (black crosses, black line). For the algorithms using no filter or a gaussian filter, resizing by a factor of two increases performance, but further oversampling does not confer any additional benefit. For the laplacian of gaussians program, the effect varies with eccentricity: oversampling by a factor of two increases $F1$ score for 1.5-deg images, has little effect 3 deg, and is detrimental at 6 deg. On average, no benefit is conferred.

eccentricity and quadrant combination. For illustrative purposes, Fig 6.9 shows these plots for superior retina without cone merging only, as the shape of the curves was similar for a given filter type across quadrant and merge status. The maximum of each curve indicates the best filter standard deviation for that condition. Note that the curve maxima exceed those of the no-filter algorithm (crosses) for the laplacian of gaussians filter, but not for the gaussian filter.

To visualise the performance benefit of cone merging for different conditions, $F1$ scores for no merging were subtracted from those for merging for each filter type, eccentricity and quadrant.

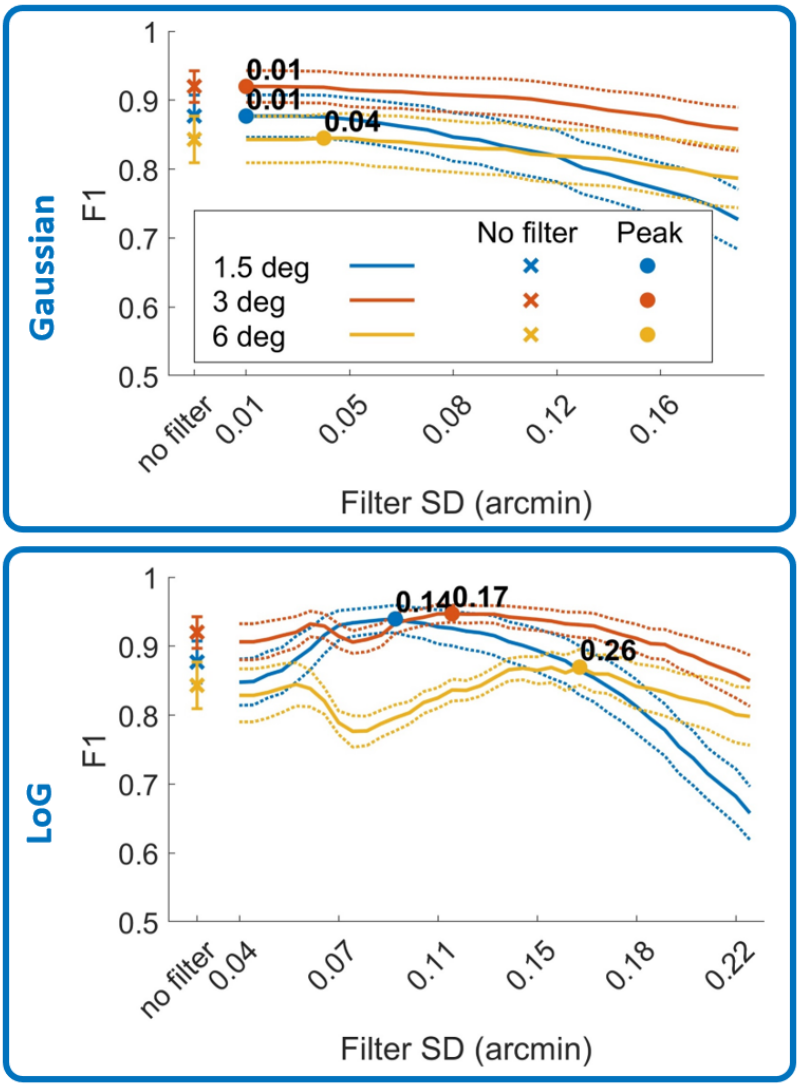


Figure 6.9: Effect of filter standard deviation on performance for gaussian and laplacian of gaussians (LoG) filtering algorithms with cone merging for superior retina. The $F1$ score for each line and marker represents the mean $F1$ across eyes for that condition, which is plotted against filter standard deviation. The maximum $F1$ score for each curve is labelled with its corresponding standard deviation, and the $F1$ scores for the associated conditions for the no-filter algorithm are also plotted on each graph for comparison. It can be seen that the application of a gaussian filter confers little advantage for our images compared to not using a filter, and that increasing standard deviation usually worsens performance. In contrast, the laplacian of gaussians filter confers an advantage provided that its standard deviation is suitable. Plots for unmerged cones and/or temporal retina are extremely comparable, and are therefore not presented here.

The bar charts in Fig 6.10 (left panel) show that for no filter or a gaussian filter, merging benefits performance at 3 and 6 deg, and does not negatively impact it at 1.5 deg. Thereafter, these filters were used with cone merging. For the laplacian of gaussians filter, merging either did not benefit or negatively impacted performance, motivating the use of this filter without merging. Note that the performance benefit values for 3 and 6 deg were always higher than those for 1.5 deg regardless of filter type, perhaps because at the lowest eccentricity, where cones are closer together, the algorithm was inappropriately merging nearby cones.

As it was desirable to choose a single merging approach across all imaging conditions for a

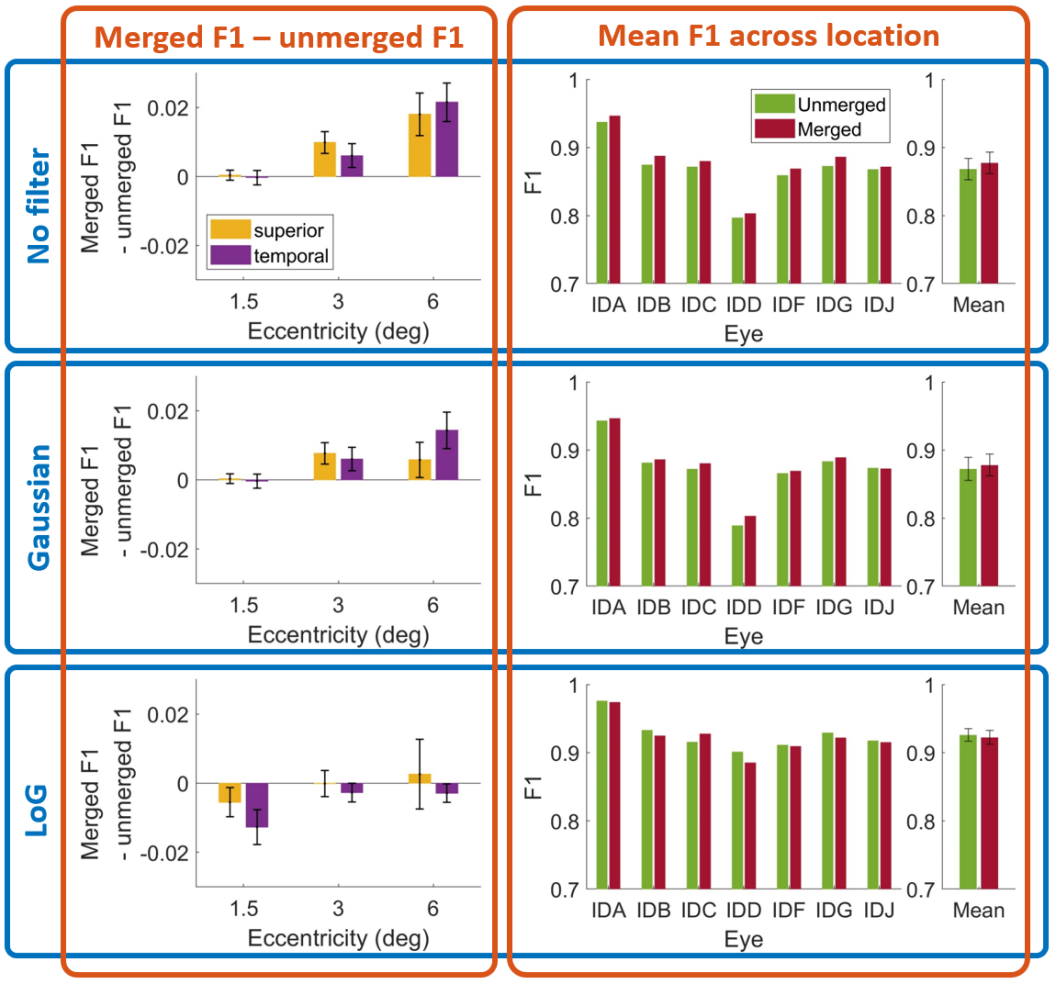


Figure 6.10: Effect of merging implausibly close cones on mean performance ($F1$ score). Left panel: Bar charts showing the advantage conferred by merging cones following the application of each of the filter types (rows: no filter, gaussian and laplacian of gaussian). The y axis represents $F1$ score following cone merging minus the $F1$ score before cone merging and values are given separately for each eccentricity (x axis) and quadrant (yellow bars, superior; purple bars, temporal). Error bars represent the standard error. For the no-filter and gaussian conditions merging shows a small but consistent advantage across eccentricities and directions from the PRL, but for the laplacian of gaussians algorithm cone merging has no or a detrimental effect.

Right panel: Bar charts showing the mean performance (y axis: $F1$ score) across all six imaging locations (three eccentricities in two quadrants) for each of the seven eyes (left subpanel), and the means across the eyes (right subpanel). Error bars cannot be shown for a given eye as the six values that are averaged to produce each bar are not independent of one another, because each of the six values shares grouping factors of eccentricity and direction from the fovea with other values in the average. Error bars for the mean $F1$ score across eyes (right subpanel) represent the standard error. The left subpanel shows that the advantage conferred by merging is consistent across all eyes for the no-filter and gaussian algorithms.

given algorithm, for each eye the algorithm's performance was averaged across the six locations (three eccentricities in two quadrants; see Fig 6.10, right panel, left subpanel). The difference in $F1$ scores obtained with and without merging was assessed separately for each filter type. All contained outliers, but the distribution of the differences between the merged and unmerged groups was approximately symmetrical in shape for all filter types, so the non-parametric Wilcoxon signed-rank test was used. For the no-filter algorithm, the merged output (mean 0.868) performed significantly better than the unmerged output (mean 0.877, $p = .018$) with a matched-pairs rank-

biserial correlation (a non-parametric estimate of effect size) of $r = 1$, indicating that the merging step improved the algorithm's performance for all eyes tested. For the gaussian algorithm, the merged output (mean 0.872) performed significantly better than without merging (mean 0.878, $p = .028$) with $r = 0.714$. For the laplacian of gaussians algorithm, the outputs with and without merging (means of 0.926 and 0.922) were not significantly different ($p = .176$). No corrections for multiple comparisons were made because the decision to merge or not merge cones as a final step was made separately for each algorithm. Therefore, the no-filter and gaussian filter algorithms proceeded with the cone merging step and the laplacian of gaussians algorithm did not.

Fig 6.11 shows the success of the different filter types for the six retinal locations. The no-filter algorithm had a mean $F1$ score of 0.877; the gaussian algorithm, 0.878; and the laplacian of gaussians, 0.926. All three groups contained outliers, defined as more than three scaled median absolute deviations (MAD) away from the median. Therefore, a Friedman test was performed. The difference in the performance of different algorithms was statistically significant, $\chi^2 = 11.185$, $p = .004$. In order to find out which filter or filters were driving the significant result, post-hoc tests were run. The distribution of the differences between the two related groups was not the same, so a sign test was used for the post-hoc tests instead of the Wilcoxon signed-rank test. Following a Bonferroni correction, the α level was set to .0167. The difference between the no-filter and gaussian algorithms was not significant ($p = .688$), but the differences between the laplacian of gaussians and each of the other two filters was significant (both $p = 0.016$). For both comparisons, the matched-pairs rank-biserial correlation was $r = 1$, meaning simply that the laplacian of gaussians filter outperformed the other two algorithms for 100% of eyes.

Cone statistics

The optimum filter standard deviation for each of the six retinal locations can be seen in Fig 6.12. Within both superior (yellow bars) and temporal (purple bars) retina, the optimum standard deviation increases with eccentricity. Across eccentricities, superior retina calls for wider filtering than temporal retina.

A repeated-measures two-way ANOVA could not be run on the data as there were some outliers, so instead the non-parametric ANOVA-type statistic (ATS) from the R package nparLD (Noguchi et al., 2015) was conducted to examine the effect of eccentricity and direction from the fovea on optimum filter width. There was a significant effect of eccentricity, $F(1.46, \infty) = 19.9$, $p < .001$, but no significant effect of direction, $F(1.00, \infty) = 2.12$, $p = .146$, and no significant interaction, $F(1.04, \infty) = 0.424$, $p = .523$.

