



City Research Online

City St George's, University of London

Citation: Evans, B. (2023). Researches on colour vision assessment including the development and validation of a new colour vision screener. (Unpublished Doctoral thesis, City, University of London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://city-ref-test.eprints-hosting.org/id/eprint/30468/>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

Researches on colour vision assessment including the development and validation of a new colour vision screener

Benjamin Edward William Evans

Doctor of Philosophy

City, University of London

Centre for Applied Vision Research

Department of Optometry and Visual Science

April 2023

Declaration

I, Benjamin Evans, confirm that the work presented in this thesis is my own. All sources used, and information derived, is fully acknowledged and all quotations are properly identified.

As specified in the City, University of London Senate Regulation 25 (20220713), I also agree that Library Services or any third party with whom CRO has an agreement to do so may, without changing content, transfer the thesis to any medium or format for the purpose of future preservation and accessibility.

Signed:

Date:

Sunday 16th April 2023

Abstract

Functional colour vision is the result of a symphony of parallel biological processes that enable humans with normal trichromatic colour vision to make effective use of colour encoded information. The use of colour coding in transport and other visually demanding working environments has increased significantly in recent years and it has become more important to be able to screen efficiently for colour vision deficiency and to quantify accurately the severity of colour vision loss.

The outcome of conventional methods of colour vision assessment in current single- and multi-test protocols were investigated to determine the current state of colour vision assessment. An exploratory study evaluated the outcome of a test commonly used in these protocols; the Farnsworth D-15. The new ‘Colour Vision Screener’ (CVS) test was developed to address many of the issues present in conventional methods of colour vision assessment and a preliminary study was carried out to determine the optimal test parameters. Upon determination of the test parameters the CVS was evaluated and validated across three separate experiments to: establish the performance of the CVS in least affected individuals with congenital colour vision deficiency, determine the reliability of the CVS, and examine the validity of the CVS through an international multi-centre study.

Current single- and multi-test colour assessment protocols fail to achieve their aims and are unable to achieve high sensitivity, or reliably quantify the severity of red/green or yellow/blue loss. The CVS was found to have very high test-retest reliability, and high sensitivity and specificity. The test does not suffer from several limitations present in conventional methods of colour vision assessment and allows for the swift, accurate, and reliable detection of congenital and acquired colour vision deficiencies.

Table of contents

List of figures	vii
List of tables	ix
1 Colour and human colour vision	1
1.1 Introduction	2
1.2 Human colour vision	3
1.2.1 Trichromacy	6
1.2.2 Opponency	7
1.2.3 Neural colour (mid- and high-level processing)	10
1.2.4 Aging	11
1.3 Colour vision deficiency (CVD)	12
1.3.1 Congenital CVD	13
1.3.2 Acquired CVD	16
1.4 The value of colour	17
1.5 Outlook	18
2 Colour vision assessment	19
2.1 Introduction	20
2.2 A technology driven history of colour vision assessment	22
2.3 Protocols for comprehensive colour vision assessments	30
2.3.1 Pseudoisochromatic plate tests	31
2.3.1.1 The Ishihara plates	32
2.3.1.2 The AO-HRR plates	35
2.3.2 Lantern tests	36
2.3.2.1 The Holmes-Wright lanterns	36
2.3.3 Hue discrimination tests	38
2.3.3.1 Farnsworth-Munsell 100 hue	38
2.3.3.2 Farnsworth D-15	40
2.3.3.3 City University test	41
2.3.4 Spectral anomaloscopes	43

2.3.4.1	The Nagel Anomaloscope	43
2.3.5	Display-based tests	46
2.3.5.1	Cambridge Colour Test (CCT)	47
2.3.5.2	Colour Assessment and Diagnosis (CAD) test	48
2.4	‘Full’ colour vision assessment and the ‘Gold standard’	53
2.4.1	The challenges of ‘full’ colour assessment	53
2.4.2	The ‘Gold standard’ of colour vision assessment	54
2.5	Summary	56
3	The current state of conventional colour vision assessment	57
3.1	Introduction	58
3.2	Single- and multi-test protocols	59
3.2.1	Introduction	59
3.2.2	Methods	63
3.2.3	Results	66
3.2.4	Discussion	75
3.2.4.1	Single-test protocols	77
3.2.4.2	Multi-test protocols	81
3.2.5	Conclusions	84
3.3	Exploratory Farnsworth D-15 study	85
3.3.1	Introduction	85
3.3.2	Methods	86
3.3.3	Results	87
3.3.4	Discussion	93
3.3.5	Conclusions	95
3.4	Alternatives to single- and multi-test protocols	95
3.5	Implications	98
4	The development and validation of the Colour Vision Screener	100
4.1	Introduction	101
4.2	The Colour Vision Screener (CVS)	103
4.2.1	CVS implementation	104
4.2.2	CVS design and theory	104
4.3	Preliminary study	108
4.3.1	Methods	109
4.3.2	Results	110
4.3.3	Discussion	112
4.4	Modifications made to the CVS following the preliminary study	113
4.5	Main study	115

4.5.1	Experiment 1 - the performance of the CVS for normal trichromats, least-affected deuterans, and least-affected protans	116
4.5.1.1	Methods	118
4.5.1.2	Results	119
4.5.1.3	Discussion	121
4.5.2	Experiment 2 - the reliability of the CVS	126
4.5.2.1	Methods	127
4.5.2.2	Results	127
4.5.2.3	Discussion	130
4.5.3	Experiment 3 - the validity of the CVS	133
4.5.3.1	Methods	133
4.5.3.2	Results	134
4.5.3.3	Discussion	135
4.6	Summary	141
5	The Colour Vision Screener: A multi-centre study	143
5.1	Introduction	144
5.2	Methods	146
5.3	Results	148
5.4	Discussion	156
6	Conclusion and outlook	163
6.1	Summary of findings	163
6.2	Reflection	164
6.3	Future directions	166
6.3.1	Continuing validation of the CVS at external centres	166
6.3.2	Investigation into performance of the CVS on uncalibrated visual displays	167
6.3.3	Usability and validity of the CVS in children	168
6.4	Conclusion	169
Appendix A	The statistically equivalent CAD threshold	170
Appendix B	Mathematical description of statistical terms	172
Appendix C	Supplementary figures	176
References		179

List of figures

1.1	Basic optics of the human eye	4
1.2	Simplified retinal layer schematic	5
1.3	Retinal ganglion cell centre-surround receptive field	8
1.4	Cone pathways	10
1.5	Neural visual pathways	11
2.1	Graphical representation of silent substitution	25
2.2	Human colour vision assessment development timeline	29
2.3	Ishihara pseudoisochromatic test plates	34
2.4	Circular Farnsworth D-15 results diagram	40
2.5	City University test plates (2nd Edition)	41
2.6	Schematic diagram of the Nagel Type I anomaloscope	44
2.7	CCT stimulus	47
2.8	Dynamic LC noise levels	50
2.9	CAD test stimulus	51
2.10	CAD thresholds as a function of age	52
3.1	A theoretical multi-test protocol, ‘Protocol C’	60
3.2	Farnsworth D-15 ‘visual inspection’ classification grid	65
3.3	Comparison between box plots and violin plots	66
3.4	Single- and multi-test protocol study CAD thresholds	67
3.5	CAD thresholds in those who pass and fail the City University test (2nd Edition)	69
3.6	CAD thresholds in those who pass and fail the Farnsworth D-15	73
3.7	CAD thresholds in those who pass and fail the police multi-test protocol	73
3.8	Ranked CAD thresholds in those who pass and fail the police multi-test protocol	74
3.9	CAD thresholds in those who pass and fail the fire service multi- test protocol	74
3.10	Ranked CAD thresholds in those who pass and fail the fire service multi-test protocol	75

3.11	Farnsworth D-15 data and theoretical photoreceptor excitations for a normal trichromat	88
3.12	Farnsworth D-15 data and theoretical photoreceptor excitations for a deuteranope	89
3.13	Farnsworth D-15 data and theoretical photoreceptor excitations for a protanope	90
3.14	Farnsworth D-15 data and theoretical photoreceptor excitations for a tritanope	91
3.15	Farnsworth D-15 data and theoretical photoreceptor excitations for a rod monochromat	92
4.1	Task carried out and colours used in the CVS	106
4.2	Preliminary study CAD thresholds and age distribution	110
4.3	Preliminary study CVS results	111
4.4	Changes in RG CAD thresholds as a function of age	117
4.5	Fitted Weibull functions for all study participants	121
4.6	Fitted Weibull functions for a normal, deutan, and protan . .	122
4.7	Reliability study CAD thresholds	128
4.8	Reliability study CVS data in normal trichromats	129
4.9	Reliability study CVS data in deutans and an individual with acquired loss	130
4.10	Validity study CAD thresholds	136
4.11	Validity study CVS results	137
4.12	Validity study CVS results as a function of CAD threshold . .	138
5.1	Multi-centre study age distribution	149
5.2	Multi-centre study CAD thresholds	153
5.3	Multi-centre study CVS results	154
5.4	Multi-centre study CVS results as a function of CAD threshold	155
A.1	Extract from Chpt3_RankPlotFigures.m.	171
C.1	CAD thresholds obtained at each external CAD centre	177
C.2	CVS outcome distribution at each external CAD centre	178

List of tables

1.1	Prevalence of congenital colour vision deficiency	15
2.1	Colour vision test battery used at City, University of London .	31
2.2	Ishihara pseudoisochromatic plate design and function	33
3.1	Outcome of the Ishihara pseudoisochromatic plate test	68
3.2	Classification made by the Ishihara pseudoisochromatic plates	69
3.3	Outcome of the City University test (2nd Edition)	70
3.4	Classification made by the City University test (2nd Edition) .	71
3.5	Outcome of the Farnsworth D-15 test	71
3.6	Classification made by the Farnsworth D-15 test	72
3.7	Outcome of UK police and fire service multi-test colour vision protocols	72
4.1	Preliminary CVS classification parameters	107
4.2	Final CVS classification and confidence rating parameters . . .	115
4.3	CVS outcome predicted by fitted Weibull functions	120
4.4	Goodness of fit statistics for fitted Weibull functions	120
4.5	CVS reliability data collected at City, University of London . .	128
4.6	CVS validity data collected at City, University of London . . .	135
5.1	Multi-centre CVS study contributing CAD centres	148
5.2	Combined CVS multi-centre validity data	151
5.3	Individual centre CVS multi-centre validity data	151
5.4	Multi-centre RG CVS statistical outcome	152
5.5	Multi-centre YB CVS statistical outcome	152
5.6	Ishihara and multi-centre RG CVS statistical outcome	152
B.1	General representation of a diagnostic colour vision test	173
B.2	Descriptors for a general representation of a diagnostic colour vision test	173

Acknowledgements

I would like to begin by expressing my profound gratitude to my supervisors Professor John Barbur and Dr Marisa Rodriguez-Carmona. Both have too many laudable qualities to list, but I am particularly thankful to John for his wit, awe-inspiring breadth of knowledge, and untamed enthusiasm and Marisa for her empathy, gordian wisdom, and reposeful candour. Whilst the debt accrued for the opportunities they have provided is one I fear I may never be able to repay I hope to one day pass it forward.

I wish to thank Dr James Sadler for the many discussions, pragmatic advice, and great company, Umar Abd-Alghani and Li-Ying Ng for assisting with the data collection for the preliminary and reliability studies, and all research participants for giving up their time to participate.

I thank and acknowledge the COLT Foundation, UK Civil Aviation Authority, Rank Prize, Worshipful Company of Tin Plate Workers *alias* Wire Workers, and Colour Group (GB) for their financial support.

My thanks also extend to all the external Colour Assessment and Diagnosis centres who took part in the multi-centre study. There are many people working at these centres to acknowledge and a special mention goes to Dr Vilhelm Koefoed. I also wish to thank Joshibi University, Tokyo, particularly Miss Mari Terasaki and Miss Miharu Shinohara for producing the graphical illustrations used in the Colour Vision Screener.

My personal thanks to AT, CF, H/IM, JP, FM, and MB for their support and encouragement over the course of this project.

Finally, I would like to thank my family for their support particularly A&B, C&S, D&G, LE, J&F, A&P, and my parents for their love, encouragement, and sage advice. Upon explicit threat of physical and emotional injury, I would like also to hereby formally affirm my thanks to Josephine Evans. Over the course of my time at City she has remained hopelessly cheerful, tenaciously optometric, and ‘the best’ sister.

To my parents. . .

∞

“You live and learn.

At any rate, you live”

Douglas Adams - Mostly Harmless

Publications

The following research papers associated with the work described in this thesis have been published or are in preparation:

Publications

- **Evans, B. E. W.**, Rodriguez-Carmona, M. & Barbur, J. L. (2020), ‘Color vision assessment-1: Visual signals that affect the results of the Farnsworth D-15 test’, *Color Research & Application* **46**(1), 7–20.
- Rodriguez-Carmona, M., **Evans, B. E. W.** & Barbur, J. L. (2020), ‘Color vision assessment-2: Color assessment outcomes using single and multi-test protocols’, *Color Research & Application* **46**(1), 21–32.
- Barbur, J. L., Rodriguez-Carmona, M., & **Evans, B. E. W.** (2020), ‘Color vision assessment-3. An efficient, two-step, color assessment protocol’, *Color Research & Application* **46**(1), 33–45.

In preparation

- Evans, B. E. W. *et al.* (*in preparation*), ‘The Colour Vision Screener: A multi-centre study’.
- Evans, B. E. W., Rodriguez-Carmona, M. & Barbur, J. L. (*in preparation*), ‘A technologically driven history of colour vision assessment’.

Disclosure of interests

The colour vision screener described in this thesis is produced by City Occupational Limited, a spinout company set up by City, University of London to develop and market the advanced vision and optometric tests. I have no relationship with City Occupational Limited. Professor John Barbur, one of my supervisors, is a company director for City Occupational Limited.

Glossary

2AFC	Two alternative forced choice
4AFC	Four alternative forced choice
AFO	Authorised firearms officer
AVOT	Advanced Vision and Optometric Tests
CAA	Civil Aviation Authority
CAD	Colour Assessment and Diagnosis
CCD	Congenital colour deficiency
CCT	Cambridge Colour Test
CVD	Colour vision deficiency
CIE	Comission Internationale de l'Eclairage
<i>CNGB3</i>	Cyclic Nucleotide Gated Channel Subunit Beta 3 - a protein coding gene associated with achromatopsia
CVS	Colour Vision Screener
D	Deutanopia
DA	Deuteranomaly
EM	Electromagnetic
ERG	Electroretinogram
ipRGC	Intrinsically photosensitive retinal ganglion cells
IQ	Intelligence quotient
IQR	Interquartile range
LC	Luminance contrast
MCA	Maritime Coastguard Agency
NPV	Negative predictive value
P	Protanopia
PA	Protanomaly
PPV	Positive predictive value
RG	Red/Green
RLM	Random luminance modulation
STU	Specially trained units
TfL	Transport for London
VEP	Visually evoked potential
YB	Yellow/Blue

Chapter 1

Colour and human colour vision

This chapter begins this thesis with an outline of colour and the underlying physiology that enables nearly all humans to perceive colour. It moves on to describe cases, both acquired and congenital in nature, where human colour vision may be disrupted.

Chapter 2 provides an introduction to the various methods designed and used to assess human colour vision.

Chapter 3 describes experimental work carried out to establish the current state of conventional colour vision assessment.

Chapters 4 & 5 describe experimental work carried out to develop and validate a new colour vision test, the Colour Vision Screener (CVS), including the design and implementation of an international multi-centre study.

Chapter 6 contains a summary of the work contained within this thesis along with a description of several potential avenues for future study.

1.1 Introduction

The ‘visual spectrum’ is limited to a narrow range of electromagnetic (EM) radiation which extends from approximately 380nm to 780nm. The preferential absorption of light as a function of wavelength in cone photoreceptor pigments triggers a cascading symphony that starts with the generation of electrical signals that carry information about the spatial distribution of both the amount and spectral composition of light in the retinal image. This information is then encoded in the discrete firing rate of ganglion cells for efficient transmission to the cortex. Any interference with the normal functioning of the retina through the introduction of electrical stimulation at any point along this pathway can result in the perception of visual phosphenes and the perception of coloured flashes (Brindley & Lewin 1968, Brindley 1970, Murphey et al. 2008, Lewis & Rosenfeld 2016, Yue et al. 2021).

Such ‘indirect’ colour can be found in some cases of synaesthesia where non-chromatic visual or auditory stimuli are associated with the perception of colour (Bargary & Mitchell 2008, Brang & Ramachandran 2011, Smees et al. 2019) (and with optical ‘illusions’ such as Benham’s top (von Campenhausen & Schramme 1995, Kaiser & Boynton 1996)). Although some suggest that synaesthesia has a neural basis (Rouw et al. 2011, Rothen et al. 2012), a review by Hupé & Dojat (2015) “*did not find any structural or functional anomaly in the brain of synesthetes that could explain synaesthesia*” (Hupé & Dojat 2015, page 13), indicating that if any neural differences are present, they vary widely across individuals and/or current measurement techniques are insufficient to detect such differences. Non-biological mechanisms by which different wavelengths of EM radiation can be converted and conveyed using other stimuli (such as auditory stimuli) have also been explored (Alfaro et al. 2015, Bonnardel 2021).

In this thesis colour only generated as a direct consequence of the chromatic properties of visual stimuli is considered, investigated, and described. The relationship between the physical properties of a stimulus and the resulting sensory response is studied through the discipline of *psychophysics*.

1.2 Human colour vision

Human vision, of which colour vision is an important aspect, begins when light (EM radiation with wavelengths $\sim 380\text{-}780\text{nm}$ (Sliney 2016)) enters the eye and is focussed, filtered and projected on to the retina – a thin ($\sim 0.5\text{mm}$ thick (Kaiser & Boynton 1996, Kolb 2003)) cellular mesh lining the back of the eye. The process of colour perception begins with opponent comparisons of cone photoreceptor excitations generated by phototransduction – a process that converts light into a change in the electrical potential across a cell membrane (Stockman et al. 1993, Kolb 2003).

Focussing in the eye involves refraction at the cornea and crystalline lens (as shown in Figure 1.1A) and optical filtering occurs as light passes through the cornea, crystalline lens, and the macular pigment. The macular pigment, present in the inner retinal layers of the central retina (Trieschmann et al. 2008), only acts as a filter for photoreceptors within the macula (central ~ 18 degrees of the retina (Masri et al. 2021)). The filtering of light by both the crystalline lens and the macular pigment change as a function of age (Birch 2001, Artigas et al. 2012, Vemala et al. 2017), but both act generally to filter wavelengths $< 380\text{nm}$ (Gegenfurtner et al. 2001, Sliney 2002, 2016, Barbur et al. 2010, Stockman 2019).

The retina is a multi-layered cellular network responsible for the first stages in visual processing (Figure 1.2). The layout and distribution of the components that make up the retina vary as a function of foveal eccentricity.

At the fovea, one finds the maximum density of photoreceptors, bipolar, and ganglion cells. The bipolar cells receive inputs from a small number of photoreceptors, and consequently the receptive field of the retinal midlayer ganglion cells is small. In the central region of the fovea, the foveola, all other retinal layers are laterally displaced, and light is incident directly on to the photoreceptors (as shown in Figure 1.1A). Cones in the fovea are also smaller than their peripheral counterparts (Curcio et al. 1990). The foveola subtends ~ 1 degree, the fovea subtends ~ 5.5 degrees, the parafovea surrounds the fovea

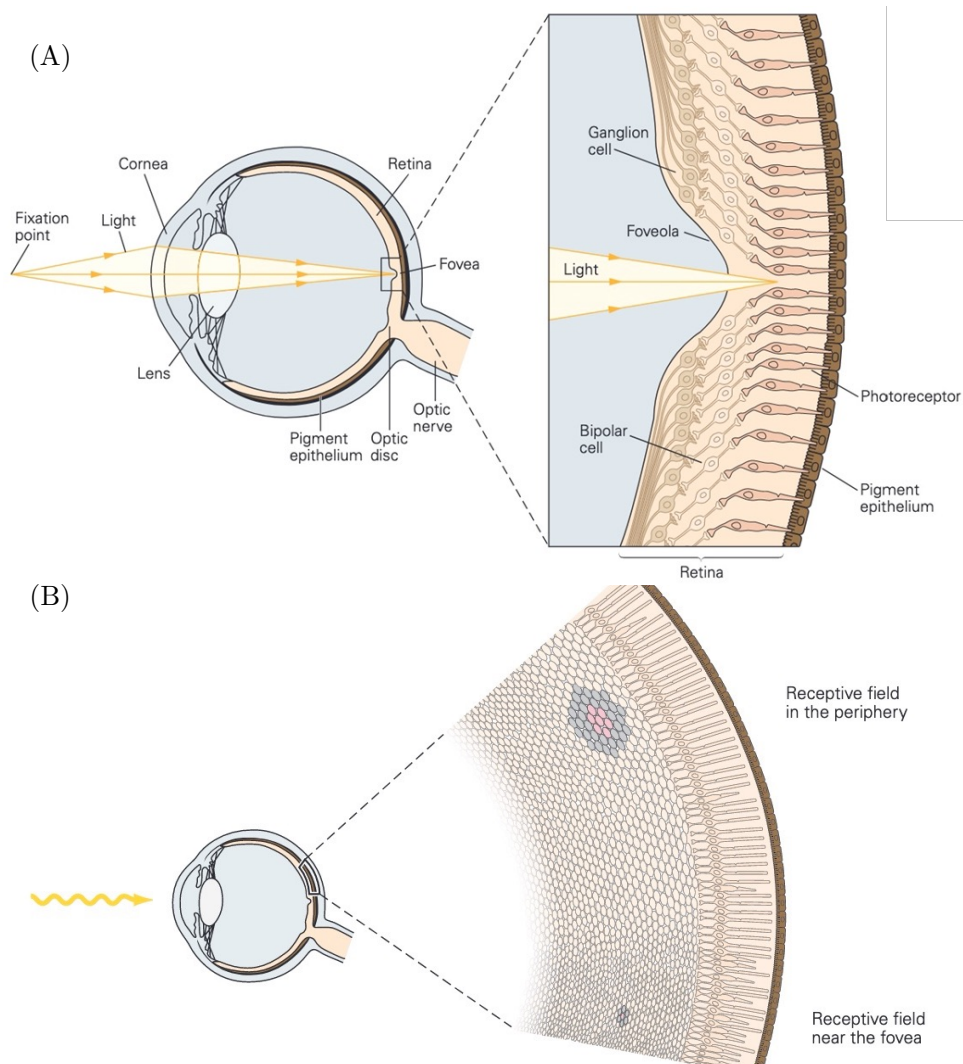


Figure 1.1: A. Light from an object is focussed by the optics of the eye onto the retina. B. The receptive field of retinal ganglion cells varies with foveal eccentricity. The receptive field is smallest at the fovea and increases in the periphery (adapted from Kandel et al. (2021)).

and subtends ~ 8 degrees, and the remainder of the macula is the perifovea (Kolb 2011, Masri et al. 2021).

In the perifovea and more peripheral retina, light is incident upon cone and rod photoreceptors after passing through almost all other retinal layers (Figure 1.2). The cone photoreceptor density decreases and the retinal ganglion cell receptive field increases as one moves away from the fovea (Kandel et al. 2021).

There are five kinds of light sensitive cell found in the retina: three classes of cone photoreceptors, rod photoreceptors, and intrinsically photosensitive retinal ganglion cells (ipRGCs) (Gegenfurtner et al. 2001, Allen et al. 2019).

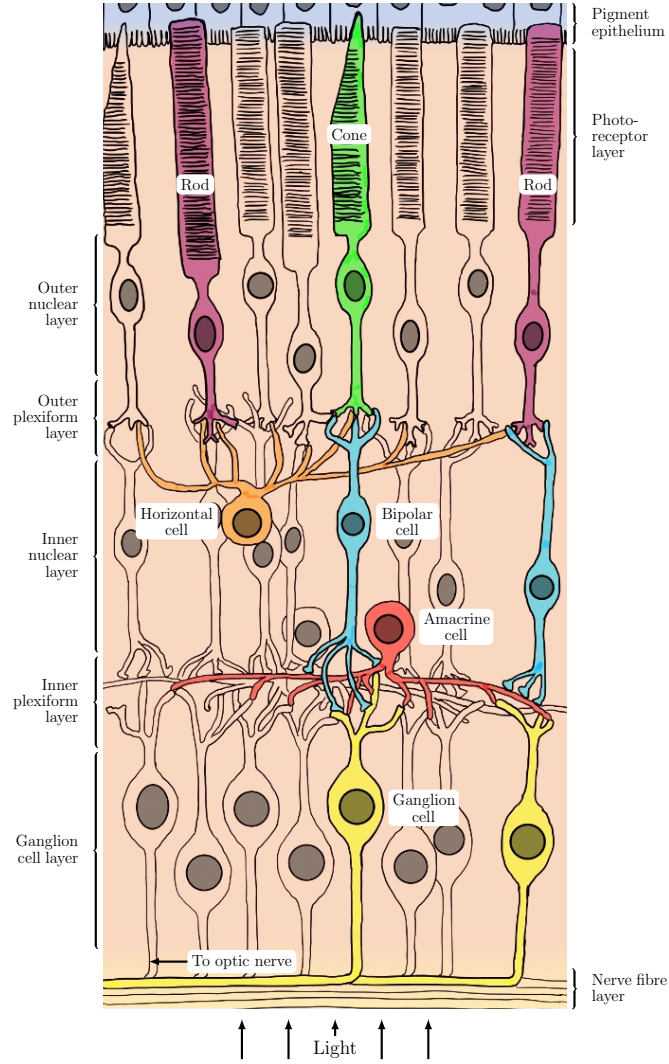


Figure 1.2: A simplified schematic showing the layers present within the retina. Light enters from the bottom of this diagram passing through the nerve fibre layer, ganglion cells, inner plexiform layer, amacrine, bipolar, horizontal cells, and outer plexiform layer before reaching the outer segments of the rod and cone photoreceptors (adapted from Purves (2004)). Note: this simplified view does not show Müller (supporting glial) cells which extend from the outer nuclear layer to form the internal limiting membrane beyond the nerve fibre layer, astrocytes present in the nerve fibre layer, or microglia present in the outer and inner plexiform layers, the ganglion cell layer and the inner nuclear layer (Peters 2002, Guo et al. 2022).

The three classes of cone photoreceptors are classified by their peak sensitivity and are split into long-, middle-, and short-wavelength sensitive (L, M, and S) categories with peak spectral responsivities (λ_{max}) of $\sim 419\text{nm}$ (S), $\sim 531\text{nm}$ (M), $\sim 559\text{nm}$ (L) (Kolb 2013). The difference in spectral responsivity is a result of specific sequences in the genetic code for the L, M, and S photopigments (Neitz & Neitz 2011, Wandell et al. 2022). Cones are the

photoreceptors associated with, and responsible for, human colour vision and are used in photopic and mesopic conditions (Gegenfurtner et al. 2001).

Rod photoreceptors extend visual sensitivity in mesopic and scotopic conditions and are mostly absent from the fovea (Kaiser & Boynton 1996, Birch 2001). Rod photoreceptor density peaks at approximately 5 degrees from the fovea (Kaufman et al. 2003). Although prevalent throughout the retina and invaluable in scotopic conditions, rod photoreceptors do not mediate colour vision (Gegenfurtner & Rieger 2000).

ipRGCs are a subclass of retinal ganglion cells that contain melanopsin. They are absent from the fovea and have the highest density in the perifovea (Do & Yau 2010). Although ipRGCs are an active area of research (Schmidt et al. 2011, Lucas et al. 2014, Allen et al. 2019) historically their input has not been considered when describing human colour vision due to their low temporal resolution and retinal position. Recent findings, however, have suggested that melanopsin may have an impact, specifically for short wavelength stimuli in extrafoveal conditions (Barrionuevo et al. 2022).

1.2.1 Trichromacy

At the photoreceptor level, human colour vision is trichromatic and directly proportional to the relative input from the three classes of cone photoreceptors. Each individual class of cone photoreceptors has no sensitivity to colour as any absorbed photon triggers the hyperpolarisation of the cell. What changes across the three classes of photoreceptors is the probability that a photon will be absorbed. As such, any information regarding the intensity and wavelength of the photon is lost upon its absorption (Stockman et al. 1993, Gegenfurtner et al. 2001, Stockman 2019). This principle of univariance, exploited in several methods designed to assess human colour vision, is discussed in more detail in section 2.2 (Rushton 1972).

The absorption of a photon by a cone’s opsin results in the isomerisation (a change in the arrangement of the atoms in the molecule) of the 11-*cis*-retinal molecule within the cell (Gegenfurtner et al. 2001). The isomerisation con-

verts to the 11-*cis*-retinal to all-*trans*-retinal triggering a series of biological processes that result in a reduction in the influx of positively charged ions into the cell and, consequently, the hyperpolarisation of the cell (Lamb & Pugh 2006, Jiang & Mahroo 2022, Lamb 2022).

The three classes of cones, L-, M-, and S-cones, are distributed across the retinal mosaic and the relative ratio of cone photoreceptors present in the retinal mosaic varies amongst individuals (Gegenfurtner et al. 2001, Kremers et al. 2000, Hofer et al. 2005, Sabesan et al. 2015). The total proportion of S-cones has been reported to vary between approximately 4-10%, with S-cones being absent from the central 0.3 degrees of the fovea (Curcio et al. 1991, Williams 2011, Sabesan et al. 2015). The L:M cone ratio has been found to vary significantly with observed ratios ranging from 0.37:1 to 16.5:1 in individuals with normal colour vision (Carroll et al. 2002, Hofer et al. 2005). Extreme variations in L:M cone ratio can be caused by carriers of congenital colour vision deficiency (CVD) (Gegenfurtner et al. 2001, Hofer et al. 2005).

Cone photoreceptors synapses link with horizontal and bipolar cells. The horizontal cells, as part of the lateral pathway, are mostly inhibitory and, amongst other functions, enable the human visual system to respond over a wide range of illumination levels (Werblin 1973, Lee et al. 1999, Smith et al. 2001, Grünert & Martin 2020). The bipolar and horizontal cells are the start of a biological system which transforms the photoreceptor signals generated into opponent visual information channels.

1.2.2 Opponency

Colour opponency occurs in the neural layers of the retina. The trichromatic signals generated from the three classes of cone photoreceptor are processed and transformed through a three-neuron chain (cones $\rightarrow$ bipolar cells $\rightarrow$ ganglion cells) into three separate pathways: an L-M-opponent pathway, S-opponent pathway, and a luminance pathway (Kaufman et al. 2003, Solomon 2021). The axons of the retinal ganglion cells form the optic nerve and the

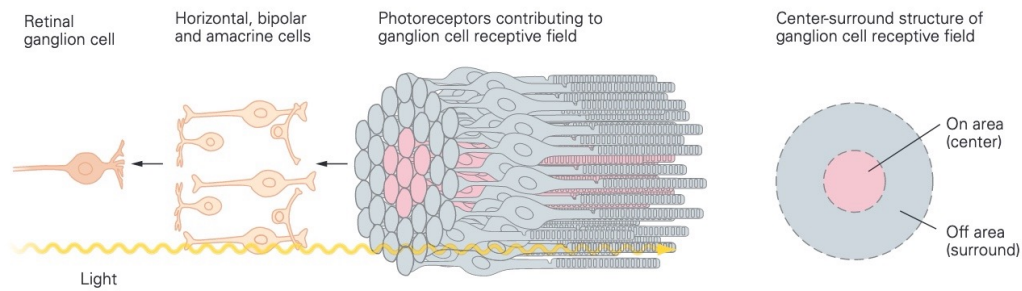


Figure 1.3: The receptive field of some retinal ganglion cells. The inputs from several photoreceptors are transmitted to retinal ganglion cells by horizontal, bipolar, and amacrine cells. As shown in Figure 1.1B, the receptive field shown varies as a function of foveal eccentricity (adapted from Kandel et al. (2021)).

neural pathway that projects through the lateral geniculate nucleus (LGN) of the thalamus (Stockman & Brainard 2015, Neitz & Neitz 2016).

A large number of amacrine, bipolar, and ganglion cells have been catalogued (at least 12 types of bipolar cell, 17 types of ganglion cell, and 40 types of amacrine cells in primate retina) (Kolb 2003, Grünert & Martin 2020, Masri et al. 2021). For brevity, a simplified view of the retinal circuits responsible for human colour vision is considered here.

There are two primary classifications of bipolar cells, ON-centre and OFF-centre, each classified by the response induced by cone photoreceptors (Kaufman et al. 2003, Stockman & Brainard 2015). OFF-centre bipolar cells hyperpolarise in light, conserving the sign generated by the hyperpolarisation of the cone photoreceptor. ON-centre bipolar cells depolarise in light, inverting the sign generated by the initial hyperpolarisation of the cone photoreceptor (Kolb & Marshak 2003). These bipolar cells are further subdivided into midget bipolar cells which connect to one photoreceptor and diffuse bipolar cells which connect to multiple photoreceptors (Kandel et al. 2021, Solomon 2021). The opposing side of these bipolar synapses interface with amacrine and retinal ganglion cells.

Three kinds of retinal ganglion cell, parasol, midget, and bistratified ganglion cells, each project to a different layer of the lateral geniculate nucleus (LGN) (Kandel et al. 2021). The retinal ganglion cells and subsequent pathways are categorised by the location to which they project in the LGN (magnocellular,

parvocellular, and koniocellular [etymology: magno – large, parvo – small, konio – dust]).

In the fovea each M- and L-cone contacts a midsize ON- and OFF-bipolar cell, compared to ~ 2 -3 cones per bipolar cell in the periphery. Each midsize ON- and OFF-bipolar cell in turn contacts a single midsize ganglion cell (Gegenfurtner et al. 2001).

Information transmitted via L- and M-cones connected to midsize bipolar and midsize retinal ganglion cells forms an L-M-opponent pathway that projects to the LGN via the parvocellular pathway (Derrington et al. 1984, Stockman & Brainard 2015).

S-cones have been considered to input only to an S-ON-centre bipolar cells which synapse with S-cone bistratified retinal ganglion cells (Dacey & Lee 1994, Calkins et al. 1998, Kaufman et al. 2003, Stockman & Brainard 2015). This input is then combined in S-cone bistratified retinal ganglion cells that also receive inputs from L- and M-cones via diffuse bipolar cells, resulting in the formation of an S-opponent pathway (Kaufman et al. 2003, Szmajda et al. 2006). This S-opponent information projects to the LGN via the koniocellular pathway (Szmajda et al. 2008). However, more recent evidence suggests that S-cones may transmit OFF-centre information via midsize ganglion cells (Wool et al. 2019). Analysis of macaque retina has also provided evidence for transmissions between S-cones and OFF bipolar cells in primate retina (Klug et al. 2003, Field et al. 2010).

An additional pathway is formed from the combined inputs of L- and M-cones (under photopic conditions) via ON- and OFF-diffuse bipolar cells to ON- and OFF-parasol retinal ganglion cells (Stockman & Brainard 2010). These signals, from two classes of photoreceptors with very similar spectral sensitivities, form an achromatic photopic luminance pathway that projects to LGN via the magnocellular pathway (De Monasterio & Gouras 1975, Kandel et al. 2021). S-cones are not considered to contribute significantly to this luminance pathway (Sun et al. 2006, Ripamonti et al. 2009).

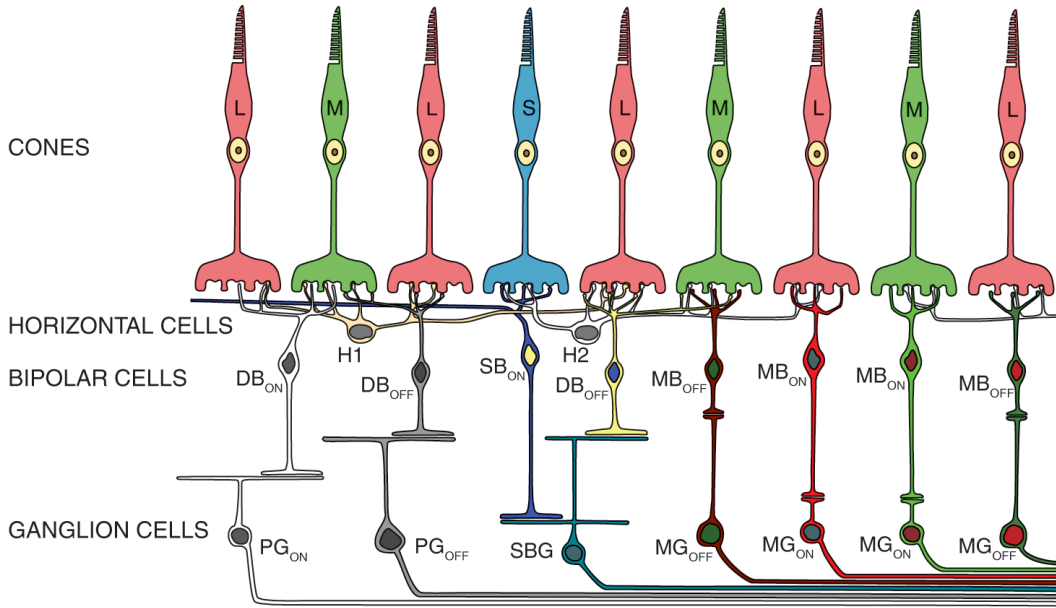


Figure 1.4: Human cone pathways responsible for the transformation of trichromatic cone signals to two opponent chromatic channels and one achromatic channel. The R/G opponent channel originates from L- and M-cones interacting with midget bipolar cells (DB) which synapse midget ganglion cells of the same polarity (ON $\rightarrow$ ON & OFF $\rightarrow$ OFF). Inputs from S-cones via S-cone bipolar cells (SB) are combined with diffuse bipolar cells (connected to L- and M-cones) in S-cone bistratified ganglion cells (SBG) to form a YB opponent channel. An achromatic luminance channels is formed from the combined inputs from L- and M-cones via diffuse bipolar cells to parasol ganglion cells (PG) of the same polarity. Lateral connections made by amacrine cells are not shown (adapted from (Stockman & Brainard 2015)).

The cone-based retinal circuits that result in these pathways are shown in Figure 1.4. The L-M-opponent pathway projects red/green (RG) opponent information and the S-opponent pathway projects yellow/blue (YB) opponent information to the LGN (Kandel et al. 2021).

1.2.3 Neural colour (mid- and high-level processing)

After passing through the optic nerve the ganglion cell axons from each eye meet at the optic chiasm. Here the ganglion cells representing the temporal visual field of either eye cross the chiasm and separate the left and right visual hemifields to right and left hemispheres respectively. En route from the optic chiasm the axons extend through the optic tract to the LGN - a region in the thalamus. The LGN is made up of six primary ovoid layers, two magnocellular

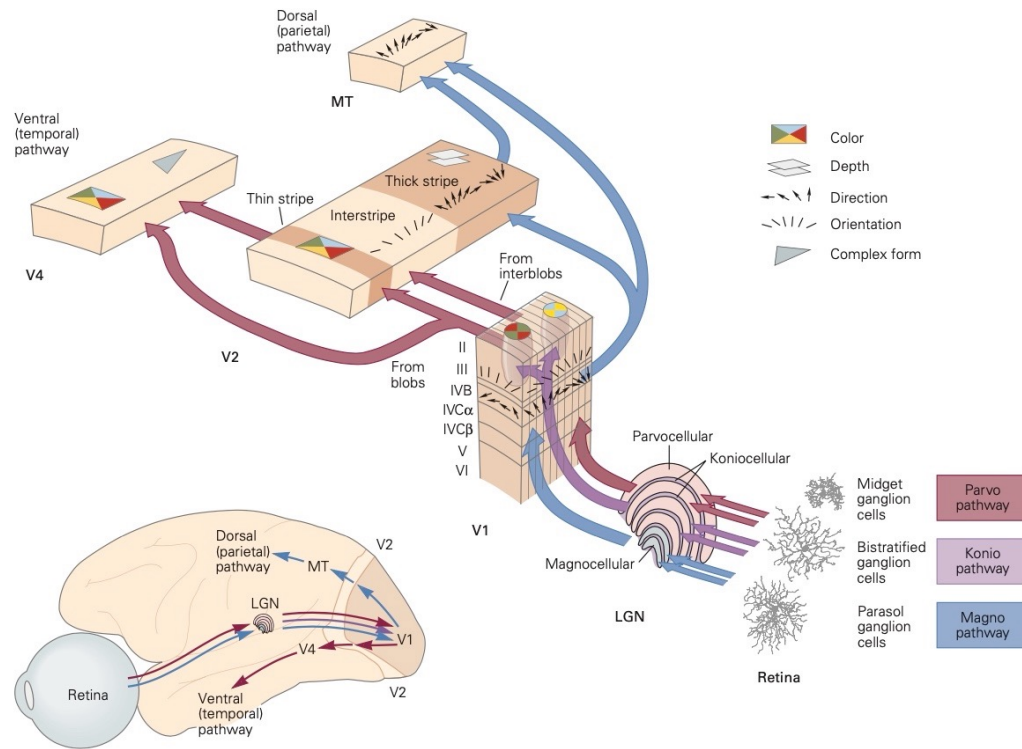


Figure 1.5: R/G and Y/B channels are processed via the ventral stream (V1→V4). It should be noted that although pathways are shown as distinct there is considerable interconnection between them (adapted from (Kandel et al. 2021)).

and four parvocellular. All six of these layers are interlaced with koniocellular layers (Figure 1.5) (Kaufman et al. 2003). RG and YB opponency is preserved by the neurons in the LGN (Dacey 2000) and with further processing in the visual cortex (V1 to V4 and higher association areas) colour perception begins to emerge from RG and YB opponency (Li et al. 2022).

Cortical chromatic mechanisms are complex and the high-order processing that occurs in and beyond the visual cortex is responsible for a number of perceived chromatic phenomena, including colour constancy (Gegenfurtner 2003, Smithson 2005, Barbur & Spang 2008, Foster 2011).

1.2.4 Aging

Changes in colour vision as a result of ‘normal healthy aging’ have been investigated in a number of studies using several methods of colour vision assessment (described in more detail in chapter 2) (Verriest et al. 1982, Kinnear &

Sahraie 2002, Paramei & Oakley 2014, Barbur & Rodriguez-Carmona 2015). Psychophysical testing has revealed two phases of variation of chromatic sensitivity over the human life span:

- I. from birth to ~ 20 years old
- II. from ~ 20 years old to death

From birth there is an increase in chromatic sensitivity until the minimum threshold is reached at an age of approximately 20 years. Part of this may be due to an increase in performance at colour discrimination and detection tasks with age and some methods of colour vision assessment have been shown to be positively correlated with individuals nonverbal intelligence quotient (IQ) (Cranwell et al. 2015). After approximately 20 years chromatic sensitivity slowly decreases as a function of age (Knoblauch et al. 2001, Barbur & Rodriguez-Carmona 2015).

Several factors contribute to the degradation of chromatic sensitivity as a function of time, including the increase in density and yellowing of crystalline lens (Pokorny et al. 1987), the reduced optical density of photopigments (Werner & Steele 1988, Shinomori et al. 2001, Wuerger et al. 2010), and other generalised aging changes in the visual pathway (Peters 2002).

1.3 Colour vision deficiency (CVD)

A subsection of the human population possesses a reduced ability to detect, distinguish, identify, and consequently make use of colour. The extent to which an individual's sensitivity to colour (or an individual's 'chromatic sensitivity') is reduced can vary from just outside the 'normal range' to a complete absence of colour vision (Knoblauch et al. 2001, Barbur & Rodriguez-Carmona 2015). Individuals who have reduced sensitivity to colour are said to have a colour vision deficiency (CVD), or 'Daltonism' (Birch 2001, Parry 2015).

CVDs occur in two flavours, congenital and acquired. The terms congenital and acquired refer to the nature of the deficiency and the two are not mutually

exclusive - it is possible to have no CVD (normal trichromatic colour vision), congenital CVD, acquired CVD, or both acquired and congenital CVD.

1.3.1 Congenital CVD

Congenital CVDs are present from birth and arise from the encoding, or lack thereof, of different opsins and the subsequent production and expression of opsins within cone photoreceptors (Neitz & Neitz 2000, 2011). Due to congenital CVD being a direct consequence of the photoreceptors produced in the retina, congenital CVD affects both eyes equally and is generally stable over time (Birch 2001).

Interestingly, although there are inherited conditions that can indirectly affect colour vision, the blanket term ‘congenital colour vision deficiency’ is used exclusively to refer to hereditary cases where only colour vision is affected. Changes in colour vision as a result of progressive congenital conditions, including inherited retinal diseases such as Leber congenital amaurosis and Stargardt macular dystrophy, although technically congenital in nature, are categorised as acquired CVDs.

Congenital CVDs are categorised via a parallel classification system based upon the number and type of photoreceptors produced in the retina. The first layer of this parallel classification system is determined by the classes of cone photoreceptors present. Three classes of cones are expressed and produced in forms of trichromacy (normal trichromacy and anomalous trichromacy), two classes of cones are expressed and produced in forms of dichromacy, and a single class of cones or only rod photoreceptors are found in cases of cone and rod monochromacy respectively. The second parallel layer is used to classify the class of cone photoreceptors that are absent in cases of anomalous trichromacy and dichromacy (Birch 2001).

For cases of anomalous trichromacy and dichromacy the prefixes *protan-*, *deutan-*, and *tritan-* refer to the class of cone photoreceptor that is not expressed (L-, M-, and S-cones respectively). The suffixes *-omaly* and *-opia* are used to identify cases of anomalous trichromacy and dichromacy respectively

(i.e., *deuteranomaly* is a case of anomalous trichromacy where the M-cone is not produced, instead being ‘replaced’ by a variant L-cone (L’) (Barbur & Rodriguez-Carmona 2017)). Protan and deutan (*deutan-*) deficiencies are forms of RG CVD and tritan deficiencies are YB CVDs (Parry 2015).

Anomalous trichromats express and produce three classes of cone photoreceptors across their retinal mosaic, however one of the three cones is abnormal with a shifted λ_{max} (Gegenfurtner et al. 2001). For protanomaly and deuteranomaly (deficiencies in which the L- and M- cone are replaced with M’- and L’-cones respectively) the deficiency can be quantified by the degree to which the cone is shifted compared to the λ_{max} absent photoreceptor ($\delta\lambda_{max}$) (Barbur et al. 2008, Barbur & Rodriguez-Carmona 2017). In a normal trichromat $\delta\lambda_{max}$ for the L- and M-cone photoreceptors is 28nm (Birch 2001). In individuals with protanomaly the $\delta\lambda_{max}$ between M- and M’-cones ranges from ~ 1 to 7 nm and in individuals with deuteranomaly the $\delta\lambda_{max}$ between L- and L’-cones ranges from ~ 1 to 12 nm (Nathans et al. 1986, Neitz et al. 1996, Barbur & Rodriguez-Carmona 2017). The ‘least affected’ protanomalous and deuteranomalous trichromats have the largest $\delta\lambda_{max}$ (~ 7 and ~ 12 nm respectively) (Neitz et al. 1996, Rodriguez-Carmona & Barbur 2017).

Dichromats express and produce only two classes of cone photoreceptors. Protanopes express M- and S-cones, deuteranopes express L- and S-cones, and tritanopes express L- and M-cones. As such a dichromat’s colour vision is limited at the first stage of retinal processing and their post-receptoral mechanisms must reassemble and interpret two chromatic dimensions. This results in a larger range of potential colour confusions compared to normal and anomalous trichromats - a property that is exploited in spectral anomaloscopes and other methods of colour vision assessment described in chapter 2 (Birch 2001).