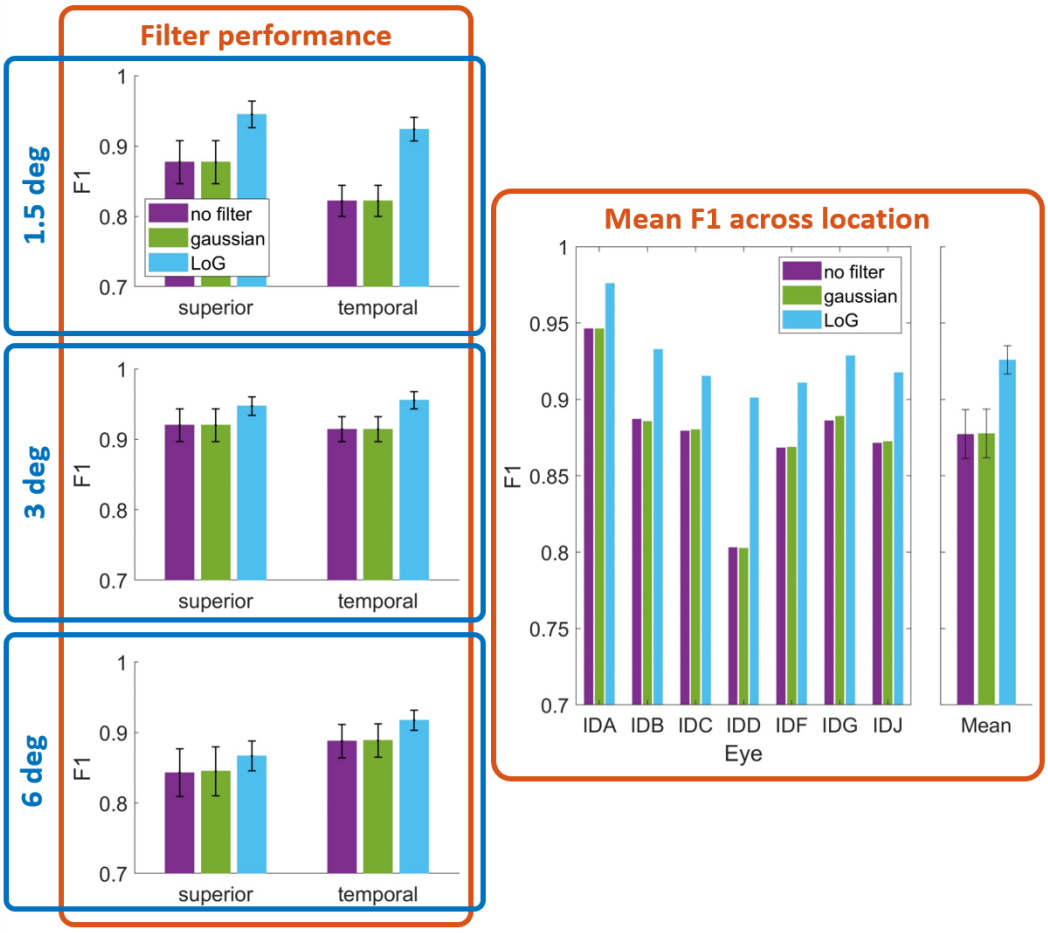


Figure 6.11: Effect of different filters on mean performance ($F1$ score). Left panel: Bar charts at the three eccentricities comparing mean $F1$ scores across the seven eyes between optimised algorithm types, separately for superior and temporal retina. Error bars represent the standard deviation of $F1$ scores across the seven eyes. The laplacian of gaussians (LoG) filter (blue bars) performed better than either no filtering (purple bars) or the gaussian filtering (green bars). Mean $F1$s for the no-filter and gaussian algorithms are consistently comparable to one another, while the LoG algorithm $F1$s are consistently higher than either of the other two methods. Note that at an eccentricity of 3 deg, $F1$ scores for superior and temporal retina are similar. Considering the algorithms with no filter or a gaussian filter, $F1$ scores are depressed for temporal retina at 1.5 deg (where cone spacing should be the smallest of all the imaging conditions) compared to superior retina at the same eccentricity (due to lower recall, data not shown), while the performance of the LoG algorithm suffers less from this effect. For all algorithms, $F1$ scores are also depressed for superior retina at 6 deg (where cone spacing should be the largest of all the imaging conditions) compared to temporal retina at the same eccentricity (due to lower precision and lower recall, data not shown).

Right panel: Bar charts showing the mean performance (y axis: $F1$ score) across all six imaging locations (three eccentricities in two quadrants) for each of the seven eyes (left subpanel), and the means across the eyes (right subpanel). Error bars cannot be shown for a given eye as the six values that are averaged to produce each bar are not independent of one another. Error bars for the mean $F1$ score across eyes (right subpanel) represent the standard error. The left subpanel shows that the higher performance of the LoG algorithms is consistent across all eyes.

Fig 6.13 shows line plots for the optimal filter widths and estimates of cone-to-cone spacing, transformed from published histological measurements of cone density (Curcio et al., 1990) using the equation published by Zhang et al. (2015). Optimal filter width (square markers, solid lines) and cone-to-cone spacing (circular markers, dashed lines) increase with eccentricity in both superior (yellow) and temporal (purple) retina. Both optimal filter width and cone-to-cone spacing are greater for superior compared to temporal cones at each eccentricity. A least-squares linear

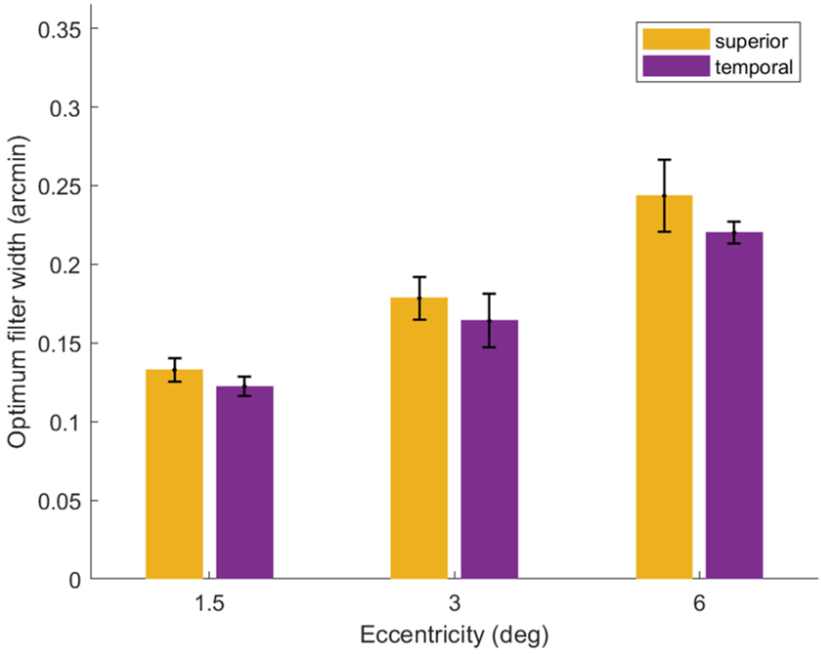


Figure 6.12: Bar chart showing the mean optimum laplacian of gaussians filter standard deviation (SD) across the seven individual eyes, separated by eccentricity and direction from the PRL. The laplacian of gaussians algorithm was applied using an oversampling factor of one and without merging nearby cones as a final step, as these had been identified as the best oversampling and merging parameters for this filter type. Error bars represent the standard error of the best filter SD across eyes. Optimal filter width increases with eccentricity, and superior retina (yellow bars) calls for a larger filter than temporal retina (purple bars) at all eccentricities tested.

Note that these mean filter SDs are not the same as those arrived at earlier (see Fig 6.9). Here, the filter SD with the best $F1$ score for each eye is found, then these filter SDs are averaged, because the aim is to investigate the underlying statistics of the retina and also because allows the calculation a filter SD SE. In contrast, for Fig 6.9, the $F1$ scores for each filter SD are averaged across eyes to yield the filter SD with the best mean $F1$ score, because the purpose is to optimise the algorithm.

regression provides the following equation:

$$\sigma_{LoG} = 0.103 s_{cc} - 0.040 \tag{6.4}$$

where σ_{LoG} is the optimal filter width (arcmin) for the laplacian of gaussians algorithm and s_{cc} is the estimated centre-to-centre spacing of cones (arcmin) at a given eccentricity. Optimal filter width (square markers, solid lines) and optimal filter width predicted from cone-to-cone spacing (circular markers, dashed lines) are plotted in Fig 6.14 for superior (yellow) and temporal (purple) retina. Predicted and actual optimal filter width are similar in each case, but the predicted values are an overestimate at 3 deg and an underestimate it at 6 deg.

The advantage of transforming cone spacing to filter width rather than eccentricity to filter width is that the relationship between the two is likely to be more linear as they are likely to share similar underlying nonlinearities.

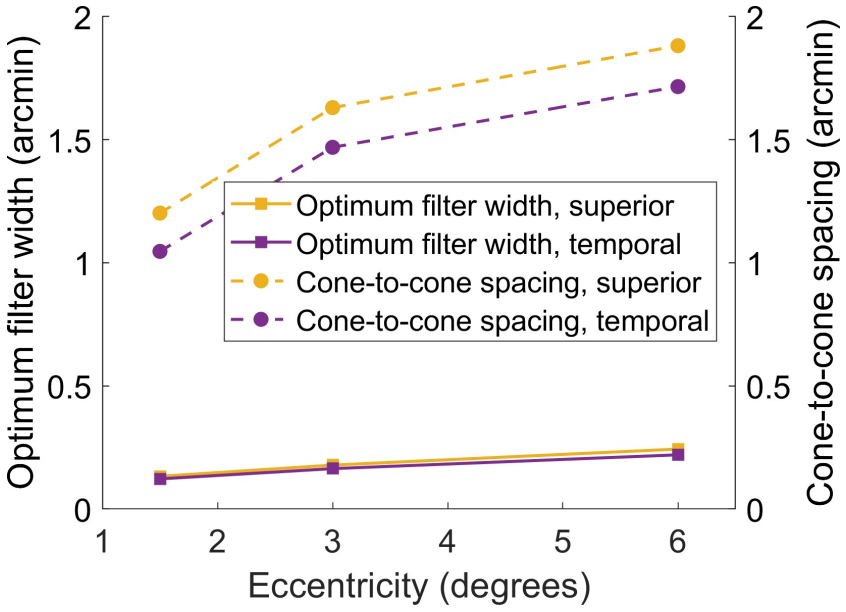


Figure 6.13: Plot comparing the optimum filter width for the laplacian of gaussians algorithm (square symbols) with estimates of cone-to-cone spacing derived from histological measurements of cone density (Curcio et al., 1990) and converted into spacing (Zhang et al., 2015; circular symbols).

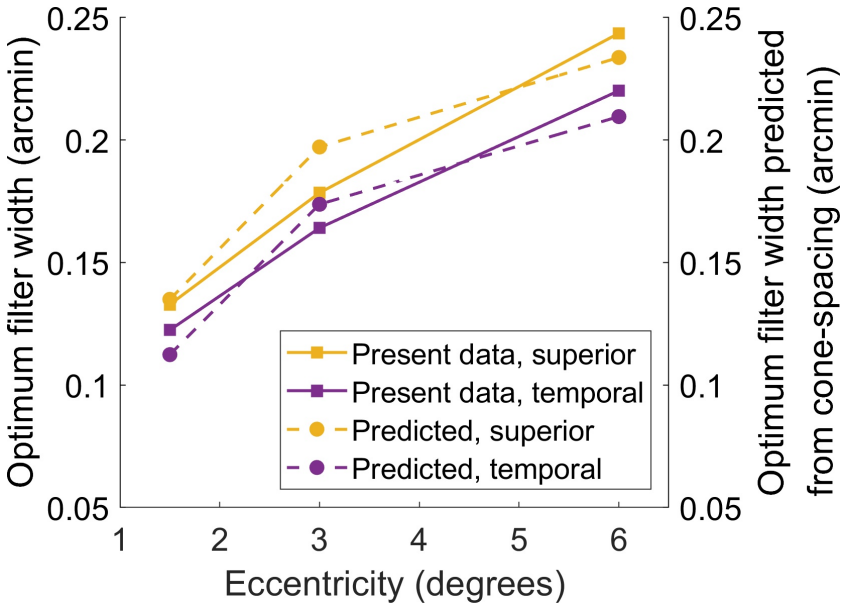


Figure 6.14: Plot showing the effect of transforming histological measurements of cone density (Curcio et al, 1990) into predicted optimum filter width for the laplacian of gaussians algorithm (circles, dashed lines) alongside the obtained optimum filter width (squares, solid lines).

6.4 Discussion

Automatic cone detection methods either require parameter intervention (non-machine learning) or the creation of a gold standard (human-labelled dataset) for each instrument (machine learning). To date, parameter intervention has not been guided by a gold standard, only tested against it (Chiu et al., 2013; Li & Roorda, 2007). Nor have the quality of the gold standards been systematically investigated.

6.4.1 Manual localisation

In this work, manual cone localisations were obtained from six members of the Oxford Perception Lab who were experienced with AOSLO cone images. The cone localisations of each manual rater were compared with the set of cone coordinates agreed upon by five of the six raters to determine each rater's performance ($F1$ score). The exclusion of one manual rater from the final gold standard of cone coordinates due their low $F1$ score is indicative that not all raters (even amongst those experienced with AOSLO cone images) agree with their own colleagues on the definition of a cone. While it is true that this rater was the least experienced of the observers (as assessed by time since joining the lab before data collection), it should be noted that judgements on the appearance of cones develop through repeated exposure to and interaction with the judgements of others, and members of different labs may be more likely to make different judgements to one another. This arises through the lack of a ground truth of cone locations and the multidimensional criteria used by human raters to determine the location of cones. Such issues could be addressed by formal training: observers could be asked to manually rate images that have published manual cone localisations or simulated images that have a true ground truth, comparing their performance with the manual gold standard or simulated ground truth at regular intervals until their performance is of a high level.