Individuals with cone monochromacy and rod monochromacy (sometimes referred to as achromatopsia¹) only have access to one dimension of colour at the photoreceptor level. Rod monochromats do not produce functioning cone

¹As noted by Birch (2001) the term achromatopsia means absence of colour vision and, in addition to rod monochromacy, is also used to refer to complete acquired cortical CVD

Type of colour deficiency	Prevalence (%)	
	Male	Female
Protanomalous trichromacy	1.08	0.03
Deuteranomalous trichromacy	4.63	0.36
Protanopia	1.01	0.02
Deuteranopia	1.27	0.01
Tritanopia	$\sim 0.01 - 0.20^*$	
Rod monochromacy	$\sim 0.002 - 0.003^\dagger$	

*Incidence between 1 in 500 and 1 in 10,000

† Incidence between 1 in 30,000 and 1 in 50,000

Table 1.1: Prevalence of congenital colour vision deficiencies (adapted from data compiled by Sharpe & Nordby (1990), Gegenfurtner et al. (2001), Birch (2001, 2012)). The prevalence of CVD in different populations is discussed in more detail in section 5.1.

photoreceptors and consequently have poor visual acuity ($\sim 20/200$), in addition to reduced chromatic sensitivity (Sharpe & Nordby 1990, Birch 2001). Rarer cases of incomplete, or atypical, achromatopsia have also been reported in which residual cones provide some chromatic discrimination (Pokorny et al. 1982, Nathans et al. 1989, Schneck et al. 1990, Sharpe & Nordby 1990, Birch 2001, Aboshiha et al. 2016). Mutations in the *CNGB3* gene account for approximately 50% of rod monochromacy (Nathans et al. 1989, Kohl et al. 2000, 2005).

The inheritance mechanism, and consequently the prevalence, varies across different types of congenital CVD. Blue cone monochromacy is inherited as an X-linked recessive trait (Nathans et al. 1989). Rod monochromacy and incomplete achromatopsia are autosomal recessive conditions (Pokorny et al. 1982, Aboshiha et al. 2016). The photopigment present in S-cones is coded on chromosome 7 and the photopigments present in L- and M- cones are coded in an array on the X-chromosome (Nathans et al. 1986, 1992). Congenital RG CVDs follow an X-linked recessive inheritance, and are subsequently more common in individuals with one X-chromosome, and congenital YB CVD are inherited via an autosomal dominant trait. The prevalence of different types of congenital CVD are shown in Table 1.1.

1.3.2 Acquired CVD

Acquired CVDs, unlike congenital RG CVDs, affect men and women equally, can change rapidly over time, and affect each eye separately depending upon the underlying cause (Birch 2001). Whereas the loss of chromatic sensitivity in congenital CVD (excluding cases of monochromacy) is limited to either RG or YB loss, acquired CVDs can result in a reduction in both RG and YB chromatic sensitivity. Acquired loss can occur from injury, damage, or disruption to any part of the visual pathway and as a result can impact other visual functions, such as acuity or visual fields (Verriest 1963, Pacheco-Cutillas et al. 1999, O'Neill-Biba et al. 2010, Rauscher et al. 2013, Simunovic 2016, Jolly et al. 2021).

Efforts have been made to quantify acquired CVD, notably the 4 type system by Verriest (1963). Unfortunately, given the large range of potential causes and manifestations an elegant acquired CVD classification system, comparable to the parallel classification categories for congenital loss, does not exist. In practical terms, the simplest and most actionable classification is a monocular classification that, for each eye, quantifies the severity of any RG and/or YB loss.

Some drugs, or ‘external agents’, that are known to cause acquired CVD include ethambutol (used to treat tuberculosis), chloroquine and hydroxychloroquine (malaria, autoimmune conditions, and connective tissue disorders), digoxin and digitoxin (cardiac arrhythmias), sildenafil citrate (erectile dysfunction), and vitamin A deficiency (Birch 2001, Parry 2015, Simunovic 2016, Fraunfelder & Fraunfelder 2020). Additionally, a review by Iregren et al. (2002) highlighted that industrial exposure to styrene (used in industrial plastics), perchloroethylene (dry cleaning), and *n*-hexane (vegetable oil extraction and printing) all result in an acquired loss of chromatic sensitivity.

1.4 The value of colour

From an evolutionary perspective the value of trichromatic colour vision has been associated with foraging, specifically the detection of ripe fruit in foliage (Sumner & Mollon 2000, Regan et al. 2001). Although this specific case does still occur, rapid advances in colour display technologies and lighting systems have replaced foraging in foliage with new and novel visually demanding working environments. Such environments have seen an increase in the use of colour, as a direct consequence of technological advancements, and this has spilled over into almost all aspects of society. In the modern world even daily urban navigation relies frequently upon the use of colour for identifying and selecting routes on public transport maps (Parry 2015).

The use of colour for encoding information can enable an individual to swiftly and efficiently identify and group stimuli in visual search tasks – an ability that is essential and safety critical for those who work as air traffic controllers (Barbur et al. 2016). Other commonly noted occupations that have safety critical tasks that rely upon the use of colour include commercial and private aviation, seafarers, and electricians (Barbur et al. 2009, Carter & Barbur 2015, Barbur & Rodriguez-Carmona 2017).

Other, often less considered occupations also utilise colour. Andreani et al. (2021) described the selectively utilisation of colour for the purpose of monitoring and controlling insect populations, such as deer flies. A control mechanism that is valuable in the field of agriculture. For those dealing with antiquities, Farini et al. (2019) demonstrated the value of colorimetric analysis, when combined with the analysis of textiles and weave, in identifying and discriminating between real and fake historical clothing.

Although not the focus of this thesis, one should also note that, in addition to the utilitarian examples considered here, colour also has significant aesthetic value in a number of human activities.

1.5 Outlook

Colour provides and encodes valuable information about one's environment. The ability to see colour is the result of a symphony of parallel biological processes acting in harmony. When the harmony of this perceptual symphony is disrupted the amount of available information, and value of any available information, from one's environment is diminished. Determining the extent to which one can make use of this information and extract the value conveyed by colour, both for pleasure and for safety, is an area of considerable interest and impacts the lives of hundreds of millions of people.

This thesis is centred around the methods that can be used to determine if the symphony produced by the biological orchestra responsible for colour vision, is disrupted, or, in this analogy, out of tune. Such methods allow us to detect, diagnose, and classify 'normal' and 'abnormal' colour vision.

Chapter 2

Colour vision assessment

A broad range of tests and methods for assessing colour vision and colour perception have been developed and used. These methods range from psychophysical tests, such as the Ishihara pseudoisochromatic plate test and the Nagel Anomaloscope, to electrophysiological measurements, such as electroretinograms (ERGs) and visually evoked potentials (VEPs), and genetic tests. A common reason for colour vision assessment is to determine occupational suitability. Community optometrists frequently assess colour vision in children to detect congenital colour vision deficiencies which can impact study and learning efficiency. Colour vision assessment plays an important role in the diagnosis and monitoring of ocular pathology, particularly in cases of diabetes, macular retinopathy, cone dysfunction or drug toxicity.

This chapter describes the various methods that have been developed for the purpose of measuring chromatic sensitivity. The inception of the first colour vision tests, the discovery and development of novel methods of assessment, and the many forms of colour vision assessment used today, both in clinical practice and for research purposes, are discussed.

2.1 Introduction

Clinically, colour vision tests are used to identify, differentiate and measure the severity of congenital and acquired CVD, often with the aim of determining if an individual has sufficient chromatic sensitivity to perform colour-related tasks in a given occupation as well as normal trichromats (Birch & Patel 1995, Margrain et al. 1996, Birch 2001, Cole 2007, Mollon & Cavanus 2012, Rodriguez-Carmona et al. 2012, Dain et al. 2015, Stockman 2018). Although there are hundreds of separate occupational colour vision standards, the fundamental process used to determine if an applicant has the chromatic sensitivity to safely work in a given occupational environment remains relatively constant. The first stage typically involves the applicant completing some form of screening test, which if failed are followed by one or more secondary tests to determine the type and severity of any potential CVD (Birch & Rodriguez-Carmona 2014, Barbur & Rodriguez-Carmona 2017).

Screening tests are the most efficient way of testing a population, both in terms of the time and resources required (Goodman & Livingston 2013, Christ & Herriges 2018). Colour vision screening tests are no exception. They are designed to classify participants into two groups, those with normal trichromatic colour vision and those with CVD. As with other screening tests the value of a colour vision screening test is primarily determined by the test’s sensitivity and specificity. For CVD, an ‘efficient’ screening test has historically been defined as having a “*specificity of 100 per cent*” and preferably a sensitivity “*of over 95 per cent*” (Birch 2001, Chapter 6, page 39). In the context of colour vision screening tests, the sensitivity refers to the probability that the test will correctly identify individuals with a CVD and the specificity refers to the probability that the test will correctly identify individuals with normal trichromatic colour vision (Altman 1991).

Screening tests are often carried out by healthcare practitioners with limited experience, knowledge and technical ability. Most importantly, for a screening test to be widely adopted, not only must it have a high sensitivity and specificity, the test should also be practical. This means the test should be

quick to complete, easy to administer, the task should be accessible and understood by the majority, if not all, of participants and the results should be straightforward to interpret (National Research Council 1981).

Methods of colour vision assessment that aim to classify and grade the type and severity of CVDs can vary markedly in the techniques used and the ease with which the results are interpreted (Rodriguez-Carmona 2006, Birch & Rodriguez-Carmona 2014, Hickey 2015). The differences between each type of assessment are, for the most part, due to the differences in the aim of the tests. As such it is important to always keep the aim of a test in mind when examining its efficacy (and in turn when selecting it for a particular occupational protocol).

The secondary tests that follow the initial screening can range from hue discrimination tests that are conceptually straightforward and easy to administer and interpret, to spectral anomaloscopes that require a trained examiner for effective use (Birch 2001). Some tests are designed for the specific purpose of classifying the type of deficiency (e.g. into protan/deutan categories and/or further classification into protanomaly/protanopia), while other tests are designed to mimic the colour critical tasks that are required when working in particular occupational environments (e.g. the HW-A and HW-B lantern tests for aviation and maritime lookout duties respectively) (Birch 1993, Dain et al. 2015).

The development of clinical colour vision tests has always been linked to changes in occupational working environments and the recent increase in screen-based technologies has been accompanied by an increase in the amount of colour coded information in occupational environments. Perhaps then it is unsurprising that most new colour vision tests are display based tests (discussed in more detail in section 2.3.5), and mimic the ‘natural working environment’ of our time (Hasrod & Rubin 2015, Fanlo Zarazaga et al. 2019). Modern tests are a far cry from the first methods designed to assess chromatic sensitivity, developed over 300 years ago, yet the principles driving the development of colour vision assessment have remained constant since the inception of the earliest forms of colour vision assessment.

To appreciate the wide range of contemporary colour vision tests, and to understand their value to both clinicians and researchers alike, it is helpful to briefly review the history of colour vision assessment. Indeed one of today's most well-known and widely used colour vision tests, the Ishihara pseudoisochromatic plate test, was introduced over 100 years ago, in 1917 (Hardy et al. 1945).

2.2 A technology driven history of colour vision assessment

Prior to the introduction of the Ishihara pseudoisochromatic plate test in 1917 several methods of assessment were developed, with varying degrees of success, for measuring chromatic sensitivity.

Turberville (1684) pioneered one of the first methods for assessing colour vision, using the method of colour naming. The technique of colour naming (asking an individual to verbally identify colours) is still used today in lantern tests, but Turberville's method required participants to name the colours of everyday objects before comparing their answers to the answers given by an observer with 'normal' colour vision (National Research Council 1981). John Dalton (1798) refined Turberville's methodology into a more rigorous clinical test involving 20 coloured ribbons, and Dalton is credited by some to be the first to describe CVD (as Dalton was himself a deuteranope) (Birch 2001, Evans 2003). In 1837, August Seebeck built upon this principle and created a colour vision test, that involved sorting 300 coloured papers into groups (National Research Council 1981, Birch 2001). Several further variants of this test were devised between 1877-1903 using a range of different materials including skeins of wool, small square pieces of coloured cardboard and small beads/pellets (Holmgren 1877, Bickerton 1900).

In 1877 Jakob Stilling, a German ophthalmologist, developed what is nowadays recognised as the first pseudoisochromatic plate test (Stilling 1878, National Research Council 1981). It was known at the time that individuals

with CVD have difficulty finding an object when the object lies in a background where the lightness is differing randomly and the object is defined by its colour (such as in fruit or berries among foliage (Steward & Cole 1989)). Attempts by Stilling's contemporaries to replicate the natural scenarios where this occurs had proved unsuccessful, mainly because it was exceptionally difficult to eliminate all edge artefacts from the target. As a result, one couldn't determine if the observer had detected the presence of the object from its chromatic contrast or from its luminance defined edges. An additional complication arose because an 'isoluminant' stimulus, designed to be isoluminant for an observer with normal trichromatic colour vision, may not be isoluminant for other normal trichromats and, even more so, for subjects with a colour vision deficiency, due to the changes in luminous sensitivity observed in individuals with congenital colour vision deficiency (Wright 1946, Barbur et al. 2008).

To overcome these complications, when designing his test, Stilling made two key changes. First Stilling broke up both the target and the background into a number of discrete patches and then Stilling also varied the luminance of the individual patches (Stilling 1878, Regan et al. 1994). By making these modifications Stilling ensured that all the patches used in his plate test varied in terms of lightness and contour. As a result, the participant cannot make use of luminance cues or edge artefacts to differentiate between the target and the background. Stilling's innovations were the most impactful developments made to the design of pseudoisochromatic plates. These techniques have been incorporated into all subsequent sets of pseudoisochromatic plate tests produced and used today (Regan et al. 1994, Shevell 2003).

In 1881 Lord John Rayleigh introduced the concept of colour mixing to assess an individual's metamerism, with the aim of assessing CVD (Rayleigh 1881). Rayleigh found that some observers made matches that appeared to be very different from those made by most observers, in particular the differences in mixtures of monochromatic red and green light required to match to a yellow. He used this to develop the 'Rayleigh equation'; an equation that Dr Willibald

Nagel made use of in the Nagel anomaloscope, first introduced in 1907 (Nagel 1907).

The first lantern tests were designed by Donders in 1875 (Vingrys & Cole 1983), Edridge-Green in 1891 (Birch 2001) and further refined by Williams in 1903 (National Research Council 1981). These came about because of a fatal train accident in 1875 near Lagerlunda, Sweden (Mollon & Cavonius 2012). The Holgrem wool test proved inadequate, as predicted by Bickerton (1900) at discriminating red-green colour vision deficiency in train drivers as the ensuing tragic accident and court case demonstrated (Topley 1959, Birch & Patel 1995, French et al. 2008). Consequently lantern tests that replicated the occupational colour critical task completed by train drivers, by using the actual coloured signal lights that railway workers needed to identify when working on the railways, were developed (Dain et al. 2015). These lantern tests required observers to verbally identify the colour of a standard light at a fixed distance.

In 1934 Pierce developed the first hue arrangement test, initially used in the National Institute of Industrial Psychology, London (Davies et al. 1934). This new method of colour vision assessment was designed to measure an individual's ability to arrange a matching colour series to determine they possessed 'occupationally safe' colour vision, rather than screening for normal trichromatic colour vision. This method would later be used by Farnsworth in the creation of the Farnsworth-Munsell 100-Hue, Farnsworth D15, and the desaturated Farnsworth D15 test and in turn by Robert Fletcher in the creation of the various editions of the City University test (Farnsworth 1943, 1947, Birch 2001).

The next step change in technology heralded developments of novel techniques in screen-based technologies to enable both psychophysical and electrophysiological tests of colour vision.

One such technique is silent substitution, first described in the literature by Forbes, Burleigh & Neyland (1955) and Biersdorf & Armington (1957). There is some dispute about the originator of the concept (Estévez & Spekreijse 1974). However, Estévez & Spekreijse (1982) state that William A. H. Rush-

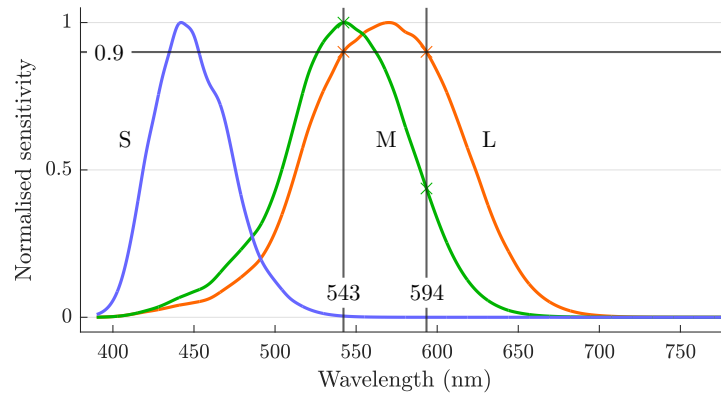


Figure 2.1: The principle behind silent substitution techniques (adapted from Foster (2010)). Two monochromatic lights with wavelengths of 543nm and 594nm produce the same L-cone response but a different M-cone response. By alternating between the two monochromatic lights it is possible to selectively excite a single class of cone photoreceptor.

ton is “*without a doubt... the one who, through his own work, definitely established ‘univariance’ and ‘light substitution’ as permanent research tools of the visual scientist*” (Estévez & Spekreijse 1982, page 682). Silent substitution exploits the theory of univariance to isolate one class of cone photoreceptors. The principle of univariance was summarised by Mitchell & Rushton (1971).

“For each class of receptor the result of light depends upon the effective quantum catch, not upon what quanta are caught”

(Mitchell & Rushton 1971, page 1041)

In other words, a single cone photoreceptor is excited to the same extent by a number of different combinations of wavelength and intensity. As a result, a single cone photoreceptor in isolation, cannot distinguish changes in the wavelength of a light from changes in its intensity (once a photon has been absorbed). Thus, by alternating between two monochromatic stimuli one can, theoretically, selectively excite a single class of cone photoreceptor cells. The use of silent substitution for the purpose of isolating single cone photoreceptors is shown graphically in Figure 2.1.

Since its introduction silent substitution continues to be used for isolating photopigments and is one of the primary methods of investigating melanopsin contributions to the human pupillary light response (Spitschan & Woelders 2018). Colour assessment techniques which make use of silent substitution

include flicker electroretinography (ERG) and psychophysical flicker photometry.

High frequency flicker ERG was first analysed in congenital colour vision deficiencies by Dodt, Copenhaver & Gunkel (1958). Numerous studies have followed since including a small sample size study demonstrating the potential of this technique to discriminate between dichromats and trichromats (Zrenner 1984, Usui et al. 1998). Flicker photometry ERG has also been used to measure the effects of chromatic adaptation on the spectral sensitivity function (Padmos & van Norren 1971, Regan 1974). In 1993 Crognale et al. also demonstrated the potential for the stimulation of selective chromatic pathways by measuring the VEPs generated by grating patterns (Crognale et al. 1993). Both ERGs and visually evoked potentials (VEPs) are extremely useful for identifying malingering and in cases where eliciting an active response is challenging (i.e., subjects with mental impairments or subjects that have difficulties with communication) (Jägle et al. 2003). However both methods are time consuming and require specialist equipment and trained examiners to carry out the measurements and interpret the results (Jacobs & Neitz 1993, Gerling et al. 1997, Jacobs et al. 1996).

The Objective Screening of Color Anomalies and Reductions (OSCAR) test and the Medmont C100 test, first introduced and described in 1983 and 1992 respectively, are designed to utilise silent substitution to screen for colour vision deficiencies (Estévez et al. 1983, Metha & Vingrys 1992). The OSCAR test is no longer produced commercially, but Le Sueur et al. (2014) evaluated both tests and concluded that the Medmont C100 is a suitable replacement for the OSCAR test. During both tests participants are shown red and green lights modulated in counterphase and asked to adjust the modulation so as to minimise the perceived flicker. Neither the OSCAR nor the Medmont C100 reliably distinguish between normal trichromats and individuals with a deutan deficiency (Zisman et al. 1987, Le Sueur et al. 2014). Both tests do, however, reliably distinguish between protans and deutans and provide an estimate of individuals' L:M cone ratio (Verriest & Uvijls 1989, Metha & Vingrys 1992, Danilova et al. 2013, Lawrance-Owen et al. 2014). Both tests

have also been reported to be able to identify carriers of congenital protan deficiencies (Jordan & Mollon 1997, Harris & Cole 2005). When evaluated on their capacity to screen for CVD, the tests suffer from similar limitations, primarily due to the natural variation in both L:M cone ratios and spectral sensitivity of cone photopigments (Le Sueur et al. 2014).

In direct contrast to monochromatic sources employed in silent substitution and instruments like the anomaloscope the use of ‘comb-filtered’ spectra have also been described for the quantification of chromatic discrimination (Bonnardel et al. 1996, Bonnardel & Varela 2003). First described by Barlow in 1982 comb-filtered spectra are illuminants with sinusoidal spectral power distributions (Barlow 1982, Bonnardel et al. 1997). Following Barlow’s initial description in 1982, Bonnardel, Martinoya & Varela (1991) described a device that allowed for the independent control of the amplitude, phase, and frequency of square wave modulations and Bonnardel & Varela (1991) used this device to measure the chromatic contrast sensitivity function in three normal trichromats in 1991. More recent devices designed to produce comb-filtered spectra have been described in 1996 and 1998 (Bonnardel et al. 1996, Bonnardel 1998). Discrimination thresholds measured using spectra produced by these devices have been used to estimate the spectral modulation sensitivity function of the human colour vision system for several normal trichromats and individuals with congenital CVD (Bonnardel et al. 1996, 1997, Bonnardel & Valero 2001).

In 1989 John Mollon and Jeremy Reffin adapted the techniques developed by Stilling in the ‘Cambridge Colour Test’ (CCT). Instead of the paper design of Stilling, the Landolt C stimuli in the CCT are controlled by a computer display. During the test the chromaticity and orientation of the stimuli are varied in order to measure the observers’ chromatic threshold along three predetermined axes of colour space (Mollon & Reffin 1989).

A further innovation in computerised testing was the first published use of ‘random luminance masking’ (RLM) for the purpose of colour assessment by Barbur, Birch & Harlow (1992). This technique employs moving, ‘isoluminant’ coloured stimuli buried in dynamic luminance contrast (LC) noise to

mask more effectively the use of residual LC signals (Birch et al. 1992, Barbur et al. 1993, 1994). The development of this technique led to the Colour Assessment and Diagnosis (CAD) test. Although previous publications by Barbur et al. had also made use of this technique to isolate the use of chromatic signals, these publications had focused primarily on pupil responses (Barbur, Birch & Harlow 1992, Barbur, Harlow & Sahraie 1992). More information regarding the origin of this technique and the work that led to the use of RLM as a method of colour vision assessment, is discussed in section 2.3.5.2 - Colour Assessment and Diagnosis (CAD) test (page 48). In March 2009 the CAD test was made commercially available through a City, University of London spin out company ‘City Occupational Limited’.

In 1986, following the isolation and sequencing of the cone photopigment opsin genes, genetic testing became a new method that could be used to identify both individuals with and carriers of congenital colour vision deficiency mutations (Nathans et al. 1986, Stockman 2018). Thirty years later, in 2016, a large scale study (N=1872) highlighted the viability of using genetic testing as an method of diagnosing congenital colour vision deficiency in large populations (Davidoff et al. 2016). Similar to ERGs and VEPs genetic testing is particularly useful because it offers an objective insight into an individuals colour vision. Genetic testing can also be used to detect carriers of congenital colour vision deficiency, along with the OSCAR and Medmont C100 tests, but it is unable to detect acquired colour vision deficiencies (unless the origin of the acquired deficiency is due to a genetic mutation). However, genetic testing introduces numerous ethical concerns, particularly if genetic data are used for occupational screening.

In summary, the past 350 years have seen significant changes in the methods used to measure chromatic sensitivity and diagnose colour vision deficiencies, in no small part due to the striking technological advances over this time. From the wool and ribbons of Tuberville and Dalton to modern computerised tests and genetic testing, the fundamental goals of colour vision assessment, to assess the severity and classify the type of any congenital or acquired CVD, have remained unchanged.



Figure 2.2: Timeline showing the development and introduction of some of the methods designed and employed for human colour vision assessment.

Unsurprisingly in a clinical setting, where time and efficiency are critical factors, only a small selection of the many available tests are regularly used. In most cases these clinical tests are carried out to determine occupational suitability. They have to be easy to use, simple, straightforward and, at least perceived as being, an accurate indicator of an individual’s level of chromatic sensitivity and thus an accurate indicator of the ability to work unimpeded by any potential colour vision deficiency in a given occupational environment (Margrain et al. 1996, Carter & Barbur 2015, Barbur & Rodriguez-Carmona 2017, Stockman 2018).

2.3 Protocols for comprehensive colour vision assessments

Historically, the recommended strategy for a comprehensive colour vision assessment involved completing a battery of tests (Verriest 1968). The types of tests included in the test battery depend upon the purpose of the examination, as often each colour vision test would have a specific function and the tests included in the examination may be limited by *“availability, location, cost, the time available and the ability of the subject to respond”* (Birch 2001, Chapter 7, page 82).