6.4.2 Automatic cone localisation

Final performance

Automatic cone localisation algorithms with a range of parameters were applied to a set of AOSLO images collected at three eccentricities along both the horizontal and vertical meridians. The algorithms were optimised with respect to oversampling factor, filter standard deviation (if applicable) and the inclusion of a merging step for cone coordinates that were implausibly close together. The algorithm using a laplacian of gaussians filter significantly outperformed those using a gaussian or no filter, with an $F1$ score around 0.05 higher.

The performance of the gaussian algorithm is, at first glance, puzzling, in that its performance is no different to using no filter at all and that the optimum gaussian width is so low (indeed, following pilot testing of parameters, the highest filter width of 0.18 arcmin tested here was lower than standard filter width of 1 μm, or approximately 0.20 arcmin, of Li and Roorda (2007)). It is impossible to make a direct comparison between the performance metrics presented here with those obtained for the Li and Roorda (2007) method by Li and Roorda themselves or by Mariotti and Devaney (2015), in part because they each used different, poorly-defined performance metrics, but

perhaps more importantly because the former used images gathered on a different instrument to the images in the present study and the latter group used simulated images. The six image examples given by Li and Roorda appear more noisy than the present images, which would indeed require smoothing to avoid marking flecks of noise as cones. That the gaussian filtering here fails to improve performance relative to using no filter at even extremely small filter widths may speak to the balance that must be made in noisy images between smoothing away noise and blurring cones together: for Li and Roorda's noisier images there should be what Mariotti and Devaney refer to as the 'operating point' where further increase in the filter width begins to add false negatives at a faster rate than it reduces false positives. In contrast, the present images comprise many frames averaged together, resulting in much smoother images that have nothing to gain from the smoothing achieved by a gaussian filter. They can therefore only lose performance with an increase in its width. This speaks to the importance of validating parameters against multiple datasets from different groups when attempting to develop a widely-applicable cone detection algorithm.

For the images presented here, the optimised laplacian of gaussians algorithm significantly outperformed the other two filter types. Examining the underlying data more closely, Fig 6.11 (left panel) suggests that this may be true at all eccentricities. The advantage of the laplacian of gaussians approach is most marked at 1.5 deg in spite of the fact that the lack of oversampling used with this algorithm is not ideal at this eccentricity (see Fig 6.8, bottom panel, blue lines). The fact that the benefit is reduced at higher eccentricities is perhaps due to the nature of the advantage of the laplacian of gaussians: its strength may lie in the way it uses the dim surrounds of cones to separate cones that appear to touch from one another, and cones appear to touch less often with increasing eccentricity as cone density decreases. It is notable that the greatest performance advantage of the laplacian of gaussians occurs for the *temporal* 1.5-deg images, the group of images where the cones should be closest to one another. This is driven by depression of the $F1$ scores of the other two filter types. Fig 6.10 (left panel) indicates that this is not due to the merging step inappropriately merging nearby cones at low eccentricities. It may instead be explained by the lesser effectiveness of the gaussian and no-filter methods at separating touching cones at low eccentricities, even with the aid of bicubic interpolant oversampling described in § 6.4.2.

Parameter optimisation

For the no-filter and gaussian algorithms, and for the laplacian of gaussians algorithm at 1.5 deg, upsampling the images improved cone localisation performance (see Fig 6.8). Mariotti and Devaney (2015) attribute the performance benefit conferred by upsampling to improved determination of

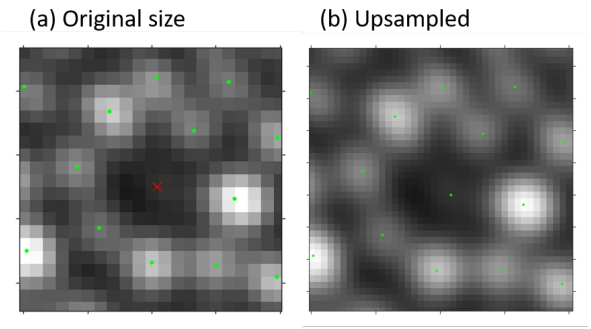


Figure 6.15: AOSLO cone image taken at an eccentricity of 1.5 deg at (a) its original size, and (b) upsampled by a factor of two. The images have been processed using the no-filter algorithm which has correctly detected all of the cones in (b) and most of the cones in (a; green dots) but has missed one cone in (a; red cross).

cone positions. In the case of algorithms with cone merging, imprecise cone position could lead to missed or inappropriate links between the test and reference coordinates. Regardless of the use of merging, oversampling using cubic (Mariotti & Devaney, 2015) or bicubic (the present study) interpolation can reveal the peak of a dimmer cone against a brighter neighbour, as shown in Fig 6.15, because the middle of the dimmer cone, when bicubically interpolated with its relatively bright immediate surroundings, will return a higher value than the slim connection with the brighter cone which has dimmer pixels above and below it. It stands to reason that the more closely packed cones found at lower eccentricities have more to gain from this: their connections to brighter cones are shorter and are therefore more likely to hide the peak of a dim cone that can then be revealed through oversampling.

Although the decision was taken before parameter optimisation began to use one oversampling factor for all images, it can be seen from Fig 6.8 (bottom panel) that retinal eccentricity affects the optimum factor for the laplacian of gaussians method: oversampling improves algorithm performance at 1.5 deg but makes no difference or actually impedes accurate detection at higher eccentricities. A direction for further analysis of the data presented here may be to find an optimum minimum cone separation in pixels such that the oversampling factor can be adjusted based on metrics extracted from the image.

Despite averaging $F1$ across eyes for each filter width, the resulting curves (see Fig 6.9) still show evidence of noise, which presents problems for simply using the filter width with the maximum $F1$ score. The tops of most of the curves are fortunately relatively flat, pointing to a degree of robustness to small deviations from the ideal filter width, and smoothing would require a further experimenter intervention as to the method of smoothing. But the choice of the best filter width may yet benefit from smoother curves, and a parameter decision analogous to the necessary degree of smoothing (the filter width step size) has already been taken by the experimenter.

For the laplacian of gaussians algorithm, the decision was taken not to include the cone merging

step due to the negative effect of merging at 1.5 deg and its lack of benefit at higher eccentricities. It is interesting that the performance reduction at 1.5 deg did not occur when the image size was doubled (data not shown), in line with the other filter types at that degree of upsampling. This suggests that when the cones are closer together in terms of pixels, merging runs the risk of incorrectly combining separate cones.

6.4.3 Cone statistics

The maximum of the laplacian of gaussians algorithm's filter width-performance curve is expected to be the point at which the laplacian is small enough that it enhances rather than merges neighbouring cones and large enough that it smooths away high-frequency noise. As expected, this maximum increased with eccentricity, and therefore decreased with cone density (Curcio et al., 1990). Furthermore, as cone density falls off more rapidly along the vertical meridian, the optimal filter width should be greater for the temporal images compared to superior. This was the direction of effect that was found consistently across eccentricities, although this could not be confirmed as a statistically significant difference.

6.4.4 Future directions

Manual localisation assessment

To assess manual raters' self-consistency we asked them to rate a set of retinal patches five times, the images being rotated and reflected at random to reduce the chances of mere replication of remembered ratings. Analysis of these data is beyond the scope of this chapter, but the approach would echo that used to make the gold standard of cone coordinates: for each patch, coordinates would be linked together and the percentage of times a coordinate was identified as a cone would be calculated. A rater's average percentage would indicate their overall self-consistency. These data would help in the evaluation of the performance of an automatic algorithm, with the average or highest self-consistency of expert raters representing good performance.

Furthermore, these data could be used together with the percentage agreements used in the creation of the gold standard to uncover factors that may drive uncertainty within and between raters, such as noise, intensity, proximity to neighbouring cones and local anisotropy.

As no histological ground truth is available for the human images, manual ratings were also collected for simulated images. These ratings can be compared to the simulation ground truths, and differences between the manual gold standards and simulation ground truths will be useful in developing better simulations and in furthering discussions about the appearance of cones in

AOSLO images, for example in the case of 'dark' cones (Bruce et al., 2015).

Algorithm

This chapter has presented an algorithm that is more effective at localising cones in images taken by the Oxford AOSLO than a similar method derived from Li and Roorda (2007) and has offered an equation to convert the cone-to-cone spacings derived from Curcio et al. (1990) by the equation in Zhang et al. (2015) to an optimum standard deviation for this algorithm.

This equation may be improved by refining the extraction of the optimum filter width from the eccentricity-filter width data presented in Fig 6.9 by fitting curves to the data to smooth any remaining noise, and by fitting curves to the cone spacing data derived from Curcio et al (Curcio et al., 1990) rather than linearly interpolating to arrive at values for eccentricities intermediate to those directly measured.

The aim of this line of work is to provide an algorithm that does not require parameter intervention by the experimenter or the creation of further gold standards. To this end, the algorithm and eccentricity-filter width equation presented here should be tested against existing cone image/gold standard sets that have been made available by other labs. It would be particularly interesting to discover if the equation generalises to images collected outside the eccentricity range tested here.

Finally, the ideal algorithm would be able to extract its optimum filter width from the image itself to account for inter-individual variation in cone density at a given eccentricity. Therefore, the relationship between a range of image metrics and the optimum filter width for an image should be studied on an image-to-image basis, rather than collapsing across participants. The most promising such metric is Yellott's ring, the peak in the Fourier spectrum that corresponds to the modal cone spacing (Cooper, Langlo, Dubra & Carroll, 2013).

Chapter 7

General discussion

This thesis focuses on colour processing by the human visual system. Chapter 2 uses an adaptation paradigm to investigate the possibility of separable neural channels carrying signals for hue and saturation. Later chapters consider the impact of differences in the relative numbers of long- and middle-wavelength cone photoreceptors (L:M cone ratio) on spatiochromatic vision. Differences between and within observers are considered in light of vision as an active process, constantly resampling the scene with eye movements. Further work on these issues would benefit from improvements to both the automated localisation of cones in adaptive optics (AO) retinal images and the generation of manual benchmarks against which the algorithms are tested; this is addressed in Chapter 6.

7.1 Hue and saturation

Thresholds for hue discrimination had been found to be smaller than those for saturation when arranged in two-dimensional colour spaces (Munsell and the biologically relevant MacLeod-Boynton space) and measured away from the neutral point (Judd, 1968). This had given rise to the hypothesis that hue and saturation may be carried in separate neural channels; however, this had not been directly tested. Chapter 2 introduced a novel method of modulating hue or saturation in a way that holds the time-averaged excitation of the cone cells and cardinal channels constant and allows hue and saturation stimuli to be expressed in common units. By measuring thresholds for hue or saturation modulation following adaptation to suprathreshold hue or saturation modulation, evidence for separate neural channels was directly sought but not found.

The stimuli used to selectively adapt and test hue or saturation were generated by modulating the two cardinal channels in different phase relations. This phase-combination adaptation approach has opened up a new way forward to study hue and saturation. In its original form, it could

be used to further investigate the potential for separate neural channels for hue and saturation by taking measurements of suprathreshold, rather than threshold, hue and saturation discrimination thresholds following adaptation to hue and saturation stimuli that circle or pass through the neutral point. Alternatively, phase shifts of the modulation of the S/(L+M) axis by 0 and 180 deg relative to that of the L/(L+M) axis, both of which constitute saturation changes when operating through the neutral point, could be made relative to points displaced along 45-, 135-, 225- and 305-deg lines from neutral to create hue and saturation adaptation stimuli to be presented prior to test measurements in that local area of MacLeod-Boynton space.

Recent work (Danilova & Mollon, 2016, 2020), including ours, has found that hue super-importance does not hold throughout colour space, and where it does occur, it varies in magnitude. Studies show lower thresholds for saturation than for hue when measurements are made close to the neutral point in the upper right or lower left (our 0-deg phase shift, diagonal right line) of MacLeod-Boynton space. The two Danilova and Mollon studies also find that measurements made away from the neutral point in the upper left and lower right quadrants do show that hue thresholds are lower than saturation thresholds. Danilova and Mollon (2020) suggest that the difference between their own measurements and those of Judd (1968) lies in how they were made: the stimuli for the modern measurements were self-luminous stimuli displayed on monitors, while the historical measurements were made using slips of Munsell paper. Like any surface that is not entirely matte, Munsell paper should show specular highlights such that the colour reflected ranges from a point near the illuminant towards the surface colour, a variation in reflected light termed 'Mongean noise'. This would increase noise along lines of saturation and potentially elevate saturation thresholds more than hue threshold when compared to the modern, self-luminous stimuli. To determine whether Mongean noise could account for the magnitude of hue super-importance found in the earlier measurements, the modern measurements for static chromatic discrimination and modulation detection, both pre- and post-adaptation, could be repeated with rendered images of pieces of paper with differing degrees of specularity, and therefore of Mongean noise.

7.2 L:M ratio

To date, the wide variation in human L:M cone ratios had been found to have only limited perceptual consequences: in elevated thresholds for red-green contrast for those with very asymmetrical L:M ratios (Miyahara et al., 1998), and a relationship between L:M cone ratios and L:M spatiochromatic threshold ratios for local targets presented to the parafovea (Danilova et al., 2013). The latter experiment had not been attempted in the fovea because there, where packing densities are higher,

'spatial resolution is known to be limited primarily by optical factors (Thibos, 1998)' (Danilova et al., 2013, p. 27). Heterochromatic flicker photometry (HFP) was used as an assay for L:M ratio, and thresholds for foveally-presented L- and M-isolating optotypes were gathered to determine foveal L:M spatiochromatic threshold ratios. This was undertaken in 100 participants, replicating the range of L:M ratios expected in the local population.

Until recently, measurements of or relating to L:M ratio have relied wholly or primarily on white observers. Luminous efficiency functions such as V(λ) (Stockman et al., 2008) and Y in the CIE 1931 xyY colour space (CIE, 1932), which quantify the effect of a spectrum on the luminance channel of the 'standard' observer, share this reliance, and their use extends from scientific research to display manufacturing. However, mean estimated L:M ratio in males of African descent is significantly lower than that of Caucasian males (McMahon et al., 2008), and may also be lower for Asian observers (Yamauchi et al., 2013). The longstanding assumption that the data from Caucasians holds true across ethnic groups has resulted in estimates that are not well-representative of much of the world's population. The data obtained from the small number of Asian/Asian British participants in the experiments of Chapter 3 are not inconsistent with the conference report of Yamauchi et al. (2013), and encourages the expansion of this sample of opportunity to a deliberate sample of 30 observers of Asian descent.

The data presented in Chapter 3 do, in fact, demonstrate that L:M ratios impact spatiochromatic vision in the fovea in spite of the fact that a greater share of the sampling limitation would derive from optical factors compared to the parafovea. L- vs M-isolating threshold ratios for both size and contrast were correlated with L:M ratios. One possibility is that L and M spatial resolution should be linearly related to the cone-to-cone spacing of the relevant cone class. While instrumental limitations in the present study prevented this from being evaluated at low L:M ratios, the data at higher ratios were consistent with a linear relationship between class-specific cone-to-cone spacing and L and M size thresholds up to an observer's optical acuity.

The correlations between L:M ratios and spatiochromatic threshold ratios were moderate ones. Partly this can be attributed to the choice of primaries in the HFP apparatus in which the separation of higher, more asymmetrical L:M ratios were prioritised at the expense of lower and more even ones. However, factors affecting the relationship at more asymmetrical ratios have yet to be explained. Spatiochromatic threshold ratios corresponding to higher flicker-photometrically measured L:M ratios were compressed towards more balanced threshold ratios. This could be attributed to the limits of optical acuity, to gain control, or to some additional measure that brings the spatiochromatic output of the visual systems of those with the most imbalanced L:M ratios

towards the central tendency.

7.3 Eye movements and L:M ratio

One potential such mechanism could be fixational eye movements (FEM). Until recently, the study of visual perception has not generally considered the motion of the eye, which constantly translates the stimulus across the retinal mosaic even during fixation. However, it is now known that not only do FEM improve acuity for high-spatial frequency targets, but in some cases they are actually tuned to the requirements of the task (Intoy & Rucci, 2020).

The L and M cones offer a uniquely interesting avenue for the exploration of the functional role of FEM. In a di- or trichromatic retina, any given class of cones will have a lower sampling resolution than that of the overall cone mosaic. Therefore, for stimuli that are predominantly visible to one class of cones in particular, the extraction of the image may be expected to benefit more from resampling than in the case of black-and-white images visible to all classes. Such advantages may occur when cones from the class in question do not immediately fall on an informative part of the stimulus and FEM translate the stimulus into a better position; they may also occur from the mere resampling of the image to increase the overall number of relevant receptors that have passed over the image, or by moving at a rate closer to the most effective temporal frequency to extract the spatial characteristics of the target (Intoy & Rucci, 2020). The particular power of studying L and M cones lies in (a) their imbalanced relative numbers in the general population, meaning that on average at a population level any such effect should be differentiable between the two classes, and (b) the huge variation in L:M ratio across the population, allowing individual differences to be studied.

Chapter 5 advances the use of adaptive optics (AO) retinal imaging as a tool for recording and studying FEM (Vogel, Arathorn, Roorda & Parker, 2006; Young et al., 2017). Using its unprecedentedly fine spatial resolution, the technique has been leveraged to obtain measurements of eye movements during a spatial discrimination task with small, local stimuli that target either the L or the M cones.

Although the regression of eye movement on HFP (or L:M ratio) was not significant, the amount of eye movement increased with increasing HFP score; that is, it decreased with increasing L:M ratio. As performance on a high-acuity task has been seen to peak with a lower-than-average amount of eye movement (Agaoglu et al., 2018), the lower amount of FEM in participants with L:M asymmetry may reflect the importance of smaller or more precise eye movements in those with sparse M cone mosaics. Interestingly, in spite of one study's finding that FEM can be closely tuned to a task's spatial requirements (Intoy & Rucci, 2020), there was no significant relationship between the difference in

eye movement on L- and M-isolating trials and HFP score (or L:M ratio), indicating that FEM do not differ when stimuli target cone classes of different sampling densities.

In the literature to date on FEM and acuity, it has been shown with retinal stabilisation that some movement is better than no movement (Rucci et al., 2007) and that the amount of movement leading to optimal performance is rather less than the average amount (Agaoglu et al., 2018), while a magnitude, speed and curvature of movement that seems to have functional relevance is more likely during a particular task (Intoy & Rucci, 2020). However, no attempt has yet been made to directly compare the different amounts of FEM that occur naturally with observer performance. Intoy and Rucci drew specific conclusions that a certain peak temporal power of drift and amplitude of microsaccades were important for performance on a high-acuity task, and that the task elicited such eye movement attributes. However, they did not actually test whether those values were associated with improved performance when compared with higher or lower levels of peak temporal power or microsaccade amplitude. In contrast, Chapter 5 directly tested the hypothesis that the median amount of eye movement that occurs during the task is associated with the highest performance. However, there was no evidence for better performance at the median levels of eye movement, or at the lowest or highest amounts of eye movement. This is, instead, consistent with the study of Agaoglu et al. (2018), which suggested that peak performance should occur with an amount of eye movement less than the observer's average quantity.

The work presented here on spatiochromatic resolution and eye movements tests, to my knowledge, the strongest hypothesis advanced to date regarding functional FEM. It asks not only whether such movements improve resolution, and whether particular stimuli elicit certain degrees of movement, but also: Are FEM tuned to the specific biological needs of an observer, adapting to counteract the greater pressure on class-specific sampling that may be expected in retinas with particularly skewed L:M cone ratios? Furthermore, it represents one of the most ambitious experiments to date in terms of the number of individuals tested (20) and analysed (10) using AO retinal imaging. The largest group of new observers in an AO study of which the author is aware is eight, in a study of direct measurements of L:M cone ratio with an AO flood system (Hofer et al., 2005). Developing the analysis pipeline for AO retinal images is an ongoing challenge both in the Oxford Perception Lab and in other labs internationally, and it is hoped that at a future date, it will be possible for the lab to analyse the data of additional participants from the dataset collected for this thesis.

7.4 Cone localisation: The gold standard and an automated algorithm

7.4.1 Gold standard for cone detection

Performance of automatic cone detection algorithms appears to vary substantially when applied to different datasets. Meanwhile, most such algorithms are evaluated – and machine learning algorithms are trained – against what is currently considered the gold standard: manual localisation by a human expert in retinal images. Not only has this gold standard not been critically evaluated, but it has never yet been questioned by any authors using manual localisations for this purpose.

However, Chapter 6 found that at least one of the six experts, who all rated the same set of cones and were from the same lab, had a substantially lower $F1$ (performance) score when compared to the manual localisations of the other raters. It notes that our ideas about the appearance of a cone arise through the exposure to and interaction with the ideas of others, and therefore that there may be even wider discrepancies between manual localisations by raters from different labs. Indeed, some of the performance differences seen when a given algorithm is tested against images from different instruments may arise not only from differences in the images, but also from the manual gold standards against which algorithm performance is measured. It argues that manual raters should undergo formal training, localising cones in images for which published cone coordinates are available, or in simulated images which have a true ground truth.

7.4.2 Automatic algorithm

This thesis presents an automatic cone localisation algorithm that significantly outperforms the classic algorithm of Li and Roorda (2007) for the images produced by the Oxford AOSLO with a mean $F1$ score of 0.926 compared to 0.878. The development of this algorithm has clarified and opened up exciting further avenues of research and recommendations to other researchers.

First, the future publication of our manually labelled dataset will add to the existing dataset of Garrioch et al. (2012), meaning that there will now be ample data available for reviewers to request the validation of any new or improved automatic cone methods on labelled images from multiple instruments. The algorithm presented here must also be thus tested. Second, the image statistics of false positives and false negatives should be analysed to determine what drives discrepancies between the algorithms and the manual raters: this can not only be used to improve the algorithms such that their outputs more closely match those of the raters, but also to drive discussions among manual raters about what a cone looks like in AO retinal images, as the algorithms may in some cases be more accurate than the gold standard.

Third, the research community should maintain the precedent set by Mariotti and Devaney

(2015) of performance analysis of a given algorithm with and without added noise, and, in the case of the present algorithm, the introduction should be attempted of an image statistic that represents an image's noisiness to the equation that returns the recommended laplacian of gaussians (LoG) standard deviation (SD). Fourth, and also with regards only to the present algorithm, Yellott's ring should be determined for a given cone image and should replace the cone eccentricity term in the LoG SD equation, such that the equation uses the real mean cone spacing in a given image rather than the mean cone spacing predicted from a published mean across eyes. Thus, the LoG automatic cone localisation algorithm could reduce its reliance on external measurements and instead draw its parameters directly from the image to be analysed.

7.5 Conclusions

This thesis takes as its focus the irregularities and complications of the story of trichromatic vision. The three classes of photoreceptor which underpin human trichromacy do not exist in equal numbers; nor can the detection and discrimination thresholds arising from their comparisons be described by a Euclidean colour space. It is here demonstrated that the super-importance of hue is unlikely to be mediated by separate neural channels for hue and saturation. Owing to the variation in hue and saturation thresholds around the neutral point, and to the greater hue super-importance seen when using illuminated than self-luminous stimuli, future attempts to probe underlying channels using our adaptation paradigm should target specific areas away from the neutral point and use rendered images of illuminated surfaces.

It also demonstrates that, in spite of previous assumptions (Danilova et al., 2013), L and M packing densities do in fact affect spatiochromatic thresholds in the fovea. Yet the merely moderate correlation of L:M ratios with foveal spatiochromatic threshold ratios raises the question of what mechanisms alleviate the impact of asymmetry of L:M ratio. When considered in combination with the known impact of eye movements on spatial frequency thresholds (Agaoglu et al., 2018; Rucci et al., 2007) and spatial attributes on eye movements (Intoy & Rucci, 2020), the question arises: are FEM one such mechanism? Visual inspection of the trends of the data presented here do suggest the potential for an effect of L:M ratio on eye movements, yet despite obtaining sample sizes typical of research in this field, the association was not significant. It seems unlikely, in light of their striking importance for achromatic spatial resolution, that FEM would play little or no role in spatiochromatic vision; alternative analysis techniques may therefore shed more light on the question.

Work is currently under way in the Oxford Perception Lab to replicate existing, and develop new,

methods of cone typing. As part of the work to identify a cone's class and to track particular cones in video streams, the locations of large numbers of cones must be automatically and accurately extracted from AO retinal images. The new algorithm developed here, which performs much better for images from the Oxford AOSLO than other classic methods, represents a vital step towards tracking the translation of spatiochromatic stimuli across cones of known classes during FEM.

This thesis tackles two fundamental issues in colour perception: the post-receptoral processing of colour dimensions, hue and saturation; and the initial sampling of the visual scene by a moving, trichromatic cone mosaic. It exploits and progresses new technology, adaptive optics retinal imaging, which yields images of neural tissue and eye movement measurements with extremely high spatial and temporal resolution, and shows great promise in the investigation of the early stages of human colour processing.

References

Adler, F. H. & Fliegelman, M. (1934). Influence of fixation on the visual acuity. *Archives of Ophthalmology*, *12*(4), 475–483. Retrieved from http://archopht.jamanetwork.com/article.aspx?articleid=611153 doi: 10.1001/archopht.1934.00830170013002

Agaoglu, M. N., Sheehy, C. K., Tiruveedhula, P., Roorda, A. & Chung, S. T. L. (2018). Suboptimal eye movements for seeing fine details. *Journal of Vision*, *18*(5), 8. Retrieved from http://jov.arvojournals.org/article.aspx?doi=10.1167/18.5.8 doi: 10.1167/18.5.8

Bergmann, C. (1857). Anatomisches und Physiologisches uber die Netzhaut des Auges. *Zeitschrift fur rationelle Medicin II*, 83–108.