City, University of London has provided colour vision assessment to the general public since the 1970s along with other vision tests designed to assess other aspects of visual performance. From 2007 onwards colour vision assessments at City, University of London, together with a range of other tests, have been provided through the Advanced Vision and Optometric Test (AVOT) service. In 2022 several different types of assessment are provided. The standard ‘full’ colour vision assessment employs a test battery that includes six of the most commonly used tests in occupational colour assessment protocols. A list of the tests included in a ‘full’ colour vision assessment are shown in Table 2.1.

A) Test descriptors			
Test	Type	Version	Function
Ishihara AO-HRR	Pseudoisochromatic plate	38-Plate 2nd Ed.	Screening Screening
City University Farnsworth D-15	Hue discrimination	2nd Ed. -	Grading and classifying Grading and classifying
Nagel CAD	Spectral anomaloscope Visual display	Type I v2.8	Classifying Grading and classifying
B) UK occupations which require/accept each test			
Test	Occupation		
Ishihara	Aviation, Maritime & Coastguard Agency (MCA), Police, Fire Service, Railways, Electrical Contractors		
AO-HRR	-		
City University Farnsworth D-15	MCA (Engineering/Radio roles), Police MCA (Engineering/Radio roles), Police, Fire Service		
Nagel CAD	Aviation, Fire Service, Police (Special cases) Aviation, MCA, Electrical Contractors, Armed Forces		

Table 2.1: A) Tests routinely carried out at City, University of London as part of a ‘full’ colour vision assessment in mid 2022 and B) occupational colour vision standards which use each test.

2.3.1 Pseudoisochromatic plate tests

Pseudoisochromatic plate tests are the most widely used screening tests for abnormal colour vision (Belcher et al. 1958, Birch 1997*b*, 2001, Fanlo Zarazaga et al. 2019). They are probably the simplest type of test to use and are typically presented in book form. Pseudoisochromatic plate tests are based largely on the principle of colour confusion typically requiring the verbal identification of a coloured figure or object, although some plates require the tracing of patterns or objects (Barbur & Rodriguez-Carmona 2017, Rodriguez-Carmona & Barbur 2017). The efficiency of each plate depends upon the choice of colour, positioning of the dots, the dimensions of the figure and the luminance contrast between the elements contained in the target and those in the background. The efficiency of the entire test also varies, depending upon the pass protocol employed and the plates selected (Birch 1975, 2010, Rodriguez-Carmona & Barbur 2017).

There are many different types of pseudoisochromatic plate tests that have been developed. Two of the most widely used are discussed here: the Ishihara plates and the American Optical Corporation, Hardy, Rand and Rittler (AO-HRR) plates.

2.3.1.1 The Ishihara plates

First published in 1917 the Ishihara plates are the most widely accepted screening test for congenital red-green CVD (Belcher et al. 1958, Birch 1997*b*), although the test is not designed to identify yellow-blue deficiencies (Birch 2001, Rodriguez-Carmona 2006). Several editions have been published and whilst the quality of colour printing varies slightly in different editions the differences between editions has been found to have a minimal effect on test performance (Birch & McKeever 1993).

The standard version of the test has 38 plates, with 25 numerical designs and 13 pathway designs intended for cases where subjects are unable to verbally identify the numerical designs. Although two reduced versions of the tests are available, a 24 plate edition and a 14 plate edition, both of which use a reduced selection of the 38 plates, the 38 plate edition of the test is recommended for clinical use (Birch & McKeever 1993). It is also worth noting that there is another version of the Ishihara test for ‘unlettered persons’ introduced in Europe in 1990, in which shapes are used in place of the numerical designs (Birch 2001). Whilst this test for ‘unlettered persons’ will not be discussed here, a study by Birch & Platts (1993) found the test to be an effective method of colour vision screening in children under the age of 7 years ($N=513$).

The Ishihara plates are designed to be viewed at a distance of approximately 75cm with a viewing time of 4 seconds allowed for each plate. The subject is asked to verbally identify any number they can see for each page and subjects are made aware that some pages may appear blank. The test is designed to be viewed under daylight, however given the variation of ‘daylight’ with location and time Birch (2001) recommended the use of Commission Internationale de l’Eclairage (CIE) illuminant C (or a close approximation) in clinical settings.

Category	Plates	Function	Description
Demonstration	1	Identify malingering	To demonstrate the task and identify malingering
Transformation	2-9	Screening	A number will be seen by normal trichromats and a different number will be seen by some subjects with a red-green CVD
Vanishing	10-17	Screening	A number will be seen by normal trichromats, but no number will be seen by subjects with a red-green CVD
Hidden digits	18-21	Screening	No number will be seen by normal trichromats but some subjects with a red-green CVD will be able to see a number
Classification	22-25	Classification	Depending upon the type of CVD the subject will be unable to detect either the left or the right hand number

Table 2.2: The design and function of the 25 numerical designs of the 38-plate Ishihara test (Hardy et al. 1945, Birch 2001).

As shown in Table 2.2, the numerical designs of the Ishihara plates can be grouped into five distinct categories (classified by Hardy, Rand & Rittler (1945)).

During a normal clinical examination, the number of errors obtained on the first 25 numerical plates are recorded. Occupational environments generally have their own unique pass criteria, in which they specify the number of accepted errors for a given range of Ishihara plates. Unfortunately, this can lead to confusion as potential applicants may pass one occupational protocol yet fail another, often with little or no justification or explanation. As an example of the variation in pass protocols: the 2018 UK Police protocol allows up to 4 errors on the first 21 plates of the 38 plate Ishihara test (Stockman 2018), the Joint Aviation Requirements (JAR – a list of requirements for aviation certification agreed by several countries in Europe) allow no errors on the first 15 plates of the 24 plate edition, the USA Federal Aviation Administration allow up to 8 errors on the first 21 plates of the 38 plate Ishihara (Rodriguez-Carmona et al. 2012), and the UK fire service accepts up to 2 errors on the first 17 plates of the 24 plate Ishihara test (Margrain et al. 1996).

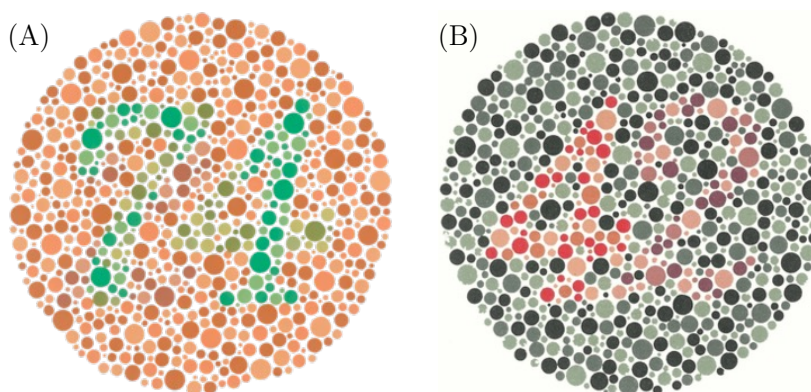


Figure 2.3: Plates used in the 38-plate edition of the Ishihara test including a (A) Transformation and (B) Classification plate. For (A) a normal trichromat would see the number ‘74’ whilst a subject with red-green CVD is expected to see the number ‘21’. For (B) deutan subjects are expected to identify the left-hand number ‘4’ and protan subjects are expected to only identify the right-hand number ‘2’.

The variation in the pass protocol employed is due to an inherent quirk of the Ishihara test and it is a quirk that is present, to some extent, in all screening tests. It is the problem of trying to maximise both sensitivity and specificity (i.e., detecting all subjects with CVD whilst also correctly passing all normal subjects). Birch (2001) reported that if one wishes to pass 100 per cent of normal trichromats, one needs to allow for up to 6 errors on the Ishihara test. In accepting 6 errors approximately 7% of deuteranomalous subjects pass (Birch 2001). Alternatively, by accepting 0 errors approximately 19% of normal trichromats fail (Rodriguez-Carmona et al. 2012). Clinically it is often assumed that the number of errors made on the Ishihara test corresponds to the severity of the colour vision deficiency, however multiple studies have demonstrated that this assumption is incorrect (Rodriguez-Carmona et al. 2012, Miquilini et al. 2019). Whilst the Ishihara test is a rapid and accurate method for screening for red-green (RG) colour vision deficiency, the information that can be obtained from examining the errors (both in terms of severity and classification) is limited (Rodriguez-Carmona & Barbur 2017).

Several modifications to the testing procedure have been suggested in attempts to increase the sensitivity and specificity of the Ishihara plates. Notable attempts were made by Rodriguez-Carmona et al. (2012) and more recently by Miquilini et al. (2019). However even with the introduction of

these modifications the overall benefits to the protocol remained relatively low (Miquilini et al. 2019).

2.3.1.2 The AO-HRR plates

Since its introduction in 1954 the American Optical - Hardy, Rand and Rittler (AO-HRR) pseudoisochromatic plate test has become the second most prominent pseudoisochromatic plate test (behind the Ishihara test). The AO-HRR plates aim to identify protan, deutan and tritan deficiencies and to grade their severity (Hardy et al. 1954, Birch 2001). An advantage of the AO-HRR plates over the Ishihara plates is that the AO-HRR includes plates designed to test for yellow-blue deficiencies, potentially enabling the test to detect acquired colour vision deficiencies where, in most cases, yellow-blue loss is more prevalent (Simunovic 2016).

The principle behind the test is the same as the Ishihara test, and all pseudoisochromatic plate tests, though the AO-HRR plates replaces the numerical stimuli used in the Ishihara test with geometric shapes in the form of a circle, a triangle and a cross. The saturation of the colours used in the shapes increase throughout the plates, resulting in the last plates having the largest chromatic contrast. Each shape is printed in neutral colours on a background matrix of grey dots and the shape is always located in one of the four quadrants of the page, with a maximum of 2 shapes per page (Hardy et al. 1954). The 24 plates are divided into four demonstration plates and 20 examination plates. The 20 examination plates can be further subdivided into 6 ‘yellow-blue’ plates designed to screen and grade yellow-blue CVD and 14 ‘red-green’ plates designed to screen and grade red-green CVD. All of the examination plates are of the vanishing design and normal subjects are expected to identify all the shapes present on all pages.

The procedure for examination is similar to the Ishihara plates, with the same recommended illumination, viewing time, and viewing distance for each plate. For each page subjects are asked to identify all visible shapes and their location. The test is normally administered in reverse order so that it has the

perception of becoming more challenging as the test progresses (Birch 2001). The number of errors made are recorded and a diagnostic analysis can be made on the basis of the results for plates 7 to 20 (National Research Council 1981).

2.3.2 Lantern tests

Lantern tests are unique in that they do not specifically screen for CVD, instead they are designed to measure the ability of individuals to identify and discriminate navigational aids and signals (National Research Council 1981, Dain et al. 2015). As such, lantern tests are easy to administer, utilising the concept of colour naming, and their value is in their ability to simulate the most demanding use of signal lights in some occupational environments.

Whilst a number of lantern tests have been produced, only the Holmes-Wright lanterns contain the actual signal colours used in navigational aids and signals (Birch 2001). Both the Holmes-Wright (HW) type ‘A’ and ‘B’ lantern tests are discussed here (Holmes & Wright 1982).

2.3.2.1 The Holmes-Wright lanterns

In recent years both the HW type ‘A’ and type ‘B’ lanterns have been phased out of use by the UK Civil Aviation Authority (CAA) and Maritime Coastguard Agency (MCA) respectively (Barbur et al. 2009, Carter & Barbur 2015). Neither of these lantern tests are currently produced (Dain 2004). Both the HW-A and HW-B lantern tests require participants to verbally identify the colour of two lights presented simultaneously in either a horizontal (HW-B) or vertical (HW-A) line. The lanterns are normally viewed at a distance of 6 metres (Birch 2001). The colours are shown in nine combinations with each combination typically being repeated 3 times in each light condition and all the colours used have x,y chromaticity co-ordinates within the CIE specifications for signal lights (Vingrys & Cole 1983).

The HW type ‘A’ lantern is currently accepted as a form of assessment in some aviation environments (Squire et al. 2005, Koefoed et al. 2020). There are

three luminance settings for the HW-A, ‘Demonstration’ (2000 μcd), ‘High’ (200 μcd) and ‘Low’ (20 μcd), with the demonstration setting only being used initially to demonstrate a red, green and white light for the participant (Birch 2001). The test aperture subtends ~ 0.9 arcmin (Carter & Barbur 2015). Different testing protocols can be employed. In the most well-established protocols the test is carried out in normal room illumination using the ‘high’ luminance setting and the nine pairs of colours are shown three times (for 27 total presentations). If any errors are made during the first 27 presentations, the subject is dark adapted for 15 minutes before the test is repeated. The luminance setting used following dark adaptation varies between ‘high’ and ‘low’ for different testing environments and a recommended test procedure was described in (Birch 2008*c*).

Until 2018 the Holmes-Wright type B lantern test was used by the MCA as a supplementary test for colour vision provided on appeal when seafarers failed the Ishihara test (Maritime and Coastguard Agency 2020). Unlike the HW-A lantern the HW-B lantern requires applicants to identify colours in two separate aperture sizes. Initially applicants identify a single coloured light using a large aperture (~ 2.9 arcmin) before identifying two colours showing using a smaller aperture (~ 0.3 arcmin) (Vingrys & Cole 1983). The entire test is performed following a minimum of 8 minutes of dark adaptation and for some testing environments the testing distance is reduced to 3 metres (Birch 2001).

The HW-B is more sensitive than the HW-A with a sensitivity of 0.98 compared to 0.83, however this is at the cost of specificity. The HW-B fails approximately 8% of normal trichromats when up to 5 errors are accepted and approximately 28% of normal trichromats if no errors are accepted (Vingrys & Cole 1983). By contrast, the HW-A passes all normal trichromats and between 1%-17% of participants with CVD (Vingrys & Cole 1983, Hovis & Oliphant 1998, Squire et al. 2005). As noted by Birch & Rodriguez-Carmona (2014), the use of different lanterns with different pass rates results in HW protocols lacking internal consistency (Birch 1999, 2008*c*, Birch & Rodriguez-Carmona 2014).

Although an advantage of the HW lantern tests is that they are designed to simulate occupational tasks, both HW lanterns possess a number of limitations. These include the potential memorisation of presentations, particularly if additional cues are available, the lack of translatability of the results to other applications or settings, and the technical challenges associated with maintaining equipment which is no longer manufactured.

2.3.3 Hue discrimination tests

Hue discrimination tests (of which arrangement tests are a subset) usually involve either the identification, or the sorting of colour samples based upon perceived differences in hue, lightness and saturation. These tests are commonly used in clinical environments as they are designed to detect all types of colour vision deficiency and can be used to examine both RG and YB congenital and acquired deficiencies. Three of these tests, the Farnsworth-Munsell 100 hue test (FM 100-hue), the Farnsworth D-15, and the City University test, are discussed here.

2.3.3.1 Farnsworth-Munsell 100 hue

The Farnsworth-Munsell 100 hue (FM 100-hue) and the Farnsworth D15 test were both initially described by Farnsworth in 1943 before being introduced in 1947, though the latest version of the FM 100-hue was published in 1957 (Farnsworth 1943, 1947, 1957). Both tests were originally intended to identify individuals with severe colour vision deficiency who are likely to experience practical difficulties in specific occupational environments. As a result, the tests are similar in design, administration and in the interpretation of the results (results are interpreted graphically using a results diagram for both tests) National Research Council (1981). As per the manufacturers' instructions, the test should be used with Illuminant C (or a close approximation) to ensure the accurate reproduction of the colours involved National Research Council (1981).

The FM 100-hue test consists of 85 Munsell samples placed into small caps separated into four groups (or boxes). Each cap subtends 1.5° at the test distance of 50cm. At the ends of each of the four boxes are two reference caps. The subject is required to arrange the remaining caps in hue order between the references making a ‘natural’ colour order. The positional order the caps are placed into is then recorded and plotted onto a results diagram from which an error score can be calculated. Computerised scoring is also available.

Although commonly used in published clinical work (Fanlo Zarazaga et al. 2019) the test is unsuitable for use as a screening test with Birch (1989) advising against the use of the test in isolation. Although Birch (1989) did recommend the use of the FM 100-hue when used as part of a test battery for the assessment of both acquired and congenital CVD, it was noted that individuals have been known to improve their score by up to 30% when retested (Birch 1989, 2001). Despite this, the test is highly sensitive. Verriest (1963) capitalised upon this sensitivity and showed that the mean FM 100-hue error score reaches a minimum at 20-24 years of age, thereafter increasing as a function of age. Such findings would later be replicated using other sensitive methods of colour vision assessment (Barbur & Rodriguez-Carmona 2015).

Limitations of the FM 100-hue include the positive correlation with non-verbal IQ and test performance (particularly with individuals that have autism spectrum disorders) and the time taken to complete the test (Cranwell et al. 2015). The correlation between non-verbal IQ and test performance may be associated with number of factors including an increased understanding of the test procedure and the ability to sustain focus and switch between global and local processing of visual information to both oversee the entire colour gradient and compare adjacent caps.

The test has been reported to take between 15-30 minutes to complete, although modifications to the testing procedure have been suggested to optimise the assessment process and reduce the time taken (Ghose et al. 2014, Murphy 2015).

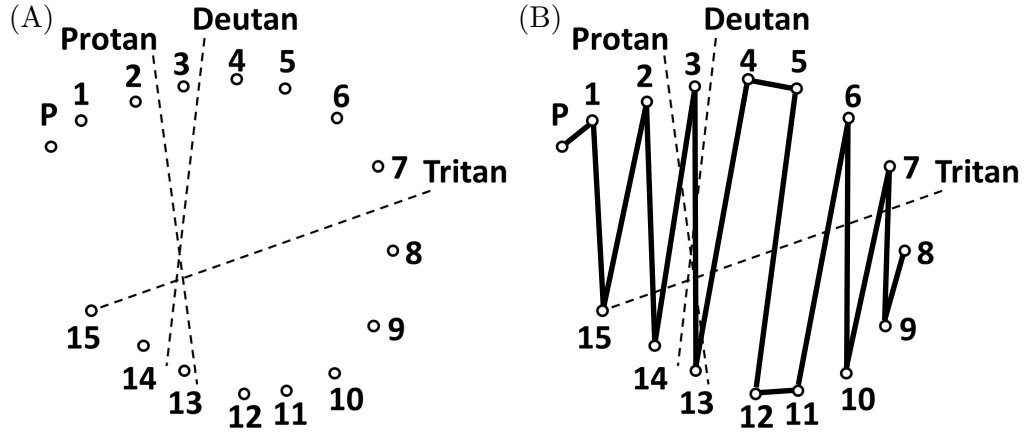


Figure 2.4: Circular results diagram used to analyse and interpret the results of the Farnsworth D-15 test. The test is analysed by plotted a subject’s cap order onto a blank diagram (A). Lines are drawn between the number of each cap in the order in which they were arranged and example results for an individual with a deutan deficiency are shown in (B).

2.3.3.2 Farnsworth D-15

The standard Farnsworth D-15 test consists of 16 Munsell hues, mounted in the same caps used in the FM 100-hue test (Farnsworth 1947). During the test participants are provided with a sole fixed reference or ‘pilot’ cap and are asked to select and arrange the remaining caps, one by one, in each instance selecting the cap that appears to be least-different perceptually to the most recently selected cap (Birch 2001).

In clinical settings the results of the Farnsworth D-15 are normally interpreted by plotting a participant’s arranged sequence on a circular results diagram (Birch 2008a, Evans et al. 2020). Errors made on the Farnsworth D-15 are classified as being either adjacent transpositions (i.e., $1 \rightarrow 3$) or major isochromatic crossings (i.e., $1 \rightarrow 15$) depending upon the appearance of errors when plotted on the circular results diagram (Figure 2.4).

Farnsworth D-15 protocols vary between accepting 1 to 2 adjacent transpositions to allowing up to 2 major isochromatic crossings (National Research Council 1981, Birch & Chisholm 2008). Monochromats, dichromats, and anomalous trichromats with reduced or absent chromatic sensitivity are expected to make multiple crossings. Those with normal trichromatic colour vi-

sion, or close to normal chromatic sensitivity, are expected to make, at most, one to two adjacent transpositions (Atchison et al. 2003, Birch & Chisholm 2008, Dain et al. 2015).

Although in theory quite a powerful and efficient test, the Farnsworth D-15 has a number of limitations. Several were highlighted by Dain, Atchison & Hovis (2019) and are described in more detail, along with a large statistical analysis of the test outcome and an exploratory study carried out using the Farnsworth D-15, in chapter 3 (sections 3.2 & 3.3).

2.3.3.3 City University test

The City University Colour Vision Test is designed to screen for moderate to severe red-green and yellow-blue deficiencies and to differentiate between protan, deutan and tritan deficiencies. The first two editions of the test aim to provide an equivalent examination procedure to the Farnsworth D15 test in booklet form (Fletcher 1972, 1978). The agreement between the second edition and the Farnsworth D-15 was found to be high if the failure criteria are set as ≥ 5 errors on all City University test plates OR ≥ 2 errors on the large plates and ≥ 2 major crossings on the Farnsworth D-15 test (Oliphant & Hovis 1998).

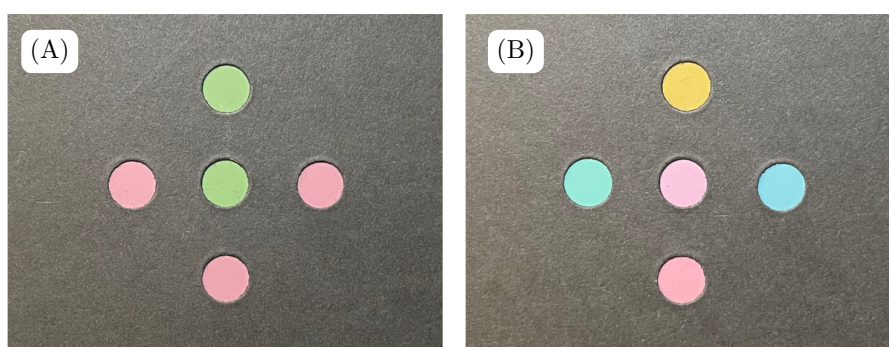


Figure 2.5: The Einstein cross plate design used for all plates in the City University test (2nd Edition), including the (A) demonstration and (B) first plate.

The second edition of the test contains 11 plates. Each plate presents five Munsell samples, a central reference colour surrounded by four test colours in an Einstein cross pattern, displayed on a black matt background (Figure

2.5). Similar to the Ishihara and AO-HRR pseudoisochromatic plate tests the first plate is a demonstration plate, followed by six ‘larger’ and four ‘smaller’ designs subtending 1.8° and 0.8° at 30cm respectively (Birch 1997*a*). Subjects are required to identify the ‘outside’ or peripheral colour they perceive as being the most similar in colour to the central reference colour. Each possible option is either an isochromatic confusion associated with protan, deutan, tritan deficiencies or the ‘correct’ option designed to replicate the next hue in the Farnsworth D-15 hue sequence (Birch 2001).

The third edition of the test introduces several plates designed for colour vision screening. Whilst the first two editions of the test are no longer produced there is some dispute over the efficiency of the third edition for use in occupational protocols (Birch 2001) and the agreement between the third edition and the Farnsworth D-15 has been reported to be lower than the second edition (Almustanyir et al. 2022). As a result in some occupational protocols, although both the second and third editions of the test are accepted, the third edition is only used if the examiner does not have access to the second edition (Stockman 2018).

The second edition of the City University test produces ‘poor’ or ‘mixed’ deutan and protan classifications with Birch (1997*a*) reporting that over 60% of those who fail the City University test obtain mixed protan and deutan classifications (Honson & Dain 1987, Birch 1997*a*). Despite its limitations the City University test remains popular, particular in optometric practice, for its perceived value as a method for dichotomising patients on the basis of their chromatic sensitivity. This popularity has, at least partially, translated to its continued use in occupational colour vision assessment (Margrain et al. 1996, Stockman 2018).

The use and implications of the continued use of the City University (2nd Edition) and the Farnsworth D-15 tests in single- and multi-test colour assessment protocols is investigated and explored in chapter 3.