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), 1961–1966. Retrieved from https://www.sciencedirect.com/science/article/pii/S0042698997003027?via%3Dihub doi: 10.1016/S0042-6989(97)00302-7

Borderies, N. & Rautureau, G. (2018). *Touchscreen-Psychtoolbox*. Retrieved from https://github.com/MBB-team/touchscreen-psychtoolbox

Bowers, N. R., Boehm, A. E. & Roorda, A. (2019). The effects of fixational tremor on the retinal image. *Journal of Vision*, *19*(11), 8. Retrieved from https://jov.arvojournals.org/article.aspx?articleid=2751568 doi: 10.1167/19.11.8

Bowmaker, J. K., Parry, J. W. L. & Mollon, J. D. (2003). The arrangement of L and M cones in human and a primate retina. In J. D. Mollon, J. Pokorny & K. Knoblauch (Eds.), *Normal and defective colour vision* (pp. 39–50). Oxford: Oxford University Press. Retrieved from http://vision.psychol.cam.ac.uk/jdmollon/papers/BowmakerParryMollon2003.pdf

Brainard, D. H., Roorda, A., Yamauchi, Y., Calderone, J. B., Metha, A., Neitz, M., … Jacobs, G. H. (2000). Functional consequences of the relative numbers of L and M cones. *Journal of the Optical Society of America A*, *17*(3), 607–614. Retrieved from http://josaa.osa.org/abstract.cfm?URI=josaa-17-3-607http://www.opticsinfobase.org/DirectPDFAccess/12AF6AC3-0542-9E88-9695F3929BD1FD5D_59178/josaa-17-3-607

.pdf?da=1&id=59178&seq=0&mobile=nohttp://www.opticsinfobase.org/josaa/abstract.cfm?uri=josaa-17-3-6 doi: 10.1364/JOSAA.17.000607

Bruce, K. S., Harmening, W. M., Langston, B. R., Tuten, W. S., Roorda, A. & Sincich, L. C. (2015). Normal Perceptual Sensitivity Arising From Weakly Reflective Cone Photoreceptors. *Investigative Opthalmology & Visual Science*, *56*(8), 4431–4438. Retrieved from http://iovs.arvojournals.org/article.aspx?doi=10.1167/iovs.15-16547 doi: 10.1167/iovs.15-16547

Calkins, D. J., Schein, S. J., Tsukamoto, Y. & Sterling, P. (1994). M and L cones in macaque fovea connect to midget ganglion cells by different numbers of excitatory synapses. *Nature*, *371*, 70–72. Retrieved from http://www.nature.com/articles/371070a0 doi: 10.1038/371070a0

Campbell, F. W. & Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *The Journal of Physiology*, *181*(3), 576–593. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1357668/ doi: 10.1113/jphysiol.1965.sp007784

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. *Journal of the Optical Society of America. A, Optics, image science, and vision*, *17*(3), 499–509. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10708031 doi: 10.1364/josaa.17.000499

Carroll, J., Neitz, J. & Neitz, M. (2002). Estimates of L:M cone ratio from ERG flicker photometry and genetics. *Journal of Vision*, *2*(8), 531–542. Retrieved from http://jov.arvojournals.org/Article.aspx?doi=10.1167/2.8.1http://www.ncbi.nlm.nih.gov/pubmed/12678637 doi: 10:1167/2.8.1

Carroll, J., Neitz, M., Hofer, H., Neitz, J. & Williams, D. R. (2004). Functional photoreceptor loss revealed with adaptive optics: An alternate cause of color blindness. *Proceedings of the National Academy of Sciences*, *101*(22), 8461–8466.

Castaño, J. A. & Sperling, H. G. (1982). Sensitivity of the blue-sensitive cones across the central retina. *Vision Research*, *22*(6), 661–673. Retrieved from http://www.sciencedirect.com/science/article/pii/004269898290102Xhttp://www.sciencedirect.com/science/article/pii/004269898290102X/pdf?md5=76dbc0880d2bfd9b1dc84a4a88c0c967&pid=1-s2.0-004269898290102X-main.pdf doi: 10.1016/0042-6989(82)90102-X

Castet, E. & Crossland, M. (2012). Quantifying eye stability during a fixation task: A review of definitions and methods. *Seeing and Perceiving*, *25*(5), 449–469. Retrieved from https://brill

.com/view/journals/sp/25/5/article-p449_5.xml doi: 10.1163/187847611X620955

Cavonius, C. R. & Estévez, O. (1975). Contrast sensitivity of individual colour mechanisms of human vision. *The Journal of Physiology*, *248*(3), 649–662. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1309545/http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1309545/pdf/jphysiol00896-0119.pdf

Chen, S. F., Chang, Y. & Wu, J. C. (2001). The spatial distribution of macular pigment in humans. *Current Eye Research*, *23*(6), 422–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12045892

Chiu, S. J., Lokhnygina, Y., Dubis, A. M., Dubra, A., Carroll, J., Izatt, J. A. & Farsiu, S. (2013). Automatic cone photoreceptor segmentation using graph theory and dynamic programming. *Biomedical Optics Express*, *4*(6), 924–37. Retrieved from http://www.opticsinfobase.org/boe/abstract.cfm?URI=boe-4-6-924 doi: 10.1364/BOE.4.000924

CIE. (1932). *Commission Internationale de l'Éclairage Proceedings, 1931*. Cambridge: Cambridge University Press.

Clowes, M. B. (1962). A note on colour discrimination under conditions of retinal image constraint. *Optica Acta: International Journal of Optics*, *9*(1), 65–68.

Cole, B. L., Lian, K.-Y. & Lakkis, C. (2006). The new Richmond HRR pseudoisochromatic test for colour vision is better than the Ishihara test. *Clinical and Experimental Optometry*, *89*(2), 73–80. Retrieved from http://doi.wiley.com/10.1111/j.1444-0938.2006.00015.x doi: 10.1111/j.1444-0938.2006.00015.x

Cooper, R. F., Langlo, C. S., Dubra, A. & Carroll, J. (2013). Automatic detection of modal spacing (Yellott's ring) in adaptive optics scanning light ophthalmoscope images. *Ophthalmic and Physiological Optics*, *33*(4), 540–549. Retrieved from http://doi.wiley.com/10.1111/opo.12070 doi: 10.1111/opo.12070

Costela, F. M., McCamy, M. B., Macknik, S. L., Otero-Millan, J. & Martinez-Conde, S. (2013). Microsaccades restore the visibility of minute foveal targets. *PeerJ*, *1*, e119. Retrieved from https://peerj.com/articles/119 doi: 10.7717/peerj.119

Crane, H. D. & Steele, C. M. (1985). Generation-V dual-Purkinje-image eyetracker. *Applied Optics*, *24*(4), 527. Retrieved from https://www.osapublishing.org/abstract.cfm?URI=ao-24-4-527 doi: 10.1364/AO.24.000527

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. *The Journal of Neuroscience*, *29*(26),

8372–8387. Retrieved from http://www.jneurosci.org/content/29/26/8372http://www.jneurosci.org/content/29/26/8372.full.pdfhttp://www.jneurosci.org/content/29/26/8372.full.pdf+html?frame=headerhttp://www.ncbi.nlm.nih.gov/pubmed/19571128 doi: 10.1523/JNEUROSCI.1218-09.2009

Cunefare, D., Fang, L., Cooper, R. F., Dubra, A., Carroll, J. & Farsiu, S. (2017). Open source software for automatic detection of cone photoreceptors in adaptive optics ophthalmoscopy using convolutional neural networks. *Scientific Reports*, *7*(1), 6620. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28747737http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5529414 doi: 10.1038/s41598-017-07103-0

Curcio, C. A. & Allen, K. A. (1990). Topography of ganglion cells in human retina. *The Journal of Comparative Neurology*, *300*(1), 5–25. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/cne.903000103/abstracthttp://onlinelibrary.wiley.com/store/10.1002/cne.903000103/asset/903000103_ftp.pdf?v=1&t=i4m2jdk9&s=12a075ecc038495007f5c63015a16895e3ec2de0 doi: 10.1002/cne.903000103

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. *The Journal of Comparative Neurology*, *312*(4), 610–624. Retrieved from http://europepmc.org/abstract/MED/1722224http://onlinelibrary.wiley.com/doi/10.1002/cne.903120411/abstracthttp://onlinelibrary.wiley.com/store/10.1002/cne.903120411/asset/903120411_ftp.pdf?v=1&t=i4m2j7sq&s=6cc5aa8c6223600e8428a6ac24952af7475c10b1 doi: 10.1002/cne.903120411

Curcio, C. A., Sloan, K. R., Kalina, R. E. & Hendrickson, A. E. (1990). Human Photoreceptor Topography. *The Journal of Comparative Neurology*, *292*, 497–523. Retrieved from https://info.cs.uab.edu/curcio/PRtopo/Curcio_JCompNeurol1990_PRtopo_searchable.pdf

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), 731–5. Retrieved from http://dx.doi.org/10.1038/367731a0 doi: 10.1038/367731a0

Dacey, D. M., Peterson, B. B., Robinson, F. R. & Gamlin, P. D. (2003). Fireworks in the Primate Retina: In Vitro Photodynamics Reveals Diverse LGN-Projecting Ganglion Cell Types. *Neuron*, *37*(1), 15–27. Retrieved from http://www.sciencedirect.com/science/article/pii/S0896627302011431http://www.sciencedirect.com/science/article/pii/

S0896627302011431/pdfft?md5=b434be3b2441c94320868cb72ac69d2f&pid=1-s2.0-S0896627302011431-main.pdfzotero://attachment/111/ doi: 10.1016/S0896-6273(02)01143-1

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. doi: 10.1016/j.visres.2012.12.006

Danilova, M. V. & Mollon, J. D. (2016). Superior discrimination for hue than for saturation and an explanation in terms of correlated neural noise. *Proceedings of the Royal Society of London. Series B, 283*, 20160164. doi: 10.1098/rspb.2016.0164

Danilova, M. V. & Mollon, J. D. (2020). Discrimination of hue angle and discrimination of colorimetric purity assessed with a common metric. *Journal of the Optical Society of America A, 37*(4), A226–A236. Retrieved from https://ezproxy-prd.bodleian.ox.ac.uk:3716/DirectPDFAccess/D8C69F99-B0C7-0155-99FA44C3162382F6_ads382382.pdf?da=1&adsid=382382&journal=1&seq=0&mobile=no

Dartnall, H. J. A., Bowmaker, J. K. & Mollon, J. D. (1983). Human Visual Pigments: Microspectrophotometric Results from the Eyes of Seven Persons. *Proceedings of the Royal Society B: Biological Sciences, 220*(1218), 115–130. Retrieved from http://rspb.royalsocietypublishing.org/content/220/1218/115.abstract doi: 10.1098/rspb.1983.0091

Davis, L. E. & Davis, L. E. (1994). *A reference guide to the iraf/daophot package*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.584.2658

De Monasterio, F. M. & Gouras, P. (1975). Functional properties of ganglion cells of the rhesus monkey retina. *The Journal of Physiology, 251*(1), 167–195. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1348381/ doi: 10.1113/JPHYSIOL.1975.SP011086

De Valois, R. L. & De Valois, K. K. (1993). A multi-stage color model. *Vision Research, 33*(8), 1053–1065. Retrieved from http://www.sciencedirect.com/science/article/pii/004269899390240Whttp://www.sciencedirect.com/science/article/pii/004269899390240W/pdf?md5=2d76e65f168d63529d7ea0e5eb7f1766&pid=1-s2.0-004269899390240W-main.pdfhttp://www.ncbi.nlm.nih.gov/pubmed/8506645 doi: 10.1016/0042-6989(93)90240-W

Deeb, S. S., Dorschner, M., Shafer, A., Kutyavin, T. & Stamatoyannopolous, J. (2004). Novel Regulatory Regions of the Human L/M Photopigment Gene Locus. *ARVO Meeting Abstracts, 45*(5), 654. Retrieved from http://abstracts.iovs.org/cgi/content/abstract/45/5/654

Derrington, A. M., Krauskopf, J. & Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *The Journal of Physiology, 357*, 241–

65. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1193257&tool=pmcentrez&rendertype=abstract

Ditchburn, R. W. (1955). Eye-movements in relation to retinal action. *Optica Acta: International Journal of Optics*, *1*(4), 171–176. Retrieved from https://www.tandfonline.com/doi/full/10.1080/713818684 doi: 10.1080/713818684

Ditchburn, R. W. & Ginsborg, B. L. (1953). Involuntary eye movements during fixation. *The Journal of Physiology*, *119*(1), 1–17. Retrieved from http://doi.wiley.com/10.1113/jphysiol.1953.sp004824 doi: 10.1113/jphysiol.1953.sp004824

D'Zmura, M. (1996). Bergmann on visual resolution. *Perception*, *25*(10), 1223–1234. Retrieved from http://journals.sagepub.com/doi/10.1068/p251223 doi: 10.1068/p251223

Eizenman, M., Hallett, P. E. & Frecker, R. C. (1985). Power spectra for ocular drift and tremor. *Vision Research*, *25*(11), 1635–1640. Retrieved from https://www.sciencedirect.com/science/article/pii/0042698985901348?via%3Dihub doi: 10.1016/0042-6989(85)90134-8

Engbert, R., Mergenthaler, K., Sinn, P. & Pikovsky, A. (2011). An integrated model of fixational eye movements and microsaccades. *Proceedings of the National Academy of Sciences*, *108*(39), E765–E770. Retrieved from https://www.pnas.org/content/108/39/E765.short doi: 10.1073/PNAS.1102730108

Eskew, R. T. (2009). Higher order color mechanisms: a critical review. *Vision Research*, *49*(22), 2686–704. Retrieved from http://www.sciencedirect.com/science/article/pii/S0042698909003289 doi: 10.1016/j.visres.2009.07.005

Farneback, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis* (pp. 363–370). Halmstad, Sweden: SCIA.

Farup, I. (2014). Hyperbolic geometry for colour metrics. *Optics Express*, *22*(10), 12369–12378. Retrieved from http://www.opticsexpress.org/abstract.cfm?URI=oe-22-10-12369 doi: 10.1364/OE.22.012369

Fechner, G. T. (1966). *Elements of Psychophysics, vol. 1 [Original work published 1860]* (H. E. Adler, Trans.). New York, NY: Holt, Rinehart & Winston.

Field, G. D., Sher, A., Gauthier, J. L., Greschner, M., Shlens, J., Litke, A. M. & Chichilnisky, E. J. (2007). Spatial properties and functional organization of small bistratified ganglion cells in primate retina. *The Journal of Neuroscience*, *27*(48), 13261–13272. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1556594http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6575807 doi: 20026215

Garrigan, P., Ratliff, C. P., Klein, J. M., Sterling, P., Brainard, D. H. & Balasubramanian, V. (2010). Design of a trichromatic cone array. *PLoS Computational Biology*, *6*(2). doi: 10.1371/journal.pcbi.1000677

Garrioch, R., Langlo, C., Dubis, A. M., Cooper, R. F., Dubra, A. & Carroll, J. (2012). Repeatability of in vivo parafoveal cone density and spacing measurements. *Optometry and Vision Science*, *89*(5), 632–643. Retrieved from https://pubmed.ncbi.nlm.nih.gov/22504330/ doi: 10.1097/OPX.0b013e3182540562

Gautier, J., Bowers, N., Banks, M. S. & Roorda, A. (2020). A functional PRL in the foveola? In *AOSLO Psychophysics Workshop 2020*.

Green, D. G. (1968). The contrast sensitivity of the colour mechanisms of the human eye. *The Journal of Physiology*, *196*, 415–429. Retrieved from http://doi.wiley.com/10.1113/jphysiol.1968.sp008515 doi: 10.1113/jphysiol.1968.sp008515

Greschner, M., Shlens, J., Bakolitsa, C., Field, G. D., Gauthier, J. L., Jepson, L. H., ... Chichilnisky, E. J. (2011). Correlated firing among major ganglion cell types in primate retina. *The Journal of Physiology*, *589*(1), 75–86. Retrieved from http://www.jphysiol.org/cgi/doi/10.1113/jphysiol.2010.193888 doi: 10.1113/jphysiol.2010.193888

Gunther, K. L. & Dobkins, K. R. (2002). Individual differences in chromatic (red/green) contrast sensitivity are constrained by the relative number of L- versus M-cones in the eye. *Vision Research*, *42*(11), 1367–1378. Retrieved from https://www.sciencedirect.com/science/article/pii/S0042698902000433 doi: 10.1016/S0042-6989(02)00043-3

Gunther, K. L., Neitz, J. & Neitz, M. (2008). Nucleotide polymorphisms upstream of the X-chromosome opsin gene array tune L:M cone ratio. *Visual Neuroscience*, *25*(03), 265–271. Retrieved from http://journals.cambridge.org/action/displayFulltext?pageCode=100101&type=1&fid=1923868&jid=VNS&volumeId=25&issueId=03&aid=1923860&fromPage=onlinehttp://journals.cambridge.org/action/displayFulltext?type=6&fid=1923864&jid=VNS&volumeId=25&issueId=03&aid=1 doi: 10.1017/S0952523808080280

Hagstrom, S. A., Neitz, J. & Neitz, M. (1998). Variations in cone populations for red-green color vision examined by analysis of mRNA. *Neuroreport*, *9*(9), 1963–1967. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9674575

Hamilton, D. K. & Sheppard, C. J. R. (1984). Differential phase contrast in scanning optical microscopy. *Journal of Microscopy*, *133*(1), 27–39. Retrieved from http://doi.wiley.com/10.1111/j.1365-2818.1984.tb00460.x doi: 10.1111/j.1365-2818.1984.tb00460.x

Hamwood, J., Alonso-Caneiro, D., Sampson, D. M., Collins, M. J. & Chen, F. K. (2019). Automatic Detection of Cone Photoreceptors With Fully Convolutional Networks. *Translational Vision Science & Technology*, *8*(6), 10. Retrieved from https://tvst.arvojournals.org/article.aspx?articleid=2755333 doi: 10.1167/tvst.8.6.10

Hauperich, A.-K., Smithson, H. & Young, L. (2020). What makes a microsaccade? A review of 70 years research prompts a new detection method. *Journal of Eye Movement Research*, *12*(6). Retrieved from https://bop.unibe.ch/JEMR/article/view/JEMR.12.6.13 doi: 10.16910/jemr.12.6.13

Hennig, M. H. & Wörgötter, F. (2004). Eye micro-movements improve stimulus detection beyond the Nyquist limit in the peripheral retina. In *Advances in Neural Information Processing Systems* (pp. 1475– 1482). Denver, CO: MIT Press.

Hering, E. (1878). *Zur Lehre vom Lichtsinne*. Wien: Gerold.

Hofer, H., Carroll, J., Neitz, J., Neitz, M. & Williams, D. R. (2005). Organization of the Human Trichromatic Cone Mosaic. *The Journal of Neuroscience*, *25*(42), 9669–9679. Retrieved from http://www.jneurosci.org/content/25/42/9669http://www.jneurosci.org/content/25/42/9669.full.pdfhttp://www.jneurosci.org/content/25/42/9669.shorthttp://www.ncbi.nlm.nih.gov/pubmed/16237171 doi: 10.1523/JNEUROSCI.2414-05.2005

Hood, S. M., Mollon, J. D., Purves, L. & Jordan, G. (2006). Color discrimination in carriers of color deficiency. *Vision Research*, *46*(18), 2894–2900. Retrieved from www.elsevier.com/locate/visreshttp://www.sciencedirect.com/science/article/pii/S0042698906001398http://www.sciencedirect.com/science/article/pii/S0042698906001398/pdfft?md5=809b28f402d86fc86ccdb43ab6c71ed7&pid=1-s2.0-S0042698906001398-main.pdf doi: 10.1016/j.visres.2006.02.028

Horwitz, G. D. & Albright, T. D. (2003). Short-Latency Fixational Saccades Induced by Luminance Increments. *Journal of Neurophysiology*, *90*(2), 1333–1339. Retrieved from https://www.physiology.org/doi/10.1152/jn.00146.2003 doi: 10.1152/jn.00146.2003

Intoy, J. & Rucci, M. (2020). Finely tuned eye movements enhance visual acuity. *Nature Communications*, *11*, 795. Retrieved from http://www.nature.com/articles/s41467-020-14616-2 doi: 10.1038/s41467-020-14616-2

Jordan, G. & Mollon, J. D. (1997). Sons and mothers: classification of colour-deficient and heterozygous subjects by counterphase modulation photometry. In C. R. Cavonius (Ed.), *Colour Vision Deficiencies XIII* (pp. 385–392). Dordrecht: Kluwer Academic

Publishers. Retrieved from http://vision.psychol.cam.ac.uk/jdmollon/papers/JordanMollon1997CounterphaseModulationPhotometry.pdf

Judd, D. B. (1968). Ideal color space. II. The super-importance of hue differences and its bearing on the geometry of color space. *Palette*, *30*, 21–28.

Kesten, H. (1958). Accelerated Stochastic Approximation. *The Annals of Mathematical Statistics*, *29*(1), 41–59. Retrieved from http://www.jstor.org/stable/pdf/2237294.pdf?refreqid=excelsior%3A9601fd0f37300108fb667e1b92f506d6

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C. & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Research*, *34*(7), 885–912. Retrieved from http://www.sciencedirect.com/science/article/pii/0042698994900396%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/8160402 doi: 10.1016/0042-6989(94)90039-6

Ko, H.-K., Poletti, M. & Rucci, M. (2010). Microsaccades precisely relocate gaze in a high visual acuity task. *Nature Neuroscience*, *13*(12), 1549–1553. Retrieved from http://www.nature.com/articles/nn.2663 doi: 10.1038/nn.2663

Krauskopf, J., Williams, D. R. & Heeley, D. W. (1982). Cardinal directions of color space. *Vision Research*, *22*(9), 1123–1131. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7147723

Krauskopf, J., Williams, D. R., Mandler, M. B. & Brown, A. M. (1986). Higher order color mechanisms. *Vision Research*, *26*(1), 23–32. doi: 10.1016/0042-6989(86)90068-4

Kremers, J., Scholl, H. P. N., Knau, H., Berendschot, T. T. J. M., Usui, T. & Sharpe, L. T. (2000). L/M cone ratios in human trichromats assessed by psychophysics, electroretinography, and retinal densitometry. *Journal of the Optical Society of America A*, *17*(3), 517–526.

Kremers, J., Usui, T., Scholl, H. P. N. & Sharpe, L. T. (1999). Cone signal contributions to electroretinograms in dichromats and trichromats. *Investigative Ophthalmology & Visual Science*, *40*(5), 920–930. Retrieved from https://iovs.arvojournals.org/article.aspx?articleid=2162025

Kuchenbecker, J. A., Neitz, J. & Neitz, M. (2014). Ethnic variation in the ratio of long- to middle-wavelength sensitive cones. *Investigative Ophthalmology and Visual Science*, *55*(13), 4539. Retrieved from https://iovs.arvojournals.org/article.aspx?articleid=2270063

Kuehni, R. G. (2003). From Color-Matching Error to Large Color Differences. In *Color Space and Its Divisions: Color Order from Antiquity to the Present* (pp. 311–335). Hoboken, NJ, USA: John

Wiley & Sons, Inc.

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. *Journal of the Optical Society of America A*, *31*(4), A226–A231. Retrieved from http://josaa.osa.org/abstract.cfm?URI=josaa-31-4-A226http://www.opticsinfobase.org/DirectPDFAccess/7AF4285A-A074-3C8F-6F4447C17CB6D651_279358/josaa-31-4-A226.pdf?da=1&id=279358&seq=0&mobile=nohttp://www.opticsinfobase.org/josaa/abstract.cfm?uri=josaa-31 doi: 10.1364/JOSAA.31.00A226

Le Sueur, H., Mollon, J. D., Granzier, J. & Jordan, G. (2014). Counterphase modulation photometry: comparison of two instruments. *Journal of the Optical Society of America A*, *31*(4), A34–A37. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24695191 doi: 10.1364/JOSAA.31.000A34

Lee, B. B., Kremers, J. & Yeh, T. (1998). Receptive fields of primate retinal ganglion cells studied with a novel technique. *Visual Neuroscience*, *15*(1), 161–175. Retrieved from http://journals.cambridge.org/download.php?file=/VNS/VNS15_01/S095252389815112Xa.pdf&code=5d7a2a4fd76a66e0bb451e4f5abcf862

Lennie, P., Krauskopf, J. & Sclar, G. (1990). Chromatic mechanisms in striate cortex of macaque. *The Journal of Neuroscience*, *10*(2), 649–669.

Li, K. Y. & Roorda, A. (2007). Automated identification of cone photoreceptors in adaptive optics retinal images. *Journal of the Optical Society of America A*, *24*(5), 1358–1363.