2.3.4 Spectral anomaloscopes

Anomaloscopes are optical instruments that utilise the method of colour matching pioneered by Lord Rayleigh in 1881 for the purpose of classifying and diagnosing colour vision deficiencies (Rayleigh 1881, Thomas & Mollon 2004). An anomaloscope presents a circular target divided by a horizontal diameter. The subject adjusts two control dials until the perceived colour and brightness of the top and bottom fields are matched. Since their introduction anomaloscopes have been considered to be the standard reference test for identifying and diagnosing red-green colour vision deficiencies (National Research Council 1981).

The efficient and valid use of an anomaloscope requires the extensive training of examiners (Birch 2001). In the hands of a skilled examiner the anomaloscope is considered to be one of the most accurate methods of classifying the type of red-green congenital colour vision deficiency. Due to the need for operator training and the time required for testing, anomaloscopes are not a common form of clinical colour vision assessment. However, anomaloscopes are used extensively in research settings, normally as a reference that other forms of testing are compared to (Cole & Maddocks 1998, Birch 1999, Squire et al. 2005).

While a Rayleigh match, or Rayleigh equation, is only designed to investigate red-green colour vision deficiencies (Rayleigh 1881), various other matches (or equations), such as the Pickford-Lakowski match and equation, have been developed for the purpose of evaluating acquired colour vision deficiencies (Lakowski 1972). These alternative matches can be used in anomaloscopes to investigate acquired and tritan deficiencies. However, only the most well-known anomaloscope, the Nagel anomaloscope (Type I) is discussed here.

2.3.4.1 The Nagel Anomaloscope

The Nagel anomaloscope is used to determine the Rayleigh equation, the mixture of red and green monochromatic lights required to match a monochromatic yellow light, for an observer. Since its introduction in 1907 the Nagel

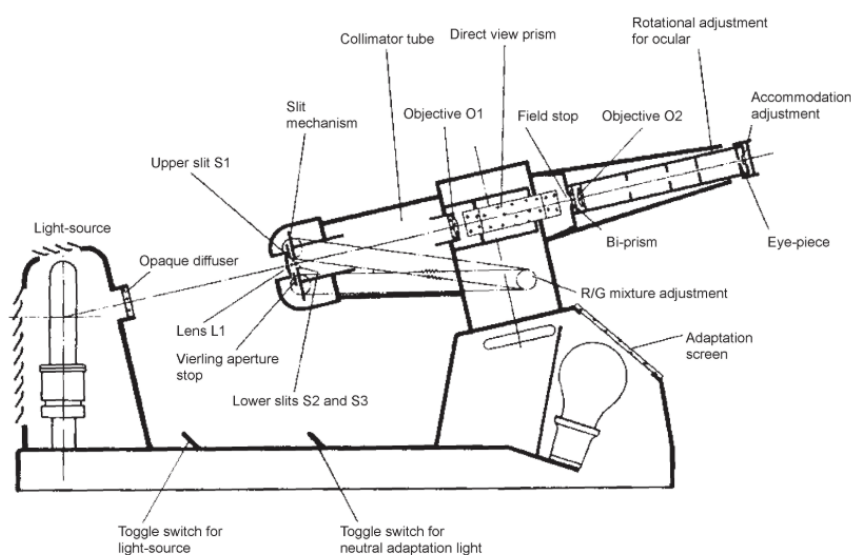


Figure 2.6: Schematic diagram of the Nagel Type I anomaloscope adapted from Jägle, Pirzer & Sharpe (2005) who in turn adapted the diagram from the original version which appears in the manufacturer’s handbook.

anomaloscope has stood almost uncontested as the ‘Gold standard’ or ‘standard reference’ for the classification of congenital red-green colour vision deficiencies (gold standards are discussed in more detail in section 2.4.2) (National Research Council 1981, Birch 2001, 2008*a*, Hickey 2015). Whilst the Nagel anomaloscope is renowned for its accuracy in distinguishing between protanomalous and deuteranomalous observers (Barbur et al. 2008) the relationship between the parameters of the match and the subject’s overall colour discrimination sensitivity is known to be generally poor (Nelson 1938, Hurvich 1972). As a result *“even the anomaloscope, therefore, is not by itself sufficient to grade the anomaly with certainty”* (Wright 1946, Chapter 24).

The Nagel anomaloscope is made up of *“a light source, a separate internal adaptation light, an ocular, a compound direct-vision prism and three entrance slits”* (Jägle et al. 2005, page 26). These components are used (in the configuration outlined in Figure 2.6) to produce a Maxwellian view system in which a subject observes a circular field, split into two; one half illuminated by a monochromatic yellow light (589nm) and the other a mixture of monochromatic red (670nm) and green (546nm). The three entrance slits are used to maintain a constant luminance for the red-green semicircle (Birch 2001). Both the temperature of the Nagel anomaloscope and the voltage used have been

shown to have an effect on the results obtained (Jordan & Mollon 1993, Jägle et al. 2005). The reported size of the field of view in the instruments produced and in use, varies slightly in the literature; some sources describe the visual field as having a diameter of 2° (Helve 1972) whilst others report the visual field of their Nagel anomaloscope to be 3° (Birch 2001, Barbur et al. 2008). Whilst this variation in individual instruments may seem to be a minor point, it could influence the results obtained (Pokorny & Smith 1976).

During the test, subjects are initially asked to make several colour matches by adjusting both the luminance of the yellow test field and the red-green mixture ratio. These initial matches are recorded by the examiner. Following the initial matches the subject is asked to make an ‘exact’ match for a range of red-green mixtures by only controlling the luminance of yellow test field. The red-green mixture ratio is controlled by the examiner, who selects a variety of different red-green ratios and records the subject’s matching range. To minimise any potential adaptation effects the subject is instructed to look away from the instrument for several seconds between each match.

In addition to the variation in visual field size between instruments there is also variation between the red-green mixture ratio and the luminance values between instruments. As such, each instrument has a slightly different ‘normal’ matching range. The normal range for a given instrument is determined by examining a large number of normal subjects to establish the normal range, and the mean and standard deviation of the range (Birch 2001). Generally normal trichromats will make a match within a small range of red-green mixture ratios and subjects with colour vision deficiency will produce matching ranges outside of the range established for normal subjects (Rodriguez-Carmona et al. 2005, Rodriguez-Carmona & Barbur 2017). In all cases the luminance of the yellow setting provides information about the subject’s relative luminous efficiency.

There are three properties of an individual’s matching range that can help to inform the examiner and assist in the diagnosis and classification of any potential red-green colour vision deficiency. These properties are; the limits

of the matching range, the mid-point of the range and the range of luminance values used in the match.

Deuteranomalous individuals require more green light in the red-green mixture (relative to a normal observer on the same instrument) and individuals with protanomaly exhibit the opposite effect, requiring more red light in the red-green mixture. Consequently the matching ranges of protanomalous and deuteranomalous trichromats form two separate distributions (Birch 2001, Thomas & Mollon 2004). Protanopes and deuteranopes will match at both extremes and the two can only be differentiated by examining the luminance of the yellow field used at each extreme. In addition to subjects that present as deuteranomalous and protanomalous there are also subjects categorised as ‘extreme anomalous trichromats’. These subjects produce matching ranges that are neither normal, deuteranomalous or protanomalous, and are unable to match at both extremes (Birch 2003).

Historically the size of the matching range has been used as a measure of the severity of RG loss (Birch 2008*a*). However, studies have shown the relationship between the severity of RG loss and the parameters of the match to be poor (Wright 1946, Hurvich 1972). Modelling work by Thomas & Mollon (2004) and Barbur et al. (2008) demonstrated that the Nagel anomaloscope matching range is affected by multiple factors including the optical density of the photopigments, the L:M cone ratio, and the noise of the L/M colour channel.

2.3.5 Display-based tests

Digital displays have permeated almost all aspects of daily life, both recreationally and in the workplace. Colour vision testing is no exception and a number of screen-based colour vision tests have been developed. These either try to emulate clinical colour tests such as the Ishihara plates, the HRR, the City University test and the Farnsworth D15 (Pacheco-Cutillas et al. 1999), or isolate chromatic signals, making use of the temporal and spatial control of stimuli of high resolution colour displays (Mollon & Reffin 1989, Birch et al.

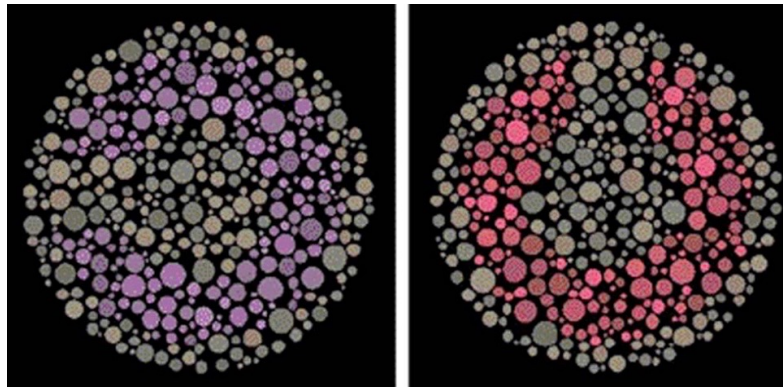


Figure 2.7: The Landolt C stimuli employed in the CCT (adapted from Paramei & Oakley (2014) who in turn adapted the diagram from the original version which appears in the CCT handbook).

1992). Two of the most well-established methods of digital display-based colour vision assessment will be discussed here; the Cambridge Colour Test (CCT) and the Colour Assessment and Diagnosis (CAD) test.

2.3.5.1 Cambridge Colour Test (CCT)

The CCT aims to replicate the techniques used in pseudoisochromatic plate tests on a computer monitor (Mollon & Reffin 1989). The commercial version of the CCT contains two separate tests or protocols, a ‘Trivector’ test and an ‘Ellipses’ test. Both tests employ a four alternative forced choice task in conjunction with a staircase procedure, to estimate observers chromatic discrimination thresholds (Mollon & Reffin 2000). The CCT displays an array of static discs of varying size and luminance (similar to the patterns used in pseudoisochromatic plate tests such as the Ishihara and AO-HRR plates), with the discs that make up the stimuli having a different chromaticity to the background. During the test observers are asked to identify the orientation of the Landolt C stimuli (out of four possible options).

The CCT has been used to quantify changes in chromatic discrimination as a function of age (Paramei 2012), measure the loss of chromatic sensitivity in acquired CVD (Feitosa-Santana et al. 2010, Zachy et al. 2017), and has even been adapted for use with animals (Mancuso et al. 2006).

Although various studies have used the CCT to evaluate chromatic sensitivity or as a standard used to compare with the results of other forms of clinical colour vision assessment, the test has yet to be adopted for use in any occupational environment (Hasrod & Rubin 2015). Whilst the CCT has clear value, and has seen significant use, as a research tool, the test is not currently used in occupational colour vision assessment¹ (Hovis & Almustanyir 2017).

2.3.5.2 Colour Assessment and Diagnosis (CAD) test

The CAD test utilises the luminance masking effect of dynamic luminance contrast (LC) noise to isolate the chromatic contrast generated by a moving colour-defined stimulus.

Dynamic LC noise, sometimes referred to as random luminance modulation (RLM), refers to use of both temporal and spatial random luminance modulation simultaneously. The principle of random spatial luminance modulations for the purpose of masking luminance cues dates back to the work of Stilling (1878). In the early 1980s multiple studies were carried out using background perturbation techniques (both spatial and temporal) for the purpose of investigating the isolation of receptive fields for motion mechanisms (Barbur & Ruddock 1980*a,b*, Barbur et al. 1980, 1981). It would take over a decade for these techniques of temporal and spatial background perturbation to be first used for the purpose of ‘clinical colour vision assessment’ (Birch et al. 1992).

During the test subjects are asked to indicate the direction of the moving coloured stimulus (out of four possible options) using a response box. Subjects are instructed to guess if they are unable to detect the stimulus and the testing normally takes 12-15 minutes to complete. The test offers an enhanced RG or YB CAD test. This requires the subject to carry out three additional RG or YB tests. The enhanced threshold option is only offered when the first threshold is within the expected uncertainty limits of the test with respect to

¹Aside from the CCT’s use in special cases (i.e., when completed privately by an applicant to a given occupational environment in an attempt to demonstrate ‘normal’ or ‘suitable’ chromatic sensitivity [before being submitted to occupational health for consideration]) resulting in the CCT falling into a similar category as electrophysiological measurements with respect to its use in occupational protocols

either the normal upper limit for their age or the allowed occupational CAD limit for a given occupational environment.

The chromatic stimuli employed in the CAD test subtend 0.8° and all the stimuli are embedded in a square array of 15×15 achromatic luminance checks that subtends 3.3° . The luminance of each check scintillates every 40-80 ms (25-12.5 Hz) with equal probability due to the way the information is transferred to the display. The mean luminance of the array remains constant in any given frame (24 cd/m^2) and the strength of the dynamic LC noise is given in terms of the maximum percentage modulation relative to the mean luminance. During the test the dynamic contrast noise is set to a value of 45%. Rodriguez-Carmona (2006) found that this was the level of LC noise required to ensure all dichromats tested were unable to detect the chromatic stimuli presented along their respective confusion axis, even when the maximum signal strength for the display was used. The larger the LC noise employed the larger the LC masking effect and chromatic detection thresholds have been shown to be relatively independent of dynamic LC noise (Barbur et al. 1994, Barbur 2004).

Each CAD stimulus is presented for 720 milliseconds and travels diagonally across a square region defined by the random LC noise (equating to a speed of $\sim 4^\circ$ per second). The direction of motion of each stimulus is randomly allocated and during the test the background is set to a chromaticity of $x=0.305$, $y=0.323$ (as defined in the CIE-1931 (x,y) colour space).

The CAD test employs a four-alternative forced choice (4AFC) task in combination with a staircase procedure to measure chromatic thresholds for 16 different directions in CIE 1931 (x,y) colour space (2 ‘yellow’, 2 ‘blue’, 6 ‘red’ and 6 ‘green’ directions each radiating away from the background chromaticity). During the test the 16 colour directions are randomly interleaved. The staircase procedure requires two correct responses to move ‘down’ (corresponding to a 6.25% or $1/16$) chance of guessing the correct response for each step in the staircase [2 staircase steps per reversal, 11 reversals total]) and one incorrect response to move ‘up’. The threshold for each of the 16 coloured stimuli is measured in terms of the chromatic displacement in CIE 1931 (x,y)

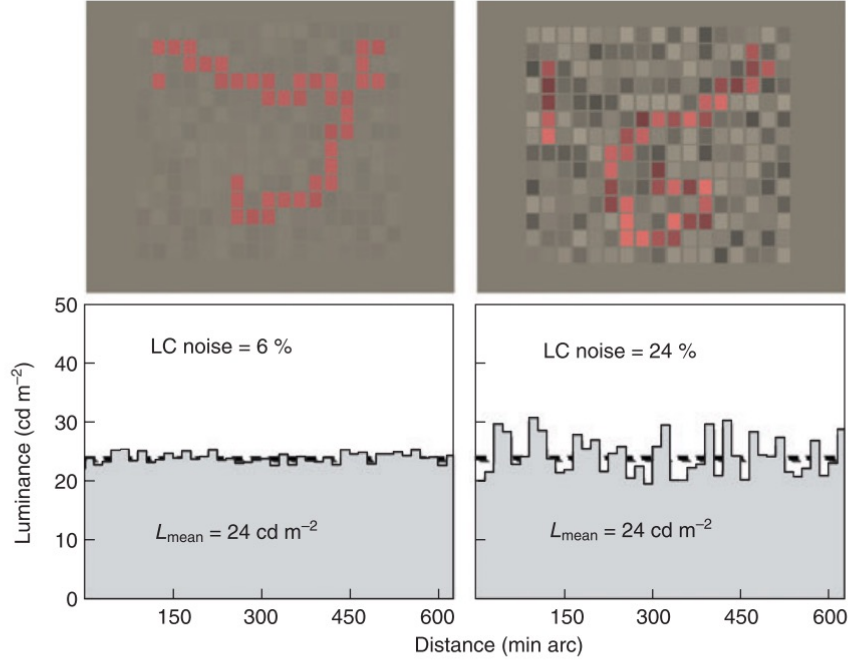


Figure 2.8: Colour-defined stimuli on two backgrounds of dynamic LC noise adapted from Barbur (2004). The difference between the strength of the dynamic LC noise can be observed between the left-hand images (at 6% modulation) and the right hand images (at 24% modulation). The mean luminance for both backgrounds represented by the black dotted line and is 24 cd/m^2 for both stimuli.

colour space between the stimuli and the background chromaticity (i.e., the difference in x, y coordinates in CIE 1931 (x, y) colour space). A move ‘down’ in the staircase corresponds to a decrease in the chromatic displacement and a move ‘up’ in the staircase corresponds to an increase in the chromatic displacement. The incremental changes for each step in the staircase procedure follow an exponential function. The first step corresponds to ~ 6 CAD units above the mean threshold for the standard CAD observer and the last reversal increment is ~ 0.25 CAD units. The threshold for each direction is calculated by taking the mean of the last 6 staircase reversals corresponding to a $\sim 71\%$ probability that subjects are able to identify the stimuli at their threshold (Barbur & Rodriguez-Carmona 2017).

The CIE 1931 (x, y) colour space is non-linear. Yet, when colour thresholds are measured as chromatic displacements relative to a background chromaticity (as in the CAD test), the cone contrasts generated are approximately linearly proportional to the thresholds measured (Rodriguez-Carmona et al. 2012).

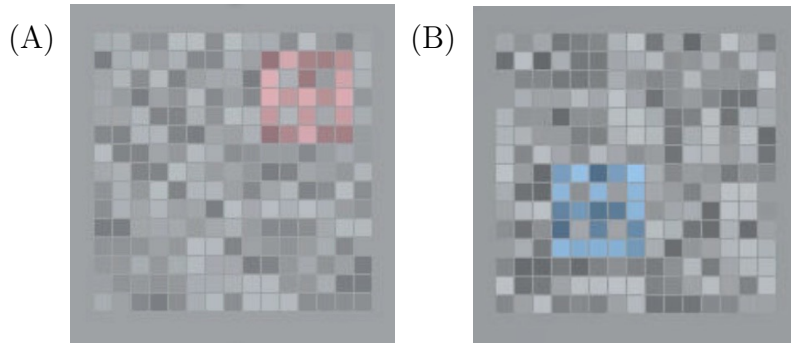


Figure 2.9: The stimulus employed in the CAD test shown on a background of achromatic luminance checks with (A) 45% and (B) 75% RLM.

Jennings & Barbur (2010) demonstrated that measured CAD thresholds can be accurately predicted based upon the cone excitation signals generated by an adapting background field. This was found to be the case regardless of the initial state of chromatic adaptation and for a wide range of luminances.

The 16 thresholds are separated into two groups, four thresholds for ‘yellow’ and ‘blue’ colours and 12 thresholds for ‘red’ and ‘green’ colours. The groups are averaged to produce the two numerical thresholds displayed in the test interface, a red-green (RG) threshold and a yellow-blue (YB) threshold. The mean YB threshold is obtained by averaging the four YB thresholds. The mean RG threshold is based on the average of the largest six thresholds measured with RG colours. Both of these thresholds are given in terms of standard normal CAD units. Standard normal CAD units were created following work carried out by Rodriguez-Carmona (2006) in which the CAD thresholds of over 330 young normal trichromats ($\mu=28$ years, $\sigma=8$ years) were recorded (Barbur & Rodriguez-Carmona 2015, 2017). These data were used to define the standard normal CAD observer, where the mean RG and YB thresholds for this sample define one standard normal CAD unit. The upper and lower age specific limits for the standard CAD observer were determined using $\pm 2.5\sigma$ limits for RG and YB CAD thresholds (Rodriguez-Carmona et al. 2005, Barbur & Rodriguez-Carmona 2017).

Following a number of studies carried out at City, University London, the CAD test was validated and is currently accepted as a form of occupational colour vision assessment in several occupational environments including com-

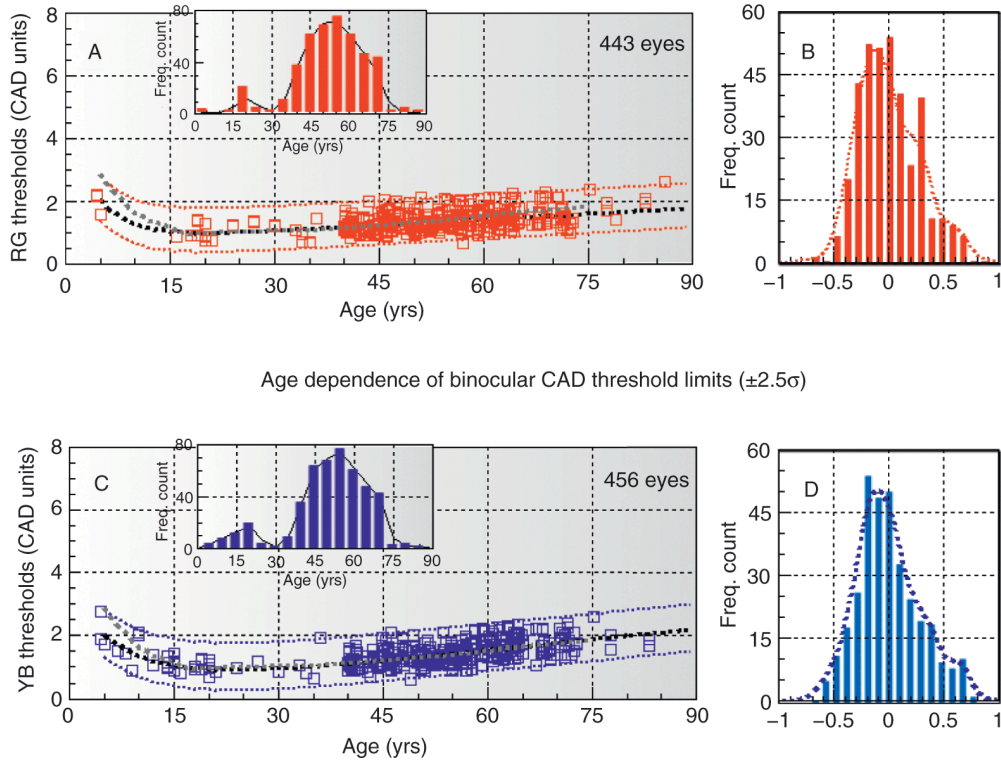


Figure 2.10: The relationship between RG and YB CAD thresholds as a function of age adapted from Barbur & Rodriguez-Carmona (2015). The normal threshold initially decreases until the age of ~ 20 years before slowly increasing for both RG and YB ($\sim 1\%$ increase/year for RG and $\sim 1.6\%$ increase/year for YB). The upper and lower normal limits are indicated on diagrams A and C by the upper and lower dotted line [A/C - relationship between RG/YB CAD thresholds and age along with a histogram showing the age range of subjects in the study, B/D the distribution of thresholds for all ages compared to the predicted mean value].

mercial and private aviation (both for pilots and air traffic controllers), the electrical contractors industry, Maritime and Coastguard agency, Transport for London (TfL), and the UK armed forces (Barbur et al. 2009, 2016, Marechal et al. 2018, Maritime and Coastguard Agency 2020). An interesting and unique property of the CAD test limits for aviation and TfL is the use of separate standards for protans and deuterans. This is a direct consequence of the experimental procedure used to establish both standards.

A widely reported inaccuracy surrounding the CAD test is worth briefly addressing. The origin of this misconception can be attributed to a paper by Seshadri et al. (2005) in which a “*a simplified web-based version of the original CAD test*” (Seshadri et al. 2005, page 882) was evaluated. The ‘web-based version of the CAD test’ reported by Seshadri et al. (2005) is a ~ 90 second

video showing the stimuli employed in the CAD test. Unfortunately the fact that the web-video is simply representative of the stimuli used in the full CAD test is not reported in a number of publications which quote Seshadri et al. (2005), along with inaccurate statements surrounding the CAD tests’ ability to detect YB loss (French et al. 2008, Fanlo Zarazaga et al. 2019).

2.4 ‘Full’ colour vision assessment and the ‘Gold standard’

Methods for assessing colour vision are often designed for a specific purpose, or to achieve a certain aim. Logically, when evaluating a test one should always do so relative to its aim as there is limited value in evaluating the performance of a test in an arena beyond it’s intended use.

It is, however, useful to consider the properties of an ‘ideal’ or ‘perfect’ colour vision test. What such a test would look like, the steps required for ‘full’ colour vision assessment, and the value and brittleness of colour vision ‘Gold Standards’ are discussed here.

2.4.1 The challenges of ‘full’ colour assessment

‘Full’ colour vision assessment starts with the full isolation of colour signals. Although perhaps obvious, this is a critical requirement as one needs to ensure that subjects can only make use of colour cues. Following the full isolation of colour signals the ‘full’ assessment would allow for the selective stimulation of RG and YB chromatic mechanisms. Such a step places limitations upon the first stage - for example, the Ishihara pseudoisochromatic plates employ YB noise to remove any YB cues from the numerical stimuli. This in turn results in the test being unable selectively to stimulate YB chromatic mechanisms (and screen for YB loss).

‘Full’ colour vision assessment would make use of the selective RG and YB stimulation to classify perfectly the class of any applicant’s colour vision.

Given the range of physiological properties present in the ‘normal’ population (including variations in the L:M cone ratio, peak responsitivity of cone photoreceptors, pre-receptor filtering, and macular pigment density (Rodriguez-Carmona & Barbur 2017)) the ideal test would need to establish a ‘normal’ range, to account for the variation observed in the ‘normal’ population.

Finally, ‘full’ colour vision assessment would quantify the severity of any RG or YB loss. Doing so, in combination with the ‘normal’ range established to classify perfectly the type of colour vision, would allow the ideal test to diagnose even cases where acquired loss had developed in an individual with a pre-existing congenital deficiency.

Unfortunately (or perhaps fortunately given that living in a perfect world would be somewhat colourless!) no test is perfect. Tests and protocols are often accepted as a compromise to fulfil practical needs, despite their flaws. This is investigated and quantified further in chapter 3, but it is important to avoid false equivalencies and note and appreciate that there are varying degrees of imperfection.

2.4.2 The ‘Gold standard’ of colour vision assessment

The term ‘Gold standard’ has different meanings with primary definitions including:

1. *“a monetary standard under which the basic unit of currency is defined by a stated quantity of gold and which is usually characterized by the coinage and circulation of gold, unrestricted convertibility of other money into gold, and the free export and import of gold for settling of international obligations”*
2. *“something that serves as a standard by which others may be measured or judged”*

(Merriam-Webster n.d.)

Here, and in the context of scientific and medical academia, particularly when dealing with statistics, the second meaning is taken (Altman 1991). When

considering a ‘Gold standard’ it is important to appreciate that its primary statistical value is in determining a frame of reference. A frame of reference is a valuable, if not essential, framework in a number of settings.

Being a ‘Gold standard’ is not an innate test property - rather an indirect measure of a test’s potential value in establishing a frame of reference. This becomes apparent when considering the evolution of colour vision assessment, and the subsequent historical ‘Gold standards’ of colour vision assessment. Dalton’s coloured ribbon test would have been the ‘Gold standard’ for colour vision assessment at the time of its introduction as it would have been the best or most accurate measure available, potentially followed by Stilling’s plate test and/or by the first lantern tests. These tests have not changed, but are no longer considered to be the ‘Gold standard’ (for good reason!) and the reality that ‘Gold standards’ can, and often should, change over time is a key point.

Within the context of colour vision assessment several current tests have been referred to as a ‘Gold standard’. The most common is the anomaloscope which is considered to be the ‘Gold standard’ for identifying and diagnosing types of RG CVD (Waalder 1927, Birch 2001, 2008a, Fanlo Zarazaga et al. 2019, Arnegard et al. 2022). The test is also sometimes referred to in a boarder sense as being the ‘Gold standard’ for the detection of CVD or colour vision diagnostics (Jurasevska et al. 2014, Karampatakis et al. 2022). The Ishihara pseudoisochromatic plate test has been referred to as the ‘Gold standard’ in screening for congenital RG CVD (Dain 2004, French et al. 2008). The CAD test has also been used as a gold-standard reference against which the performance of other methods of colour vision assessment have been evaluated (Koefoed et al. 2020).

The descriptions for these methods of colour vision assessment hint at the second, and perhaps more important, item to note regarding the ‘Gold standard’ of colour vision assessment (in addition to the fact that ‘Gold standards’ can and should change over time) - a singular ‘Gold standard’ for colour vision assessment does not exist.

A universal ‘Gold standard’ for colour vision assessment does not exist for the same reason there is no singular ‘Gold standard’ for visual assessment. Human colour vision is too complex and multidimensional for a single test to be able to, at least currently, accurately measure, quantify, and determine all aspects of colour vision (i.e., act as a sensitive screener, measure an individual’s metamerism, classify an individual’s colour vision, measure the severity of RG and YB loss, etc.).

Whilst a singular ‘Gold standard’ of colour vision assessment does not exist, there are ‘Gold standards’ for the various aspects of colour vision including the classification of CVD and measuring the severity of loss. This is important to emphasise, as judging a screening test relative to the ‘Gold standard’ for classifying types of CVD would result in a frame of reference that is obscurative at best and intentionally misleading at worst.

2.5 Summary

The past 350 years have seen a vast number of novel methods designed and introduced to measure and assess human colour vision. Despite the wide range of tests available, individuals and occupations often opt to default to a standard set of tests. This is often done for good reason, due to either the time, money or energy required to acquire and use newer methods of colour vision assessment, and/or due to the difficulty in interpreting the results of the more complex forms of colour vision assessment.

With so many available tests the evaluation of tests and protocols in current use is of considerable interest. Such an approach would crucially also enable one to examine how close current colour vision assessment protocols are to meeting current needs.

Chapter 3

The current state of conventional colour vision assessment

Chapter 2 introduced many of the creative methods that have been designed and developed for the purpose of human colour vision assessment. This chapter aims to investigate and quantify the current state of several single- and multi-test colour assessment protocols. Completed studies, including an investigation into the outcome and limitations of single- and multi-test protocols using data collected at City, University of London and an exploratory investigation using the Farnsworth D-15 are described and discussed.

Much of the work described in this chapter has been published in the following papers:

- **Evans, B. E. W.**, Rodriguez-Carmona, M. & Barbur, J. L. (2020), ‘Color vision assessment-1: Visual signals that affect the results of the Farnsworth D-15 test’, *Color Research & Application* **46**(1), 7–20.
- Rodriguez-Carmona, M., **Evans, B. E. W.** & Barbur, J. L. (2020), ‘Color vision assessment-2: Color assessment outcomes using single and multi-test protocols’, *Color Research & Application* **46**(1), 21–32.

3.1 Introduction

Modern colour vision assessment plays a vital role in occupational and clinical settings. The clinical applications are fascinating with implications for the early detection of diseases of the retina and optic nerve, including diabetes (O'Neill-Biba et al. 2010, Simunovic 2016). The occupational impact of colour vision assessment is arguably even more significant with a plethora of occupational roles involving the safety critical use of colour and therefore requiring applicants to meet specific colour vision standards. Whilst the template for occupational colour assessment has remained remarkably consistent, a screening test followed by more detailed secondary testing, implementations of this template vary widely (Birch 2001).

The development and implementation of new colour assessment protocols continues to this day - a notable recent example is the UK Police protocol for authorised firearms officers (Stockman 2018). Current occupational colour vision assessment typically makes use of multiple tests to attempt to detect if potential applicants have the level of chromatic sensitivity required to safely carry out their desired role. The continued use, development, and introduction of single- and multi-test colour assessment protocols not only demonstrates that the assessment of colour vision is as important as it ever was (potentially even more so with the increased use of colour in modern life), but it also raises several key questions:

- I. How close are current protocols to reaching the limit of ‘full colour vision assessment’?
- II. What factors contribute towards the fundamental limit of a test protocol?
- III. Do multi-test protocols, created by combining multiple single-test protocols, exceed the limitations of their individual components?

These questions are essential to address before potentially attempting to describe new methods for assessing human colour vision and/or screening for CVD (chapter 4).

The primary aims of the experimental work described in this chapter were to examine how close current single- and multi-test colour assessment protocols are to achieving their intended aims and meeting current needs, and to establish the primary factors that either enable or inhibit protocols from accomplishing their intended outcome.

3.2 Single- and multi-test protocols

3.2.1 Introduction

The conventional methods of colour vision assessment described in chapter 2 are rarely ever used in their ‘pure’ or ‘raw’ form. Instead, the tests are combined and employed using different sets of rules or protocols¹. The test outcome is interpreted through the lens of a protocol to give clinical significance to the outcome of the test and/or to achieve a specific outcome. These protocols are often so deeply intertwined with the test that in most clinical and all occupational settings the test and protocol are entirely co-dependent and inseparable.

The simplest protocols, single-test protocols, have already been introduced in section 2.3. As the simplest form of colour assessment protocols, single-test protocols are significantly less resource intensive, and consequently, more popular than multi-test protocols – particularly in clinical settings. Single-test protocols form the building blocks of more complex multi-test protocols. While single-test protocols are more common in optometric practice, in occupational settings multi-test protocols, protocols involving the sequential application of two or more single-test protocols, are more prevalent.

Evaluating the outcome of multi-test protocols can be challenging as the variation in an individual’s results across different methods of colour vision assessment cannot be reconstructed or reliably predicted from the results of a single test or protocol. As such, any study investigating multi-test protocols ideally will apply a multi-test protocol to each participant using data from multi-

¹In this context a protocol refers to a specific set of rules typically sequential in nature

ple tests for each participant, rather than attempting to combine single-test protocol statistics from separate samples.

The potential pitfalls of attempting to combine data from separate statistics for single-test protocols occur due to the different subsamples, or subpopulations, that complete each stage in the protocol. This is best explained using the following thought experiment:

Assume that you have the statistics for two single-test protocols, A and B, both taken over a large sample size that is perfectly representative of the overall population. The two single-test protocols are combined into a multi-test protocol, C, by applying A and B in the sequence shown in Figure 3.1. The percentage of the population that will complete Protocol B is known and described by the statistics for Protocol A (simply given by the percentage of the population who fail A). However, the population that completes B is no longer representative of the overall population. Subsequently the outcome of Protocol C cannot be predicted from the known statistics for Protocol A and Protocol B if the statistics for Protocol B are only known for the general population.

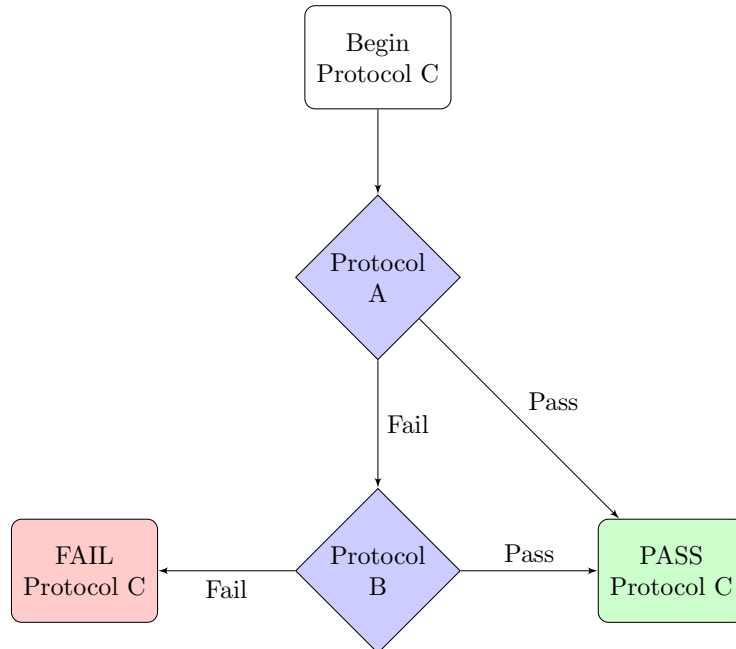


Figure 3.1: A theoretical multi-test protocol, ‘Protocol C’, consisting of two single-test protocols, ‘Protocol A’ and ‘Protocol B’.

As a rule, the statistics for a multi-test protocol can only be accurately estimated if the single-test statistics are known for the subsection of the population filtered by all prior steps in the protocol. Interestingly the agreement between the outcome of the multi-test protocol and the theoretical prediction made from a combination of the statistics for the single-test protocols, is directly proportional to the percentage of individuals who pass the initial test(s) and the number of tests employed. The lower the number of single-test protocols and the pass rate of the initial single-test protocols the more reliable the agreement between the outcome of the multi-test protocol and the predicted outcome from the combination of single-test protocols. This limitation only applies to the overall pass/fail statistics for the protocol. If the maximum level of severity of loss is known for each individual component, one can estimate the maximal level of loss that the multi-test protocol will, at least theoretically, allow and pass.

Single- and multi-test protocols that make use of three of the most commonly used conventional methods of colour vision assessment, the Ishihara pseudisochromatic plates, the City University test, and the Farnsworth D-15 are described here. As mentioned in chapter 2, when evaluating the outcome of any given method of colour vision assessment, or colour assessment protocol, it's essential to first understand its aim. The aim provides both a framework to evaluate the performance of the test or protocol and context as to the settings in which the tests and/or protocols will be, or are, employed.

Single-test Ishihara plate protocols aim to efficiently screen for RG colour vision deficiency with high sensitivity and specificity, detecting individuals with even the mildest congenital RG CVD (Birch 2001). Protocols that make use of the Ishihara plates are frequently employed as the first 'screening' step in multi-test protocols (Margrain et al. 1996, Birch & Chisholm 2008, Chorley 2015) and, as described in section 2.3.1.1, as the 38-plate edition of the Ishihara test contains the same plates as used in the 14 and 24 plate edition the results from the 38-plate edition can be used to evaluate the outcome of any version of test.

By contrast, protocols that employ the City University test and the Farnsworth D-15 test do so to identify and eliminate those with ‘moderate’ and ‘severe’ CVD (Birch 1997*a*, Dain et al. 2019). It should be noted that although the specific level of CVD associated with the terms ‘mild’, ‘moderate’, and ‘severe’ is rarely specified², the use of such terminology implicitly implies that a protocol is able to discriminate and differentiate between individuals on the basis of their level of chromatic sensitivity.

In addition to being used in protocols designed to split the population into pass/fail categories, the Ishihara pseudoisochromatic plates, City University test, and Farnsworth D-15 can, at least in theory, also be used to classify the type of CVD. This opens up the possibility to evaluate the outcome of single-test classification protocols, i.e., how well protocols are able to classify the type of CVD in addition to detecting the presence of CVD.

Two current multi-test protocols used in the UK are a firefighter protocol and a police protocol for authorised firearms officers, introduced in 1996 and 2018 respectively (Margrain et al. 1996, Stockman 2018). The police protocol includes the use of the Ishihara pseudoisochromatic plates, the City University test, and the Farnsworth D-15 and aims to pass all normal trichromats and exclude applicants with a protan deficiency and/or severe RG CVD. The colour assessment protocol for UK firefighters uses the Ishihara plates, the Farnsworth D-15, and testing using an anomaloscope. As with the police protocol, the firefighter protocol aims to pass all normal trichromats and individuals with mild deutan CVD whilst excluding applicants with a protan deficiency.

This study aimed to establish reliable statistics to investigate the outcome of single and multi-test colour assessment protocols that make use of the Ishihara, Farnsworth D-15, and City University tests.

²With one or two notable exceptions (Barbur & Rodriguez-Carmona 2017, Ministry of Defence 2018)

3.2.2 Methods

Data for 850 subjects were abstracted from anonymised records collected through the Advanced Vision and Optometric Tests (AVOT) clinic service at City, University of London. All subjects completed the CAD test, type I Nagel anomaloscope, Ishihara pseudoisochromatic plate test (38 plate edition), City University test (2nd Edition), and the Farnsworth D-15. The exclusion criteria were acquired colour vision loss, subjects under 10 years of age, and subjects who returned to the AVOT clinic for multiple assessments. For subjects who returned for repeat visits, only the data from their first visit were used. Following the application of the exclusion criteria, data for 427 subjects were analysed including 57 normal trichromats, 261 deutans, and 109 protans.

Standardised instructions were given prior to all tests. The Ishihara test, Farnsworth D-15 caps, and City University test were all illuminated with the new Macbeth easel lamp approximating CIE illuminant D65. The CAD test was displayed on a 10-bit Eizo CS2420 monitor calibrated for luminance and the chromaticity of each primary colour using a Konica CS-2000A telespectroradiometer (manufactured by Konica Minolta Inc.) with a spectral range of 380 to 780 nm. The spectrum was sampled every nm with a bandwidth of just under 1 nm. The CAD test was run on an HP ProBook 650 G1 laptop, and all participants were required to successfully completed a short “learning mode” prior to the CAD test, to ensure they understood the task. The type I Nagel anomaloscope was used to measure the Rayleigh colour match for all participants and the type of colour vision deficiency was determined from the results of the CAD test and the Nagel anomaloscope.

Single-test Ishihara protocols including accepting 0 errors on the first 25 plates of the 38-plate edition, ≤ 4 errors on plates 1-21 of the 38-plate edition, ≤ 2 errors on plates 1-17 of the 24-plate edition, ≤ 6 errors on plates 1-17 of the 24-plate edition, and the classification outcome based upon the last four plates were investigated. The classification for each of the four plates was recorded as ‘deutan’ if the left-hand number was identified and ‘protan’ if the

right-hand number was identified. The overall classification was ‘normal’ if no errors were made, ‘indeterminate’ if the same number of protan and deutan errors were made, and either ‘protan’ or ‘deutan’ based upon whether more ‘protan’ or ‘deutan’ errors were recorded.

The outcome of the second edition of the City University test was investigated using single-test protocols including accepting 0 errors on all plates, allowing ≤ 4 errors overall & ≤ 1 error on the first 6 ‘large’ plates, and the classification outcome based upon all 10 test plates. Similar to the Ishihara classification protocol, the classification was determined by the largest number of errors made in a given category (‘protan’, ‘deutan’, or ‘tritan’). If the same number of errors were made in two or more of the categories, the classification was ‘indeterminate’.

The outcome of the Farnsworth D-15 was investigated when no errors (either major crossings or adjacent transpositions on the circular results diagram [Figure 2.4]) are accepted, ≤ 2 adjacent transpositions and 0 crossings are accepted, and ≤ 1 major crossing and ≤ 2 adjacent transpositions are accepted. The classification outcome of the Farnsworth D-15 was determined by splitting the cap order into 15 pairs of sequential caps and passing each pair through the classification grid shown in Figure 3.2 to systematically determine the classification that would be made from a visual inspection of the circular results diagram.

Two multi-test colour assessment protocols, the current UK police protocol for authorised firearms officers (AFOs) and specially trained units (STUs) equipped with Tasers (introduced in 2018) and the current UK fire service protocol (introduced in 1996) were investigated.

The police protocol screens applicants with Ishihara plates, allowing up to 4 errors on the first 21 plates of the 38-plate edition. Applicants who fail the Ishihara test must pass both the City University test (≤ 4 errors overall & ≤ 1 error on the first 6 ‘large’ plates) and the Farnsworth D-15 (≤ 1 major crossing accepted). Additionally, applicants fail the protocol if they are diagnosed as having a protan deficiency on 2 or more of the tests carried out. The fire service protocol screens applicants with the Ishihara plates, allowing

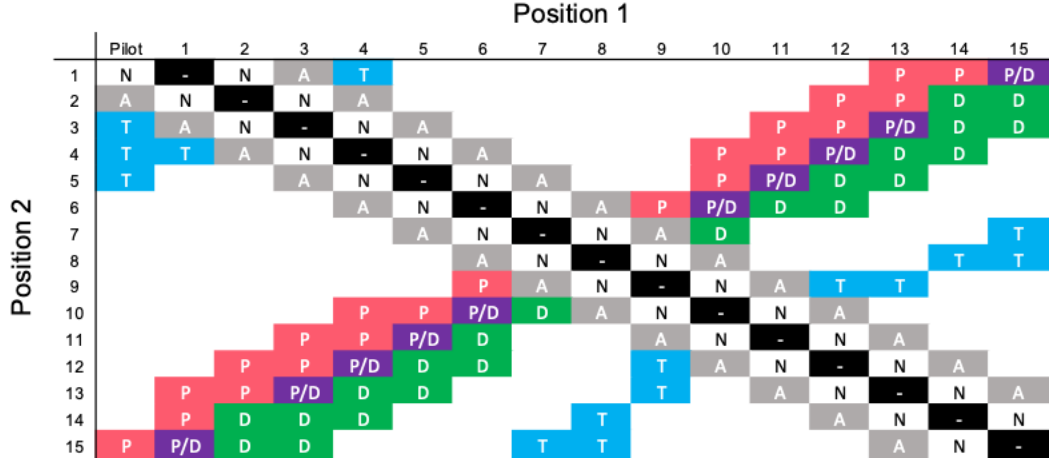


Figure 3.2: Farnsworth D-15 classification grid used to systematically apply a ‘visual inspection’ analysis to Farnsworth D-15 cap orders (adapted from Evans et al. (2020)). A participant’s Farnsworth D-15 result yields 15 sets of sequential pairs and each pair can then be linked through the classification grid to one of six types of matches (N - ‘normal’, A - ‘adjacent transposition’, T - ‘tritan’, P - ‘protan’, D - ‘deutan’, P/D - ‘protan and deutan’ [counts as both a ‘protan’ and ‘deutan’ error for the sum calculation]). The overall Farnsworth D-15 classification is determined by the largest sum of errors (e.g., if an individual makes 4 ‘protan’ crossings and 2 ‘deutan’ crossings, the classification would be “protan”). If there are the same number of errors for two types of error the classification is “indeterminate”.

up to 2 errors on the first 17 plates of the 24-plate edition. Applications who fail the Ishihara test must pass the Farnsworth D-15 (≤ 2 adjacent transpositions accepted) and be diagnosed as having normal trichromacy or deuteranomaly when tested with an anomaloscope. The statistically equivalent RG CAD thresholds (the value for which the number of participants who fail with thresholds lower than this value is equal to the number of participants who pass with thresholds higher than this value) were calculated for deutan and protans for the two multi-test protocols. A mathematical description of the statistically equivalent CAD threshold, and the algorithm used to determine this value, are described in Appendix A.

The relationship between the severity of loss of chromatic sensitivity and the outcome of single- and multi-test protocols were evaluated using RG CAD thresholds as a measure of the severity of loss. Violin plots were used to examine the distribution of RG CAD thresholds in the study (Bechtold 2016). Violin plots have been described previously and they combine a boxplot, a

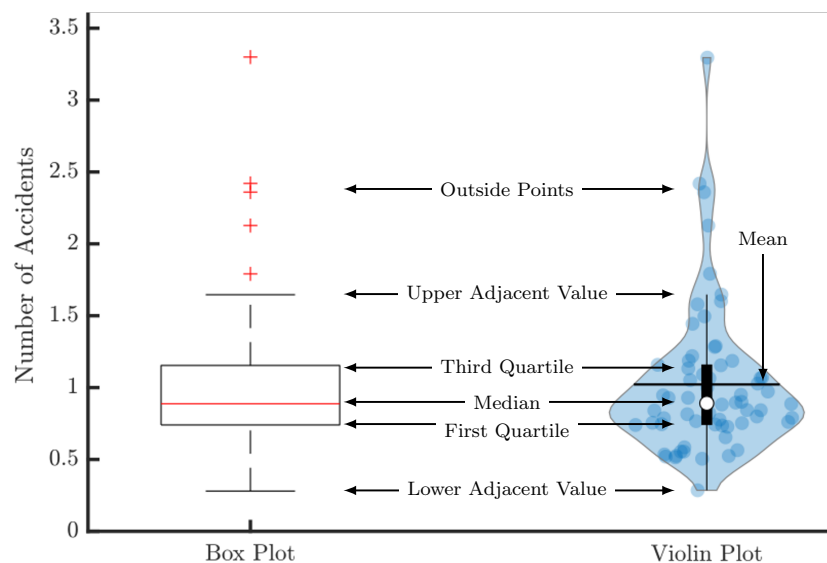


Figure 3.3: A comparison between common components in box plots and violin plots recreated from a figure published by Hintze & Nelson (1998).

density trace, and the mean into a single graphic (Hintze & Nelson 1998). A graphical comparison between boxplots and violin plots is shown in Figure 3.3.

The Nagel anomaloscope and the CAD test were used to assess the classification outcome. The study was conducted in compliance with the City, University of London research and ethical guidelines and followed the tenets of the Declaration of Helsinki.