Liang, J. & Williams, D. R. (1997). Aberrations and retinal image quality of the normal human eye. *Journal of the Optical Society of America A*, *14*(11), 2873. Retrieved from https://www.osapublishing.org/abstract.cfm?URI=josaa-14-11-2873 doi: 10.1364/JOSAA.14.002873

MacAdam, D. L. (1942). Visual Sensitivities to Color Differences in Daylight. *Journal of the Optical Society of America*, *32*(5), 247–274. Retrieved from http://www.osapublishing.org/viewmedia.cfm?uri=josa-32-5-247&seq=0&html=true doi: 10.1364/JOSA.32.000247

MacLeod, D. I. A. & Boynton, R. M. (1979). Chromaticity diagram showing cone excitation by stimuli of equal luminance. *Journal of the Optical Society of America*, *69*(8), 1183–1186. Retrieved from http://www.osapublishing.org/abstract.cfm?uri=josa-69-8-1183 doi: 10.1364/JOSA.69.001183

Marc, R. E. & Sperling, H. G. (1977). Chromatic organization of primate cones. *Science (New York, N.Y.)*, *196*(4288), 454–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/403607

Marcos, S., Werner, J. S., Burns, S. A., Merigan, W. H., Artal, P., Atchison, D. A., ... Sincich, L. C. (2017). Vision science and adaptive optics, the state of the field. *Vision Research*, *132*(February), 3–33. Retrieved from http://dx.doi.org/10.1016/j.visres.2017.01.006 doi: 10.1016/j.visres.2017.01.006

Mariotti, L. & Devaney, N. (2015). Performance analysis of cone detection algorithms. *Journal of the Optical Society of America A*, *32*(4), 497–506. Retrieved from https://www.osapublishing.org/DirectPDFAccess/11FD12EB-091F-9F89-62D09C0549BB625F_312595/josaa-32-4-497.pdf?da=1&id=312595&seq=0&mobile=no

Martinez-Conde, S., Macknik, S. L. & Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, *5*, 229–240. Retrieved from http://www.nature.com/nrn/journal/v5/n3/full/nrn1348.htmlhttp://www.nature.com/nrn/journal/v5/n3/pdf/nrn1348.pdf doi: 10.1038/nrn1348

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G. & Dyar, T. A. (2006). Microsaccades Counteract Visual Fading during Fixation. *Neuron*, *49*(6), 929. Retrieved from http://www.cell.com/article/S089662730600119X/abstracthttp://www.cell.com/article/S089662730600119X/pdfhttp://www.cell.com/neuron/abstract/S0896-6273(06)00119-Xhttps://www.cell.com/neuron/fulltext/S0896-6273(05)01056-1 doi: 10.1016/j.neuron.2006.02.007

McCamy, M. B., Macknik, S. L. & Martinez-Conde, S. (2014). Different fixational eye movements mediate the prevention and the reversal of visual fading. *The Journal of Physiology*, *592*(Pt 19), 4381. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4215783/ doi: 10.1113/JPHYSIOL.2014.279059

McCree, K. J. (1960). Colour confusion produced by voluntary fixation. *Optica Acta: International Journal of Optics*, *7*(3), 281–290. Retrieved from https://www.tandfonline.com/doi/full/10.1080/713826337 doi: 10.1080/713826337

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. Retrieved from http://jov.arvojournals.org/article.aspx?doi=10.1167/8.2.5 doi: 10.1167/8.2.5

Merbs, S. L. & Nathans, J. (1992). Absorption spectra of human cone pigments. *Nature*, *356*, 433–435. Retrieved from https://www.nature.com/articles/356433a0 doi: 10.1038/356433a0

Miyahara, E., Pokorny, J., Smith, V. C., Baron, R. & Baron, E. (1998). Color vision in two observers with highly biased LWS/MWS cone ratios. *Vision Research*, *38*(4), 601–612. Retrieved from

https://www.sciencedirect.com/science/article/pii/S0042698997883344 doi: 10.1016/S0042-6989(97)88334-4

Møller, F., Laursen, M., Tygesen, J. & Sjølie, A. (2002). Binocular quantification and characterization of microsaccades. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *240*(9), 765–770. Retrieved from http://link.springer.com/10.1007/s00417-002-0519-2 doi: 10.1007/s00417-002-0519-2

Mollon, J. D. & Bowmaker, J. K. (1992). The spatial arrangement of cones in the primate fovea. *Nature*, *360*(6405), 677–679. Retrieved from http://www.nature.com/articles/360677a0 doi: 10.1038/360677a0

Mollon, J. D. & Jordan, G. (1997). On the nature of unique hues. In *John Dalton's Colour Vision Legacy* (pp. 381–392). Retrieved from http://131.111.190.111/jdmollon/papers/MollonJordan1997UniqueHues.pdfhttp://vision.psychol.cam.ac.uk/jdmollon/papers/MollonJordan1997UniqueHues.pdf

Moreland, J. D., Robson, A. G. & Kulikowski, J. J. (2001). Macular pigment assessment using a colour monitor. *Color Research & Application*, *26*(S1), S261–S263. Retrieved from http://doi.wiley.com/10.1002/1520-6378%282001%2926%3A1%2B%3C%3A%3AAID-COL56%3E3.0.CO%3B2-6 doi: 10.1002/1520-6378(2001)26:1+<::AID-COL56>3.0.CO;2-6

Nathans, J., Davenport, C. M., Maumenee, I. H., Lewis, R. A., Hejtmancik, J. F., Litt, M., ... Fishman, G. (1989). Molecular genetics of human blue cone monochromacy. *Science*, *245*(4920), 831–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2788922

Nathans, J., Piantanida, T., Eddy, R., Shows, T. & Hogness, D. (1986). Molecular genetics of inherited variation in human color vision. *Science*, *232*(4747), 203–210. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3485310http://www.sciencemag.org/cgi/doi/10.1126/science.3485310 doi: 10.1126/science.3485310

Nathans, J., Thomas, D., Hogness, D. S., Shows, T. & Hogness, D. (1986). Molecular genetics of human color vision: the genes encoding blue, green, and red pigments. *Science (New York, N.Y.)*, *232*(4747), 193–202. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2937147 doi: 10.1126/science.2937147

Neitz, J., Carroll, J., Yamauchi, Y., Neitz, M. & Williams, D. R. (2002). Color Perception Is Mediated by a Plastic Neural Mechanism that Is Adjustable in Adults. *Neuron*, *35*(4), 783–792. Retrieved from http://www.sciencedirect.com/science/article/pii/S0896627302008188http://www.sciencedirect.com/science/article/pii/

S0896627302008188#http://www.sciencedirect.com/science/article/pii/S0896627302008188/pdfft?md5=a8e1c34e74c91ba36529179e1e1d22ca&pid=1-s2.0 doi: 10.1016/S0896-6273(02)00818-8

Nickerson, D. (1936). The specification of color tolerances. *Textile Research*, *6*(509), 505–514.

Noguchi, K., Latif, M., Thangavelu, K., Konietschke, F., Gel, Y. R. & Brunner, E. (2015). *nparLD: Nonparametric Analysis of Longitudinal Data in Factorial Experiments*. Retrieved from https://cran.r-project.org/web/packages/nparLD/

Office of National Statistics. (2016). *Ethnic group, national identity and religion*. Retrieved from https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion

Otake, S., Gowdy, P. D. & Cicerone, C. M. (2000). The spatial arrangement of L and M cones in the peripheral human retina. *Vision Research*, *40*(6), 677–693. Retrieved from https://www.sciencedirect.com/science/article/pii/S0042698999002023?via%3Dihub doi: 10.1016/S0042-6989(99)00202-3

Packer, O. S., Williams, D. F. & Bensinger, D. G. (1996). Photopigment Transmittance Imaging of the Primate Photoreceptor Mosaic. *The Journal of Neuroscience*, *76*(7), 2251–2260.

Pallikaris, A., Williams, D. R. & Hofer, H. (2003). The Reflectance of Single Cones in the Living Human Eye. *Investigative Opthalmology & Visual Science*, *44*(10), 4580–4592. Retrieved from http://iovs.arvojournals.org/article.aspx?doi=10.1167/iovs.03-0094 doi: 10.1167/iovs.03-0094

Park, S. C., Park, M. K. & Kang, M. G. (2003). Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, *20*(3), 21–36. Retrieved from http://ieeexplore.ieee.org/document/1203207/ doi: 10.1109/MSP.2003.1203207

Patrick, J. A., Roach, N. W. & McGraw, P. V. (2017). Motion-based super-resolution in the peripheral visual field. *Journal of Vision*, *17*(9), 15. Retrieved from http://jov.arvojournals.org/article.aspx?doi=10.1167/17.9.15 doi: 10.1167/17.9.15

Pokorny, J., Smith, V. C. & Lutze, M. (1989). Heterochromatic modulation photometry. *Journal of the Optical Society of America A*, *6*(10), 1618–1623. Retrieved from http://www.osapublishing.org/viewmedia.cfm?uri=josaa-6-10-1618&seq=0&html=true doi: 10.1364/JOSAA.6.001618

Prins, N. & Kingdom, F. A. A. (2009). *Palamedes: Matlab routines for analyzing psychophysical data* (No. accessed 07/01/13).

Pritchard, R. M. (1961). Stabilized images on the retina. *Scientific American*, *204*(6), 72–78. Re-

trieved from https://www.scientificamerican.com/article/stabilized-images-on-the-retina/ doi: 10.1038/scientificamerican0661-72

Purdy, D. M. (1931). Spectral Hue as a Function of Intensity. *The American Journal of Psychology*, *43*(4), 541. Retrieved from https://www.jstor.org/stable/1415157?origin=crossref doi: 10.2307/1415157

Putnam, N. M., Hofer, H. J., Doble, N., Chen, L., Carroll, J. & Williams, D. R. (2005). The locus of fixation and the foveal cone mosaic. *Journal of Vision*, *5*(7), 632–9. Retrieved from http://jov.arvojournals.org/Article.aspx?doi=10.1167/5.7.3 doi: 10.1167/5.7.3

Ratliff, F. & Riggs, L. A. (1950). Involuntary motions of the eye during monocular fixation. *Journal of Experimental Psychology*, *40*(6), 687–701. Retrieved from https://pubmed.ncbi.nlm.nih.gov/14803643/ doi: 10.1037/H0057754

Ratnam, K., Domdei, N., Harmening, W. M. & Roorda, A. (2017). Benefits of retinal image motion at the limits of spatial vision. *Journal of Vision*, *17*(1), 30. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28129414http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5283083 doi: 10.1167/17.1.30

Regan, S. E., Lee, R. J., MacLeod, D. I. A. & Smithson, H. E. (2018). Are hue and saturation carried in different neural channels? *Journal of the Optical Society of America A*.

Riggs, L. A., Armington, J. C. & Ratliff, F. (1954). Motions of the retinal image during fixation. *Journal of the Optical Society of America*, *44*(4), 315–321. Retrieved from https://www.osapublishing.org/abstract.cfm?URI=josa-44-4-315 doi: 10.1364/JOSA.44.000315

Roberts, M. R., Hendrickson, A., McGuire, C. R. & Reh, T. A. (2005). Retinoid X receptor (gamma) is necessary to establish the S-opsin gradient in cone photoreceptors of the developing mouse retina. *Investigative Ophthalmology & Visual Science*, *46*(8), 2897–904. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16043864 doi: 10.1167/iovs.05-0093

Rolfs, M. (2009). Microsaccades: Small steps on a long way. *Vision Research*, *49*(20), 2415–2441. Retrieved from http://dx.doi.org/10.1016/j.visres.2009.08.010 doi: 10.1016/j.visres.2009.08.010

Roorda, A., Metha, A. B., Lennie, P. & Williams, D. R. (2001). Packing arrangement of the three cone classes in primate retina. *Vision Research*, *41*(10-11), 1291–1306. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11322974http://www.sciencedirect.com/science/article/pii/S0042698901000438http://www.sciencedirect.com/science/article/pii/S0042698901000438/pdfft?md5=

7a0cb07130b4a851165d817ac86181b7&pid=1-s2.0-S0042698901000438-main.