3.2.3 Results

The age of the subjects in the sample ranged from 10 to 64 years with a mean of 32, standard deviation of 12, and a median of 30 years of age. The RG CAD thresholds ranged from 0.67 to 2.20 for normal trichromats, 1.81 to 26.61 for subjects with deuteranomaly, 7.75 to 28.47 for subjects with deuteranopia, 4.62 to 22.92 for subjects with protanomaly, and 11.62 to 26.67 for subjects with protanopia, as diagnosed by the Nagel anomaloscope. The distribution of RG CAD thresholds as a function of the type of colour vision are shown in Figure 3.4.

As shown in Table 3.1, when no errors are accepted on the first 25 plates of the 38-plate edition of the test less than 1% of protans and deutans pass, and

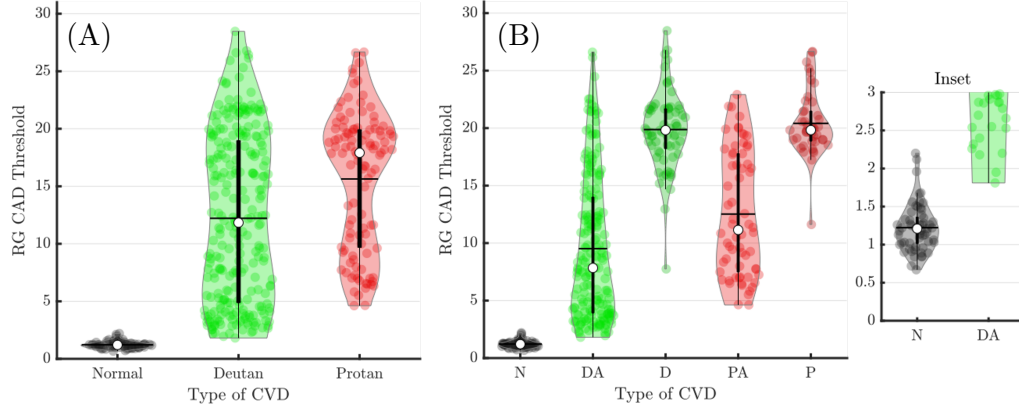


Figure 3.4: Distribution of RG CAD thresholds for the samples used in the single- and multi-test protocol study shown in violin plots where the data is separated into categories of (A) Normal Trichromats, Deutan deficiencies, and Protan deficiencies and (B) the diagnosis made using the Nagel anomaloscope (N – normal trichromacy, DA – deuteranomaly, D – deuteranopia, PA – protanomaly, P – protanopia). The violin plots contain a standard boxplot in the centre of each violin plot, the white circle is the median, the horizontal line is the mean, and the width of the violin plot shows the relative frequency. The right-hand inset contains the same data, with a reduced ordinate to enable to variation in RG CAD thresholds amongst the normal population to be examined. The overlap between the least affected deuteranomalous subjects and the least sensitive normal trichromats is due to the range of ages in the subject population and the normal age-adjusted limits employed in the CAD test.

17.5% of normal trichromats fail. When four errors are accepted on plates 1-21 of the same edition 100% of normal trichromats pass, but approximately 6.1% of deutans and 1.8% of protans also pass. The percentage of protans and deutans who pass when up to 2 errors are accepted on the 17 numerical plates in the 24-plate edition remains similar ($\sim 5.7\%$ and $\sim 1.8\%$ respectively). Allowing up to six errors on the 24-plate edition will pass 19.2% of deutans and 8.3% of protans. Some of the protan and deutan subjects who pass can have reduced RG chromatic sensitivity with thresholds up to 9.4, 13.3, and 21.6 RG CAD units, depending upon the protocol.

There was 100% agreement between the classification made by the CAD test and the Nagel anomaloscope for all subjects in the sample. The classification outcome for the Ishihara test when at least one error is made on the first 25 plates in the 38-plate edition is shown in Table 3.2. Fifty per cent of deutan

subjects and approximately 36.1% of protan subjects who make at least one error on the Ishihara test are correctly classified by the classification plates.

A) Percentage who pass (%)

Classification		Ishihara pass criteria			
		38 plate edition		24 plate edition	
		Plates 1-25 ≤ 0 errors	Plates 1-21 ≤ 4 errors	Plates 1-17 ≤ 2 errors	Plates 1-17 ≤ 6 errors
Normal	(N=57)	82.5	100.0	100.0	100.0
Deutan	(N=261)	0.4	6.1	5.7	19.2
Protan	(N=109)	0.9	1.8	1.8	8.3

B) Max RG CAD threshold in those who pass (RG CAD units)

Classification		Ishihara pass criteria			
		38 plate edition		24 plate edition	
		Plates 1-25 ≤ 0 errors	Plates 1-21 ≤ 4 errors	Plates 1-17 ≤ 2 errors	Plates 1-17 ≤ 6 errors
Normal	(N=57)	2.1	2.2	2.2	2.2
Deutan	(N=261)	3.9	13.0	13.0	21.6
Protan	(N=109)	9.4	13.3	13.3	17.9

Table 3.1: A) The outcome of four commonly used Ishihara single-test protocols for normal trichromats, deuterans, and protans. B) The maximum CAD threshold obtained by subjects who pass each of the protocols used in this study. The current UK police protocol allows up to four errors on plates 1-21 on the 38-plate edition, the UK fire service allows up to 2 errors on plates 1-17 on the 24-plate edition, and the EASA cabin crew colour vision standard allows up to 6 errors on plates 1-17 on the 24-plate edition of the test.

Classification		Ishihara classification for subjects who make ≥ 1 error (%)		
		Deutan	Protan	Indeterminate
Normal	(N=57)	0.0	0.0	100.0
Deutan	(N=261)	50.0	0.4	49.6
Protan	(N=109)	11.1	36.1	52.8
Deuteranomaly	(N=193)	39.1	0.5	60.4
Deuteranopia	(N=68)	80.9	0.0	19.1
Protanomaly	(N=66)	13.8	18.5	67.7
Protanopia	(N=43)	7.0	62.8	30.2

Table 3.2: The classification outcome of the Ishihara test based upon the digit identified on plates 21-25 (38 plate edition) for all subjects who make at least one error on plates 1-25 (38 plate edition). The DA/D/PA/P classification was determined by the Nagel anomaloscope.

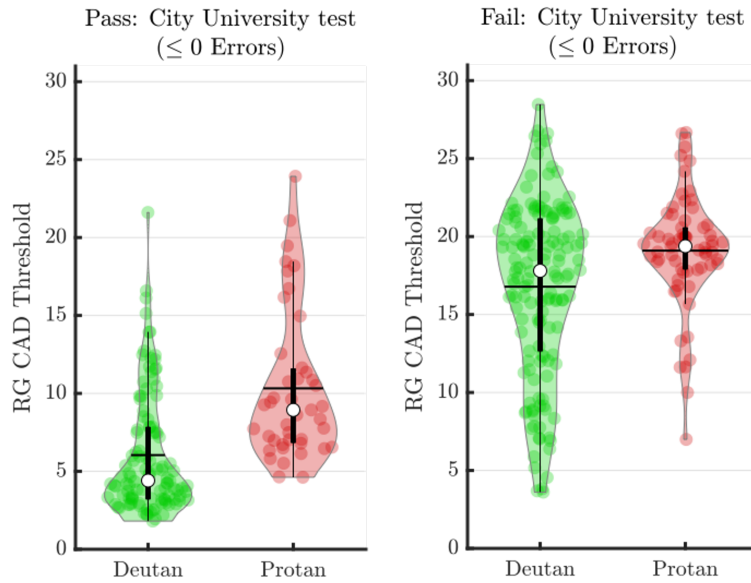


Figure 3.5: Distribution of RG CAD thresholds in protan and deutan subjects who pass and fail the City University test when no errors are accepted. The horizontal line in each violin plot indicates the mean RG CAD threshold.

Data for single test protocols using the City University test are shown in Table 3.3. All normal trichromats in the sample passed without making any errors and approximately 42.5% of deutan and 39.4% of protans did not make any errors on the City University test. When up to four total errors are accepted, with up to one of the errors being on the first 6 plates, 60.5% of deutan and 56.9% of protans passed the City University test. The maximum RG CAD

Classification		City University test pass criteria			
		≤ 0 errors		≤ 4 errors overall & ≤ 1 error on the first 6 plates	
		Outcome (%)	Max RG	Outcome (%)	Max RG
Normal	(N=57)	100.0	2.2	100.0	2.2
Deutan	(N=261)	42.5	21.6	60.5	21.8
Protan	(N=109)	39.4	23.9	56.9	23.9

Table 3.3: The outcome of two commonly used City University test (2nd Edition) single-test protocols for normal trichromats, deutan, and protans.

thresholds in deutan and protan subjects were 21.6 and 23.9 RG CAD units respectively when no errors are accepted. The maximum RG CAD thresholds were 21.8 and 23.9 RG CAD units when no more than 4 overall errors with no more than 1 error on the first six plates are accepted on the City University test. The distribution of RG CAD thresholds in deutan and protans who pass and fail the City University test when no errors are accepted is shown in Figure 3.5.

The classification outcome of the City University test, for subjects who made at least one error, is shown in Table 3.4. Approximately 92.7% of deutan subjects and 50.0% of protan subjects who made a mistake on the City University test are correctly classified. Specifically, 89.0% of subjects with deuteranomalous trichromatism, 97.1% of subjects with deuteranopia, 24% of subjects with protanomalous trichromatism, and 65.9% of subjects with protanopia are classified correctly. Thirty-four point eight per cent of protans are incorrectly classified by the City University test as having a deutan deficiency - 60.0% of subjects with protanomalous trichromacy and 19.5% of subjects with protanopia.

The outcome of the Farnsworth D-15 is shown in Table 3.5. Ninety-one point two per cent of normal trichromats, 49.8% of deutan, and 39.4% of protans do not make any major crossings or adjacent transpositions when their Farnsworth D-15 results are plotted on a circular results diagram. All normal trichromats and 54.8% and 44.0% of deutan and protans, respectively, make up to 2 adjacent transpositions. When one major crossing is

Classification		City University test classification for subjects who make ≥ 1 error (%)			
		Deutan	Protan	Tritan	Indeterminate
Deutan	(N=261)	92.7	2.0	3.3	2.0
Protan	(N=109)	34.8	50.0	1.5	13.6
Deuteranomaly	(N=193)	89.0	2.4	4.9	3.7
Deuteranopia	(N=68)	97.1	1.5	1.5	0.0
Protanomaly	(N=66)	60.0	24.0	4.0	12.0
Protanopia	(N=43)	19.5	65.9	0.0	14.6

Table 3.4: The classification outcome of the City University test (2nd Edition) for subjects who make at least one error. The DA/D/PA/P classification was determined by the Nagel anomaloscope.

A) Percentage who pass (%)					
Classification		Farnsworth D-15 pass criteria			
		Major crossings:		0	≤ 1
		Adjacent transpositions:		0	≤ 2
Normal	(N=57)			91.2	100.0
Deutan	(N=261)			49.8	58.2
Protan	(N=109)			39.4	49.5

B) Max RG CAD threshold in those who pass (RG CAD units)					
Normal	(N=57)			2.2	2.2
Deutan	(N=261)			17.7	21.8
Protan	(N=109)			23.9	23.9

Table 3.5: The outcome and maximum RG CAD threshold (‘Max RG’) obtained by normal trichromats, deutan, and protans using three commonly used Farnsworth D-15 single-test protocols.

accepted all normal trichromats, 58.2% of deutan and 49.5% of protans pass the Farnsworth D-15. The distribution of RG CAD thresholds in deutan and protans who pass and fail the Farnsworth D-15 test when up to two adjacent transpositions are accepted is shown in Figure 3.6.

When errors are made on the Farnsworth D-15 approximately 85.5% of deutan and 72.7% of protans are classified correctly. Approximately 9.1% of protans are incorrectly classified as being ‘deutan’ on the Farnsworth D-15 test when the results are plotted and determined by ‘visual inspection’.

Classification		Farnsworth D-15 classification for subjects who make ≥ 1 error (%)			
		Deutan	Protan	Tritan	Indeterminate
Deutan	(N=261)	85.5	0.0	0.8	13.7
Protan	(N=109)	9.1	72.7	0.0	18.2
Deuteranomaly	(N=193)	74.6	0.0	1.6	23.8
Deuteranopia	(N=68)	95.6	0.0	0.0	4.4
Protanomaly	(N=66)	16.7	45.8	0.0	37.5
Protanopia	(N=43)	4.8	88.1	0.0	7.1

Table 3.6: The classification outcome of the Farnsworth D-15 test for subjects who make at least one error. The DA/D/PA/P classification was determined by the Nagel anomaloscope.

The outcomes of the police and fire service multi-test colour assessment protocols are shown in Table 3.7. Both protocols pass all normal trichromats and a similar percentage of deutan ($\sim 55.6\%$ for the police protocol and $\sim 54.8\%$ for the fire service protocol). Approximately 46.8% of protans passed the police protocol and $\sim 1.8\%$ of protans passed the fire service protocol. The deutan and protan who pass the police protocol and the fire service can have RG CAD thresholds up to 21.8 and 23.9 RG CAD units and 17.7 and 13.3 RG CAD units, respectively and the distribution of RG CAD thresholds in deutan and protan who pass and fail both protocols are shown in Figure 3.7 and Figure 3.9. The statistically equivalent RG CAD thresholds for the police multi-test protocol, and the fire service protocol are shown in Figure 3.8 and Figure 3.10.

Classification		Police Protocol		Fire Service Protocol	
		Outcome (%)	Max RG	Outcome (%)	Max RG
Normal	(N=57)	100.0	2.2	100.0	2.2
Deutan	(N=261)	55.6	21.8	54.8	17.7
Protan	(N=109)	46.8	23.9	1.8	13.3

Table 3.7: The outcome of the UK police and fire service multi-test colour vision protocols in terms of the percentage of normal trichromats, deutan, and protan who pass and the maximum RG CAD threshold of individuals who pass. Both protocols consist of several single-test protocol components that include the Ishihara pseudoisochromatic plate test, Farnsworth D-15 test, City University test, and the Nagel anomaloscope.

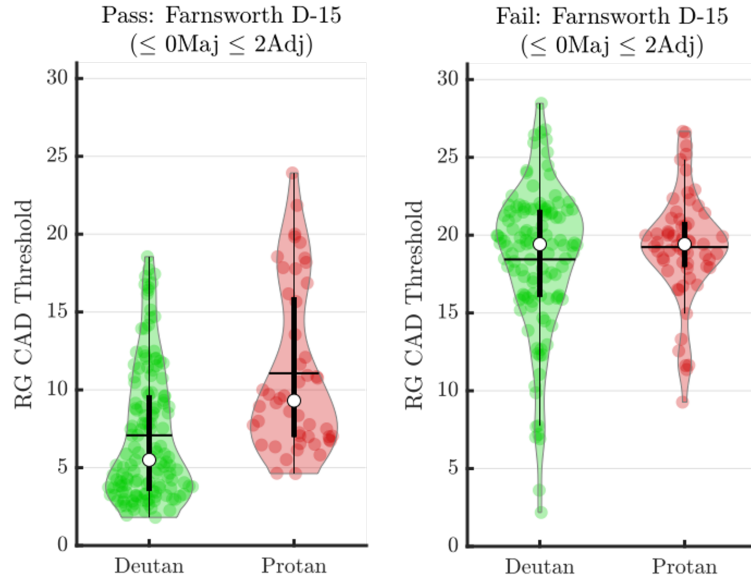


Figure 3.6: Distribution of RG CAD thresholds in protan and deutan subjects who pass and fail the Farnsworth D-15 when up to two adjacent transpositions and no major crossings are accepted. The horizontal line in each violin plot indicates the mean RG CAD threshold.

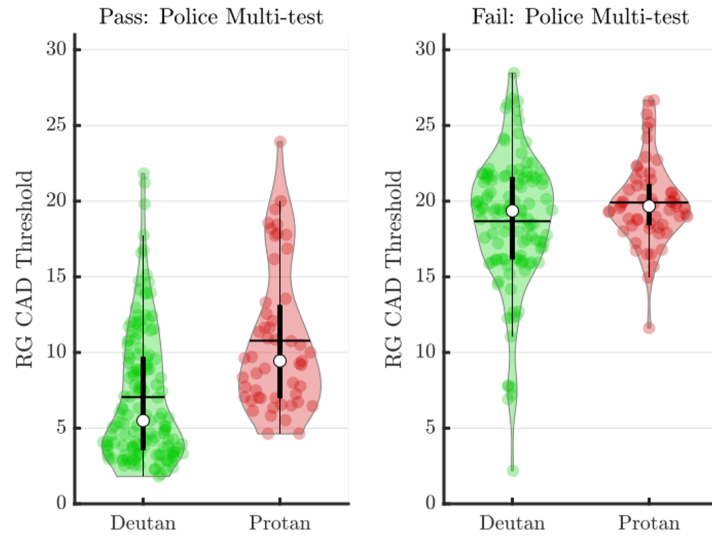


Figure 3.7: Distribution of RG CAD thresholds in protan and deutan subjects who pass and fail the police multi-test protocol. The horizontal line in each violin plot indicates the mean RG CAD threshold.

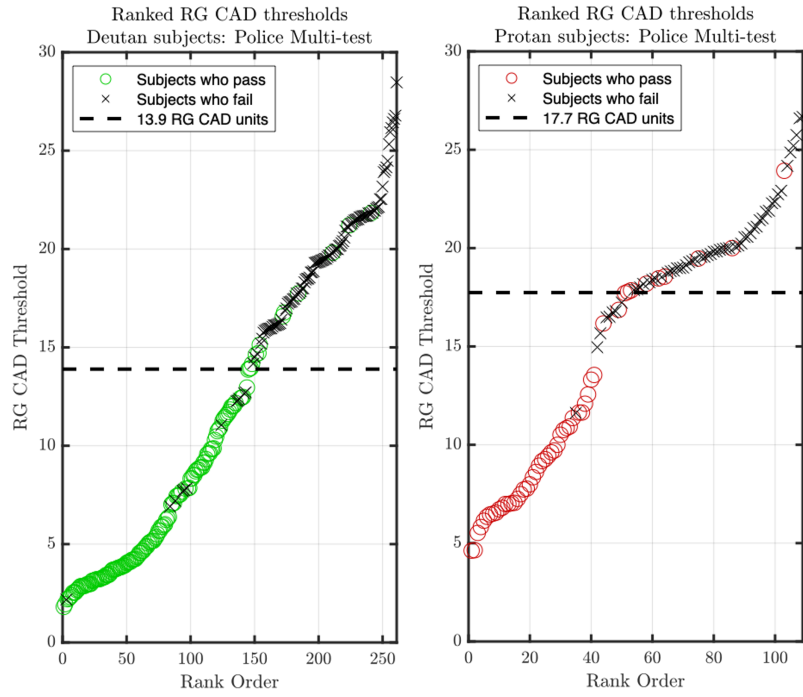


Figure 3.8: Ranked RG CAD thresholds for deutans and protans who pass and fail the police protocol (adapted from Rodriguez-Carmona et al. (2020)). The statistically equivalent RG threshold is 13.9 and 17.7 RG CAD units for deutans and protans respectively.

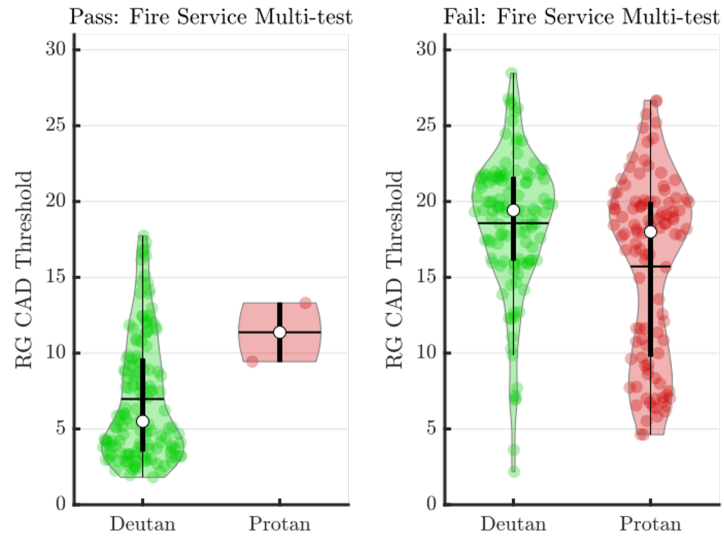


Figure 3.9: Distribution of RG CAD thresholds in protan and deutan subjects who pass and fail the UK Fire Service multi-test protocol. The horizontal line in each violin plot indicates the mean RG CAD threshold.

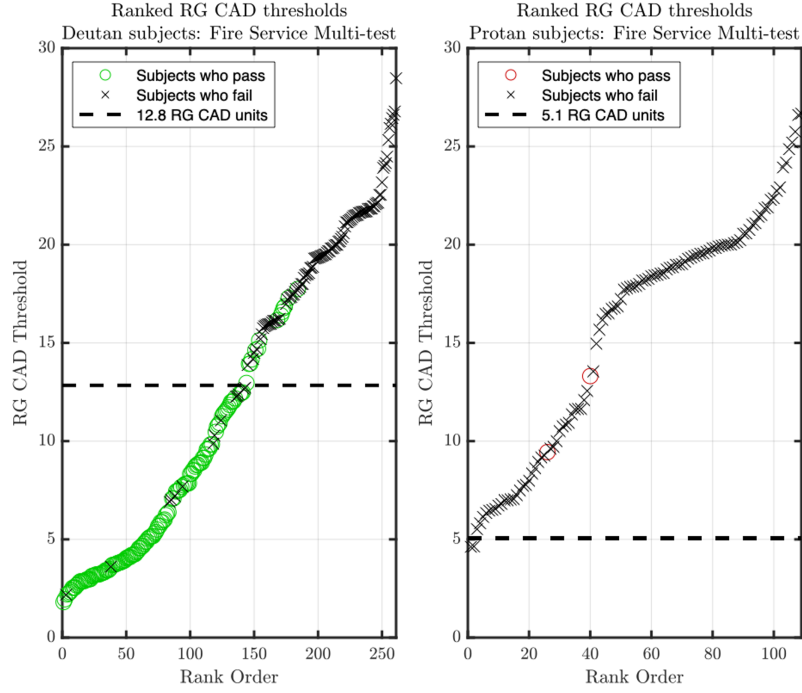


Figure 3.10: Ranked RG CAD thresholds for deutans and protans who pass and fail the Fire service protocol (adapted from Rodriguez-Carmona et al. (2020)). The statistically equivalent RG threshold is 12.8 and 5.1 RG CAD units for deutans and protans respectively. Only two protans pass the protocol.

3.2.4 Discussion

The median RG CAD thresholds for each type of CVD are shown in Figure 3.4, along with the mean, interquartile range, and the overall distribution. For subjects with CVD the median RG CAD threshold is lowest for deuteranomalous subjects, slightly higher in protanomalous subjects and considerably higher in deuteranopes and protanopes (with median values for DA: 7.84, PA: 11.15 and for D: 19.83, P: 19.83 respectively).

The variation in RG CAD thresholds in subjects diagnosed as being dichromats on the Nagel anomaloscope is of interest. Given that these are subjects who, according to the ‘gold standard’ for the classification and clinical separation of dichromats and anomalous trichromats, are dichromats, one would expect them to all have similar levels of RG loss, with small individual variation. The observed variation in these dichromatic individuals is likely due to a number of factors, including changes in the optical density of their cone opsins

and their ability to make use of additional cues, such as luminance contrast cues. These factors will also play a role in the large variation in RG CAD thresholds present in anomalous trichromats, with measured RG thresholds ranging up from 1.81 - 26.61 and 4.62 - 22.92 RG CAD units for deuteranomalous observers and protanomalous observers respectively, in addition to the variation in $\delta\lambda_{max}$ across anomalous trichromat populations (Barbur & Rodriguez-Carmona 2017).

It is important to also consider the manner in which the data were collected. Most individuals who attend the AVOT service at City, University of London do so to complete a colour vision assessment to determine whether they pass an occupational colour vision standard. Many of these individuals will have previously failed some form of screening test (typically the Ishihara test), and although not formally quantified, it is highly likely that some of the individuals in this sample will have encountered other conventional colour vision tests (such as the City University test or the Farnsworth D-15), prior to coming to the AVOT service. In particular, the impact of practicing the Farnsworth D-15 has been discussed in two publications by Ng & Liem (2018) and Ng & Morton (2018) who demonstrated that colour deficient subjects, including dichromats, who initially made errors on the Farnsworth D-15 were able to complete the test correctly after practicing the test. The confounding variable of practice effects and their influence on repeatability and reproducibility is an issue with many conventional methods of colour assessment, including the Ishihara, City University, and Farnsworth D-15 tests.