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. Retrieved from https://www.osapublishing.org/abstract.cfm?URI=oe-10-9-405 doi: 10.1364/OE.10.000405

Roorda, A. & Williams, D. R. (1999). The arrangement of the three cone classes in the living human eye. *Nature*, *397*(6719), 520–522. Retrieved from http://www.nature.com/doifinder/10.1038/17383http://www.nature.com/nature/journal/v397/n6719/full/397520a0.htmlhttp://www.ncbi.nlm.nih.gov/pubmed/10028967 doi: 10.1038/17383

Rucci, M., Iovin, R., Poletti, M. & Santini, F. (2007). Miniature eye movements enhance fine spatial detail. *Nature*, *447*(7146), 851–855. Retrieved from http://www.nature.com/doifinder/10.1038/nature05866 doi: 10.1038/nature05866

Rucci, M. & Poletti, M. (2015). Control and Functions of Fixational Eye Movements. *Annual Review of Vision Science*, *1*(1), annurev–vision–082114–035742. Retrieved from http://www.annualreviews.org/doi/10.1146/annurev-vision-082114-035742 doi: 10.1146/annurev-vision-082114-035742

Sabesan, R., Hofer, H. & Roorda, A. (2015). Characterizing the Human Cone Photoreceptor Mosaic via Dynamic Photopigment Densitometry. *PLOS ONE*, *10*(12), e0144891. Retrieved from http://dx.plos.org/10.1371/journal.pone.0144891 doi: 10.1371/journal.pone.0144891

Sabesan, R., Schmidt, B. P., Tuten, W. S. & Roorda, A. (2016). The elementary representation of spatial and color vision in the human retina. *Science Advances*, *2*(9), e1600797. doi: 10.1126/sciadv.1600797

Santini, F., Redner, G., Iovin, R. & Rucci, M. (2007). EyeRIS: A general-purpose system for eye-movement-contingent display control. *Behavior Research Methods*, *39*(3), 350–364. Retrieved from https://link.springer.com/article/10.3758/BF03193003 doi: 10.3758/BF03193003

Schmidt, B. P., Boehm, A. E., Foote, K. G. & Roorda, A. (2018). The spectral identity of foveal cones is preserved in hue perception. *Journal of Vision*, *18*(11), 1–18. Retrieved from file:///C:/Users/Sarah/Downloads/i1534-7362-18-11-19(1).pdf

Schmidt, B. P., Sabesan, R., Tuten, W. S., Neitz, J. & Roorda, A. (2018). Sensations from a single M-cone depend on the activity of surrounding S-cones. *Scientific Reports*, *8*(1), 8561. Retrieved

from http://www.nature.com/articles/s41598-018-26754-1 doi: 10.1038/s41598-018-26754-1

Schmidt, B. P., Touch, P., Neitz, M. & Neitz, J. (2015). The relative number of L and M cones shapes color experience. *Journal of Vision*, *16*, 2015. doi: 10.1167/14.15.29

Schulz, E. (1984). Binocular micromovements in normal persons. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *222*(2), 95–100. Retrieved from http://link.springer.com/10.1007/BF02150640 doi: 10.1007/BF02150640

Shaaban, S. A. & Deeb, S. S. (1998). Functional analysis of the promoters of the human red and green visual pigment genes. *Investigative Ophthalmology & Visual Science*, *39*(6), 885–96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9579468

Sharpe, L. T., Stockman, A., Jagla, W. & Jägle, H. (2005). A luminous efficiency function, V*(λ), for daylight adaptation. *Journal of Vision*, *5*(11), 948–968. Retrieved from http://jov.arvojournals.org/article.aspx?doi=10.1167/5.11.3 doi: 10.1167/5.11.3

Sharpe, L. T., Stockman, A., Jagla, W. & Jägle, H. (2011). A luminous efficiency function, VD65* (λ), for daylight adaptation: A correction. *Color Research & Application*, *36*(1), 42–46. Retrieved from https://doi.org/10.1002/col.20602 doi: 10.1002/col.20602

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. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(2), 1008–11. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11773636http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC117421 doi: 10.1073/pnas.022629799

Smith, T. & Guild, J. (1931). The C.I.E. colorimetric standards and their use. *Transactions of the Optical Society*, *33*(3), 73–134. Retrieved from http://stacks.iop.org/1475-4878/33/i=3/a=301?key=crossref.3dec82dbb24569e539ed1a7e2fdbee98 doi: 10.1088/1475-4878/33/3/301

Solomon, S. G. & Lennie, P. (2007). The machinery of colour vision. *Nature Reviews Neuroscience*, *8*(4), 276–286. Retrieved from http://www.nature.com/doifinder/10.1038/nrn2094 doi: 10.1038/nrn2094

Steinman, R. M. (1965). Effect of Target Size, Luminance, and Color on Monocular Fixation. *Journal of the Optical Society of America*, *55*(9), 1158–1165. Retrieved from https://www.osapublishing.org/viewmedia.cfm?uri=josa-55-9-1158&seq=0&html=truehttps://www.osapublishing.org/abstract.cfm?uri=josa-55-9

-1158https://www.osapublishing.org/josa/abstract.cfm?uri=josa-55-9-1158 doi: 10.1364/josa.55.001158

Stiles, W. & Burch, J. (1959). N.P.L. Colour-matching Investigation: Final Report (1958). *Optica Acta: International Journal of Optics*, *6*(1), 1–26. Retrieved from https://www.tandfonline.com/doi/full/10.1080/713826267 doi: 10.1080/713826267

Stockman, A. & Brainard, D. H. (2010). Color vision mechanisms. In M. Bass et al. (Eds.), *The Optical Society of America Handbook of Optics* (3rd ed., pp. 11.1–11.104). New York: McGraw Hill. Retrieved from http://www.researchgate.net/publication/228358887_Color_vision_mechanisms

Stockman, A., Jägle, H., Pirzer, M. & Sharpe, L. T. (2008). The dependence of luminous efficiency on chromatic adaptation. *Journal of Vision*, *8*(16), 1.1–26. Retrieved from http://jov.arvojournals.org/article.aspx?articleid=2121947 doi: 10.1167/8.16.1

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. doi: 10.1016/S0042-6989(00)00021-3

Stockman, A., Sharpe, L. T. & Fach, C. (1999). The spectral sensitivity of the human short-wavelength sensitive cones derived from thresholds and color matches. *Vision Research*, *39*(17), 2901–2927. Retrieved from http://www.sciencedirect.com/science/article/pii/S0042698998002259 doi: 10.1016/S0042-6989(98)00225-9

Sun, H., Smithson, H. E., Zaidi, Q. & Lee, B. B. (2006). Specificity of Cone Inputs to Macaque Retinal Ganglion Cells. *Journal of Neurophysiology*, *95*(2), 837–849. Retrieved from http://www.physiology.org/doi/10.1152/jn.00714.2005 doi: 10.1152/jn.00714.2005

The GIMP Team. (2018). *GNU Image Manipulation Program 2.10*. Retrieved from gimp.org

Thibos, L. N. (1998). Acuity Perimetry and the Sampling Theory of Visual Resolution. *Optometry and Vision Science*, *75*(6), 339–406.

Thibos, L. N., Still, D. L. & Bradley, A. (1996). Characterization of spatial aliasing and contrast sensitivity in peripheral vision. *Vision Research*, *36*(2), 249–258. doi: 10.1016/0042-6989(95)00109-D

Timberlake, G. T., Sharma, M. K., Grose, S. A., Gobert, D. V., Gauch, J. M. & Maino, J. H. (2005). Retinal location of the preferred retinal locus relative to the fovea in scanning laser ophthalmoscope images. *Optometry and Vision Science*, *82*(3), 177–185. Retrieved from https://journals.lww.com/optvissci/Fulltext/2005/03000/Retinal_Location_of_the_Preferred_Retinal_Locus.9.aspx doi: 10.1097/01.opx.0000156311

.49058.c8

Tuten, W. S., Harmening, W. M., Sabesan, R., Roorda, A. & Sincich, L. C. (2017). Spatiochromatic Interactions between Individual Cone Photoreceptors in the Human Retina. *Journal of Neuroscience*, *37*(39), 9498–9509. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/28871030http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5618266 doi: 10.1523/JNEUROSCI.0529-17.2017

Usui, T., Kremers, J., Sharpe, L. T. & Zrenner, E. (1998). Flicker cone electroretinogram in dichromats and trichromats. *Vision Research*, *38*(21), 3391–3396. Retrieved from https://www.sciencedirect.com/science/article/pii/S0042698997004665?via%3Dihub doi: 10.1016/S0042-6989(97)00466-5

Vogel, C. R., Arathorn, D. W., Roorda, A. & Parker, A. (2006). Retinal motion estimation in adaptive optics scanning laser ophthalmoscopy. *Optics Express*, *14*(2), 487. Retrieved from http://www.math.montana.edu/vogel/http://vision.berkeley.edu/roordalab/http://www.math.montana.edu/parker/ doi: 10.1364/opex.14.000487

Wang, Y., Macke, J. P., Merbs, S. L., Zack, D. J., Klaunberg, B., Bennett, J., ... Nathans, J. (1992). A locus control region adjacent to the human red and green visual pigment genes. *Neuron*, *9*(3), 429–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1524826 doi: 10.1016/0896-6273(92)90181-C

Webb, R. H. & Hughes, G. W. (1981). Scanning Laser Ophthalmoscope. *IEEE Transactions on Biomedical Engineering*, *BME-28*(7), 488–492. Retrieved from http://ieeexplore.ieee.org/document/4121261/ doi: 10.1109/TBME.1981.324734

Webb, R. H., Hughes, G. W. & Pomerantzeff, O. (1980). Flying spot TV ophthalmoscope. *Applied Optics*, *19*(17), 2991. Retrieved from https://www.osapublishing.org/abstract.cfm?URI=ao-19-17-2991 doi: 10.1364/AO.19.002991

Wells-Gray, E. M., Choi, S. S., Bries, A. & Doble, N. (2016). Variation in rod and cone density from the fovea to the mid-periphery in healthy human retinas using adaptive optics scanning laser ophthalmoscopy. *Eye*, *30*(8), 1135–1143. Retrieved from http://www.nature.com/articles/eye2016107 doi: 10.1038/eye.2016.107

West, D. C. & Boyce, P. R. (1968). The effect of flicker on eye movement. *Vision Research*, *8*(2), 171–192. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5729328 doi: 10.1016/0042-6989(68)90005-9

Whittaker, S. G., Budd, J. & Cummings, R. W. (1988). Eccentric fixation with macular scotoma. *Investigative Ophthalmology & Visual Science*, *29*(9), 268–278. Retrieved from https://

iovs.arvojournals.org/article.aspx?articleid=2178036

Wilk, M. A., Dubis, A. M., Cooper, R. F., Summerfelt, P., Dubra, A. & Carroll, J. (2017). Assessing the spatial relationship between fixation and foveal specializations. *Vision Research, 132*, 53–61. doi: 10.1016/j.visres.2016.05.001

Winderickx, J., Battisti, L., Motulsky, A. G. & Deeb, S. S. (1992). Selective expression of human X chromosome-linked green opsin genes. *Proceedings of the National Academy of Sciences of the United States of America, 89*(20), 9710–4. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1409688http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC50202 doi: 10.1073/pnas.89.20.9710

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. Retrieved from https://www.nature.com/articles/356431a0 doi: 10.1038/356431a0

Winterson, B. J. & Collewijn, H. (1976). Microsaccades during finely guided visuomotor tasks. *Vision Research, 16*(12), 1387–1390. Retrieved from https://www.sciencedirect.com/science/article/pii/0042698976901565 doi: 10.1016/0042-6989(76)90156-5

Wyszecki, G. & Stiles, W. S. (1982). The Eye. In G. Wyszecki & W. S. Stiles (Eds.), *Color Science: Concepts and Methods, Quantitative Data and Formulas* (2nd ed., pp. 83–116). John Wiley & Sons, Inc.

Xue, B., Choi, S. S., Doble, N. & Werner, J. S. (2007). Photoreceptor counting and montaging of en-face retinal images from an adaptive optics fundus camera. *Journal of the Optical Society of America. A, Optics, image science, and vision, 24*(5), 1364–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17429482

Yamaguchi, T., Motulsky, A. G. & Deeb, S. S. (1997). Visual pigment gene structure and expression in human retinae. *Human Molecular Genetics, 6*(7), 981–990. Retrieved from https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/6.7.981 doi: 10.1093/hmg/6.7.981

Yamauchi, Y., Yatsu, K., Kuchenbecker, J., Neitz, M. & Neitz, J. (2013). L:M Cone Ratio of Japanese Derived with ERG Flicker Photometry Method. In *Investigative Ophthalmology & Visual Science* (Vol. 54, p. 6119). Retrieved from https://iovs.arvojournals.org/article.aspx?articleid=2151136

Yang, Q., Arathorn, D. W., Tiruveedhula, P., Vogel, C. R. & Roorda, A. (2010). Design of an integrated hardware interface for AOSLO image capture and cone-targeted stimulus delivery. *Optics Express, 18*(17), 17841. Retrieved from https://pubmed.ncbi.nlm.nih.gov/20721171/

doi: 10.1364/oe.18.017841

Young, L. K., Hauperich, A. K., Morris, T. J., Saunter, C. D. & Smithson, H. E. (2017). Recording fixational eye movements with a new AOSLO: simulation, measurement and evaluation. *Journal of Vision*, *17*(7), 34. Retrieved from http://jov.arvojournals.org/article.aspx?doi=10.1167/17.7.34 doi: 10.1167/17.7.34

Young, L. K., Morris, T. J., Saunter, C. D. & Smithson, H. E. (2018). Compact, modular and in-plane AOSLO for high-resolution retinal imaging. *Biomedical Optics Express*, *9*(9), 4275–4293. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/30615719http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6157778 doi: 10.1364/BOE.9.004275

Young, L. K. & Smithson, H. E. (2018). A self-organising photoreceptor mosaic simulation for studying spatial sampling in the retina. In *AVA Christmas Meeting, Queen Mary University London, December 18 2017*. Perception.

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. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(16), 7951–7956. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/30944223http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6475411 doi: 10.1073/pnas.1816360116

Zhang, T., Godara, P., Blanco, E. R., Griffin, R. L., Wang, X., Curcio, C. A. & Zhang, Y. (2015). Variability in Human Cone Topography Assessed by Adaptive Optics Scanning Laser Ophthalmoscopy. *American Journal of Ophthalmology*, *160*(2), 290–300.e1. Retrieved from https://www.sciencedirect.com/science/article/pii/S0002939415002378#bib36 doi: 10.1016/J.AJO.2015.04.034

Zuber, B. L., Stark, L. & Cook, G. (1965). Microsaccades and the velocity-amplitude relationship for saccadic eye movements. *Science*, *150*(3702), 1459–1460. Retrieved from https://science.sciencemag.org/content/150/3702/1459.long doi: 10.1126/SCIENCE.150.3702.1459