These limitations are virtually impossible to control as one cannot stop members of the public from completing the Ishihara, City University test, or Farnsworth D-15 at other testing centres or practicing the tests in their own time. In multiple instances, applicants receive an inconclusive colour vision diagnosis at an occupational health care assessment and are referred to multiple optometrists for further colour vision assessment, before being asked to visit the AVOT clinic at City, University of London. In such cases the applicant, through no fault of their own, has multiple opportunities to practice many conventional colour vision tests including the Ishihara, City University,

and Farnsworth D-15 tests, and to improve their performance by learning how to make better use of any additional cues present in the tests (Evans et al. 2020).

3.2.4.1 Single-test protocols

The results reveal how close each single-test protocol comes to reaching its aim. Table 3.1 shows that the Ishihara test can achieve a high sensitivity with poor specificity when 0 errors are accepted on the 38-plate edition with less than 1% of deuterans and protans passing the test and approximately 17.5% of normal trichromats failing. The sensitivity of the test reduces as more errors are accepted to ensure all normal trichromats pass and even with the most stringent pass criteria some individuals with RG CAD thresholds up to 3.9 and 9.4 times larger, for deuterans and protans respectively, than the standard normal CAD observer will pass. The pass rates measured for the subjects in this sample are similar to previously reported data (Rodriguez-Carmona et al. 2012, Rodriguez-Carmona et al. 2020). Provided one is willing to accept that a sizeable percentage of normal trichromats will fail, the most stringent single-test Ishihara protocols are currently the most sensitive screener for congenital RG colour vision deficiency.

The classification outcome of the Ishihara test when only the number identified on plates 21-25 is used fails to correctly categorise the majority of anomalous trichromats, with only 39.1% and 18.5% of subjects with deuteranomaly and protanomaly respectively, being correctly classified by the Ishihara test. A larger percentage of dichromats, 80.9% of deuteranopes and 62.8% of protanopes, were correctly classified.

Birch (1997b) previously reported that 57.2% and 46.6% of subjects with deuteranomaly and protanomaly and 93.4% and 81.9% of deuteranopes and protanopes respectively, were classified correctly based upon the number identified on the last four plates. Chi-squared analysis comparing the two studies revealed no significant difference in the classification outcome of the test for dichromats, but a significance difference for anomalous trichromats at the

0.01 significance level ($\chi^2(1, N = 164) = 6.4, p = .011$ for deuteranopes, $\chi^2(1, N = 126) = 5.6, p = .018$ for protanopes, $\chi^2(1, N = 384) = 12.8, p < .001$ for deuteranomaly, and $\chi^2(1, N = 95) = 8.2, p = .004$ for protanomaly). The lower number of anomalous trichromats correctly classified in this study may be due to differences in the proportion of severe anomalous trichromats present in the sample. Unfortunately there are no available data which quantify the severity of loss in those assessed by Birch (1997*b*) and it is not possible to compare the severity of RG loss in both anomalous trichromat groups.

The findings from this study emphasise the importance of recording and using the perceived luminance of each number (i.e., asking applicants which of the two numbers appears to be brighter) as an additional metric when attempting to classify the type of CVD using the Ishihara classification plates. Given the prevalence of anomalous trichromacy, relative to dichromacy, this is of particular importance if one wishes to use the Ishihara test to classify the type of deficiency for larger populations.

Single-test protocols that employ the City University (2nd Edition) and the Farnsworth D-15 tests fail to filter out those with ‘severe’ RG loss even when the most stringent protocols are used. Deutans and protan subjects with RG CAD thresholds over 20 times larger than the standard normal CAD observer pass both protocols. In terms of the statistical outcome, 42.5% and 39.4% of deutans and protans pass the City University protocol and 54.8% and 44% of deutans and protans pass the Farnsworth D-15 protocol.

The single-test City University protocol outcome when no errors are accepted is similar to historical data from a smaller sample size (Birch 1997*a*). In the sample used in this study approximately 57.5% of subjects with deuteranomaly passed compared to the 55.5% in Birch’s data, despite Birch’s data containing 85 less deuteranomalous subjects. There were 21 more deuteranopes in this sample, but, as with Birch’s 1997 data, no deuteranopes pass the City University single-test protocol. This sample only contained 9 more protanopes and the pass rate (4.7%) was slightly higher than Birch’s data (3.8%). The largest discrepancy occurs for protanomalous subjects, in this study 62.1% passed the protocol compared to 73.3% of protanomalous trichro-

mats in Birch's dataset. However, the 1997 publication only contains data from 15 subjects, compared to the 66 used in this study. Chi-squared analysis comparing the outcome of the single-test City University protocol in the two studies revealed no significant difference in the test outcome at the 0.01 significance level for deuteranomaly, $\chi^2(1, N = 301) = 0.11$, $p = .742$, protanomaly, $\chi^2(1, N = 81) = 0.67$, $p = .414$, and protanopia, $\chi^2(1, N = 95) = 0.04$, $p = .846$. Chi-squared analysis was not possible for the two sets of deuteranope data as no deuteranopes passed the protocol in either study.

The single-test Farnsworth D-15 protocol pass rates obtained in this study, when up to two adjacent transpositions are accepted, are higher than previous reports, including those with a larger sample size $N = 710$ (Birch 2008*b*) and $N = 600$ (Evans et al. 2020), and a smaller sample size $N = 75$ (Dain & Adams 1990). Several factors may contribute to this difference, notably the range of severity of loss in the sample and the method used to collect the data. As shown in Figure 3.4 the cohort examined in this study contains a large number of mild protans and deutans, (particularly deutans) with relatively low RG CAD thresholds. Although the cohort contains some younger individuals, previous studies have shown that children aged from 5 to 12 years are capable of performing colour arrangement tests such as the D-15, albeit with a modified protocol (Dain & Ling 2009). However, in more complex tests such as the Farnsworth-Munsell 100-Hue, results in children can also be affected by their nonverbal IQ (Cranwell et al. 2015).

The collection of CAD and D-15 data allows for the direct comparison between the outcome of the D-15 and the severity of loss, as quantified by RG CAD thresholds. This approach has been employed in a recently published study designed to assess the spread in the severity of RG loss in deutan and protan subjects who pass and those who fail the D-15 test when using the Canadian Air Force color assessment protocol for the Farnsworth D-15 (Almustanyir et al. 2020). The results of the two studies reveal the large variability in the severity of RG color vision loss in both those who pass and those who fail the D-15 protocol. The large difference in sample size, 261 deutans and 109 protans examined in this study, and 40 deutan and 28 protans examined in the

study by Almustanyir, Hovis & Glaholt (2020) as well as the use of different protocols for the Farnsworth D-15, may account for the observed differences, particularly the fewer protans and deutans with less severe loss of RG color vision who fail.

CAD units can be considered to belong to a ratio scale. The median RG CAD thresholds for deutans and protans who pass City University and Farnsworth D-15 single-test protocols that maximise specificity (accepting no errors on the City University test and accepting up to 2 adjacent transpositions and no major crossings on the Farnsworth D-15) are significantly lower than the median RG CAD thresholds for deutans and protans who fail these protocols (Mann-Whitney U tests: $p < .001, r = .734$ & $p < .001, r = .687$ for deutans & protans who completed the City University single-test protocol and $p < .001, r = .764$ & $p < .001, r = .656$ for deutans & protans who completed the Farnsworth D-15 single-test protocol). However, when the outcome of the single-test City University and Farnsworth D-15 protocols are evaluated relative to the severity of RG loss, one finds an overlap in RG thresholds between those who fail and those who pass. As shown in Figure 3.5 and Figure 3.6, although at least 75% of subjects with a congenital color vision deficiency that are dichotomised by the City University and Farnsworth D-15 protocols are being split correctly, and fairly, based upon their RG chromatic sensitivity, the protocols pass some individuals with severe RG chromatic loss (with CAD thresholds up to 23.93 RG CAD units) and fail some individuals with RG thresholds below 4.00 RG CAD units. For the City University test protocol deutans who pass the protocol have RG CAD thresholds ranging from 1.81 - 21.61 RG CAD units whilst deutans who fail have RG CAD thresholds ranging from 3.61 - 28.47 RG CAD units. The RG CAD thresholds for protans who pass range from 4.62 - 23.93 RG CAD units compared to the range of 6.99 - 26.67 RG CAD units for those who fail the protocol. A similar trend is found for the Farnsworth D-15 protocol as the RG CAD thresholds for deutans who pass range from 1.81 - 18.55, deutans who fail range from 2.18 - 28.47, protans who pass range from 4.62 - 23.93, and protans who fail range from 9.27 - 26.67 RG CAD units. The protocols therefore are not dichotomising

the entire CVD population on the basis of severity of RG loss when the most stringent protocols that pass all normal trichromats, and maximise specificity, are used.

Single-test protocols that employ the City University and Farnsworth D-15 tests are not fit for purpose, primarily due to the large number of subjects with CVD who pass, in spite of severe loss of RG colour vision (as shown in Table 3.3, Table 3.5, Figure 3.5, and Figure 3.6). The general statistics for the outcome of these protocols obtained in this study are in agreement with previously reported findings yet single-test City University and Farnsworth D-15 protocols continue to be used, either as standalone protocols or as secondary components in multi-test protocols. The primary justification for the continued use of these tests is based on the desire to fail subjects with ‘severe’ loss of RG chromatic sensitivity and to facilitate the identification of protans who are considered unsafe to work in certain environments where protans may be disadvantaged (Cole 2004, Rodriguez-Carmona et al. 2020). Single-test protocols that utilise the City University or Farnsworth D-15 tests do not achieve either of these objectives.

3.2.4.2 Multi-test protocols

The results for the two multi-test protocols evaluated in this study, the police and fire service multi-test protocols, are shown in Table 3.7. The police protocol was designed for use in the UK police force, specifically for the assessment of police officers who work in specific occupational roles, such as those who carry firearms or conducted energy devices (commonly referred to by the brand name ‘Taser’). The police protocol is one of the most complex in current use and it is the least effective of the two protocols considered here in terms of achieving its stated aims. The protocol aims to pass all normal trichromats while excluding those with either severe loss of RG colour vision and/or those with a protan deficiency. To do this the protocol employs the Ishihara test (38-plate edition) allowing ≤ 4 errors on plates 1 to 21 followed by the City University and Farnsworth D-15 tests if applicants fail the Ishihara test. All applicants diagnosed as protan in at least two of the three tests

also fail the protocol. In spite of its intricacy, the protocol produces poor results allowing 55.6% of deuterans and 46.8% of protans, some with severe loss of RG colour vision, to pass.

Figure 3.7 shows the spread in the severity of RG colour vision loss of those who pass and those who fail the police protocol. Perhaps unsurprisingly considering the protocol's components, the results show the unjust potential outcomes of this protocol that fails several less severe deutan and protan applicants and passes a number of protans and deuterans with severe loss of RG chromatic sensitivity. The protocol produces large variability in RG colour vision loss in deuterans who pass (ranging from 1.81 to 21.83 RG CAD units) and fail (ranging from 2.18 to 28.47 RG CAD units) and protans who pass (ranging from 4.62 - 23.93 RG CAD units) and fail (ranging from 11.62 - 26.67 RG CAD units). The dashed black lines in Figure 3.8 correspond to the 'equivalent' thresholds for deuterans and protans. The number of deuterans and protans who fail the protocol with RG CAD thresholds below 13.9 and 17.7 RG CAD units are equal to the number of deuterans and protans who pass with thresholds above 13.9 and 17.7 CAD units. In the absence of a detailed study designed to establish pass/fail limits specific to the occupational environment in which the police protocol is employed, an 'equivalent' threshold of 13.9 RG CAD units for deuterans and 17.7 RG CAD units for protans can act as an acceptable compromise with significant advantages. Passing all deuterans and protans with RG thresholds below the 'equivalent' CAD threshold results in diminished variability, whilst ensuring that the percentage of protans and deuterans who pass remains unchanged.

The fire service protocol is to screen using a single-test Ishihara protocol which accepts ≤ 2 errors on the first 17 plates of the 24 plates edition followed by secondary testing consisting of a single-test Farnsworth D-15 protocol that accepts ≤ 2 adjacent transpositions and requires subjects to obtain a normal or deutan Rayleigh matching range on an anomaloscope. The use of an anomaloscope results in the protocol excluding almost all protans, at the cost of requiring an experienced clinician. Two protans with RG thresholds of 13.31 and 9.44 CAD units pass the fire service protocol by passing the single-

test Ishihara screening test (as shown in Table 3.1) and, under the normal protocol conditions, would not be required to complete an anomaloscope. As with the police protocol, the fire service protocol produces significant overlap in the deutans who pass and those who fail (Figure 3.9) with an ‘equivalent’ RG CAD threshold of 12.8 RG CAD units for deutans and 5.1 RG CAD units for protans as shown in Figure 3.10.

Both multi-test protocols use a similar syntax in an attempt to achieve the same aim, with the primary difference being the use of the City University test in the police protocol compared to the use of the anomaloscope in the fire service protocol. Interestingly both protocols result in a similar outcome for normal trichromats and subjects with a deutan deficiency. There is, however, a vast difference between the outcome of the two protocols for protans as only 1.8% of protans pass the fire service protocol compared to the 46.8% of protans who pass the police protocol. It’s important to note that this improvement is at the cost of the fire service protocol requiring the use of an anomaloscope and a trained examiner.

A limitation of the approach used in this study is that it only attempts to quantify the outcome of current methods of colour vision assessment, it makes no attempt to try to optimise any of the protocols considered. There is obvious value in determining the optimal protocol for a test (one would hope that this would be carried out when occupational protocols are established) particularly when optimising protocols that make use of current colour vision tests as it’s significantly easier to adapt or modify a protocol using available methods of colour vision assessment, compared to distributing new methods of colour vision assessment. Although improvements through the use of protocols based upon statistically equivalent RG CAD thresholds are described they are not the primary focus of this study. An open question still remains as to how some individuals with severe RG loss are able to complete tests specifically designed with the aim of detecting severe RG and YB loss such as the Farnsworth D-15 and the City University test.

3.2.5 Conclusions

Single- and multi-test colour assessment protocols that rely on conventional methods of colour vision assessment fail to meet current colour assessment requirements and consequently fail to achieve their aims.

The use of the Ishihara test in single-test protocols can provide high sensitivity at the cost of poor specificity. Ishihara single-test protocols do not reliably provide an assessment for the type or measurement of the severity of any potential colour vision deficiency. Single-tests City University and Farnsworth D-15 protocols fail to dichotomise the population on the basis of severity of RG loss, particularly when protocols designed to maximise specificity are used.

The deficits present in single-test protocols that make use of conventional methods of colour vision assessment are not eliminated by combining multiple single-test protocols into a multi-test protocol. This is particularly evident when the aims of the protocol are to only pass those with colour vision loss below a specific level and to eliminate all protans.

Replacing current protocols with ‘equivalent’ CAD thresholds would ensure that applicants with severe loss fail whilst simultaneously being more equitable to subjects with mild loss by reducing the variability in the protocol outcome. The use of ‘equivalent’ CAD thresholds would also ensure that the statistical outcome of the protocol remains unchanged.

3.3 Exploratory Farnsworth D-15 study

3.3.1 Introduction

Following its introduction over 70 years ago (Farnsworth 1947) the Farnsworth D-15 continues to be used to screen for individuals with reduced chromatic sensitivity in occupational and clinical environments (Margrain et al. 1996, Dain 2004, Dain et al. 2009, 2015, Almustanyir et al. 2020). The limitations of the Farnsworth D-15, including the potential effects on the Farnsworth D-15 from variation in illuminants, have recently been highlighted by Dain, Atchison & Hovis (2019) and the impact that practice can have on the outcome of the Farnsworth D-15, and the subsequent suitability of the Farnsworth D-15 has been questioned (Ng & Liem 2018, Ng & Morton 2018). Given the extensive use of the Farnsworth D-15 in the clinic, and particularly in occupational settings, the test still has relevance today. It is important to note, and appreciate, in most clinical and occupational settings, the Farnsworth D-15 is employed as a secondary test, used if applicants fail an initial screening test, such as the Ishihara pseudoisochromatic plate test (Cole 2007).

The observed variability in RG chromatic sensitivity in colour deficient subjects who pass and fail the Farnsworth D-15 test suggests that subjects make use of multiple signals to carry out the task. The extent to which subjects make use of these signals will depend both on the signals available and the attention given to any additional clues, with the latter being minimized through the use of standardized instructions. Given that all subjects described in section 3.2 had normal YB chromatic sensitivity (as measured by the CAD test), it is of interest to know the extent to which the subjects make use of YB colour signal changes to pass the test.

Since many subjects with severe RG colour deficiency pass the Farnsworth D-15, the principle aim of this exploratory study was to identify the residual signals that enable these subjects to successfully arrange the caps used in the Farnsworth D-15 test.

3.3.2 Methods

The spectral radiance of each cap when illuminated with D65 was measured using a Konica CS-2000A telespectroradiometer. The spectral radiance data were used in combination with the Smith and Pokorny cone fundamentals (DeMarco et al. 1992) to estimate cap-specific cone excitations. The spectral luminous efficiency function, $V'(\lambda)$, was used to calculate the corresponding rod photoreceptor excitations (Wyszecki & Stiles 1982).

Two reduced versions of the Farnsworth D-15 task were created to investigate the residual signals that enable subjects with severe RG deficiency to successfully arrange the Farnsworth D-15 caps. The experiment was designed to establish whether in the absence of normal RG chromatic signals, sequential, monotonic step changes in S cone and rod signals between adjacent caps can provide sufficient information to enable dichromats and rod monochromats to order the caps correctly by minimising the perceptual differences between adjacent caps. The illumination was kept constant and no new caps were introduced. The 16 caps employed in the Farnsworth D-15 test were split into two cap subgroups. Given the aim of investigating those with severe RG loss the two cap subgroups were selected based on the measured S-cone and rod excitations shown in Figure 3.11D. The first cap subgroup (caps P-8) exhibit monotonic decrements in S-cone and rod signals, whilst the second cap subgroup (caps 9-15) exhibit positive monotonic increments in S-cone and rod excitations. Note that caps 9 and 10 exhibit approximately equal rod excitations.

An individual with *CNGB3*-associated rod monochromacy³ (identified through genetic testing at Moorfields Eye Hospital London) with a visual acuity of 20/200, a tritanope, two protanopes and two deuteranopes were assessed using the two Farnsworth D-15 protocols. The protanope and deuteranope were identified using the Nagel anomaloscope. The tritanope was identified using the CAD test, in which the subject displayed normal RG colour thresholds, but exhibited specific loss along the tritan confusion axis that was only lim-

³*CNGB3* (Cyclic Nucleotide Gated Channel Subunit Beta 3) is a protein coding gene associated with achromatopsia

ited by the maximum chromatic displacements that can be achieved on the visual display, see Figure 3.14C. These subjects completed the standard D-15 protocol before the caps were separated into two cap subgroups, from caps P-8 and caps 9-15, and subjects were asked to arrange the caps in each cap subgroup as to minimise the perceived differences between each cap, repeating the D-15 protocol with a reduced selection of caps. All subjects arranged the two subgroups within two minutes of being presented with the caps. A normal trichromat also completed the Farnsworth D-15, Nagel anomaloscope and the CAD test. Informed consent was obtained from all subjects. The study was conducted in compliance with the City, University of London research and ethical guidelines and followed the tenets of the Declaration of Helsinki.

3.3.3 Results

The predicted photoreceptor excitations generated by the D-15 caps are shown in Figure 3.11D. When two protanopes and two deuteranopes completed the D-15 under the standard protocol their results produced typical error patterns, with participants making multiple major iso-chromatic crossings (Figure 3.12B and Figure 3.13B). When the caps were separated into two subgroups, both protanopes and deuteranopes were able to arrange the caps within each subgroup with no errors. When a tritanope completed the same protocol, they were unable to arrange the caps correctly under standard conditions. This was also the case when presented with each of the two subgroups (Figure 3.14B).

Not unexpectedly, the individual with *CNGB3*-associated rod monochromacy made several errors on the D-15 test under normal conditions (Figure 3.15B). When presented with each of the two subgroups, the rod monochromat was able to arrange the caps almost completely correctly, only making one minor transposition (cap 9 and 10 during the first test and interchanging the order of these two caps in repeated tests). These caps have approximately the same rod excitation, see Figure 3.15D.

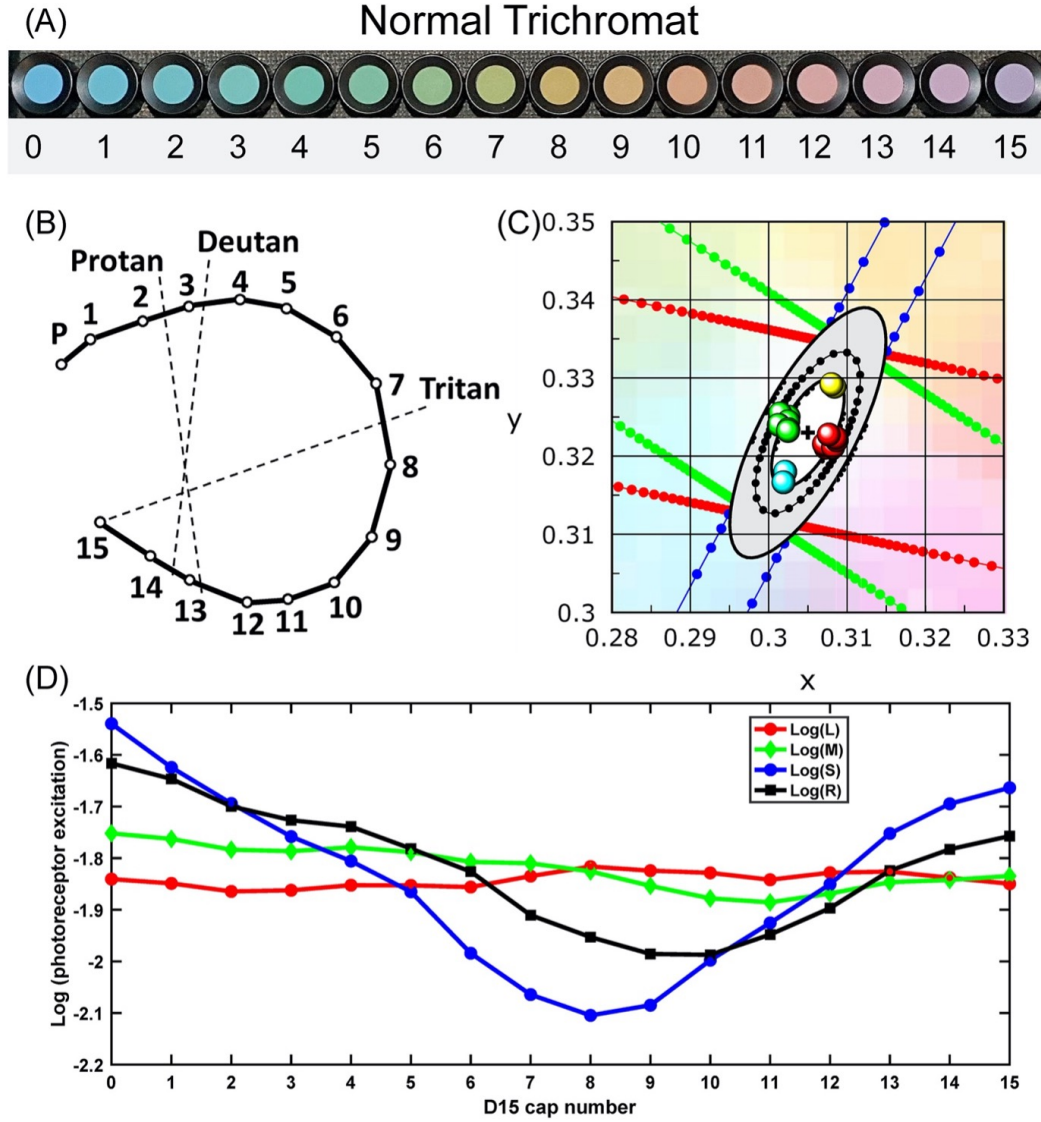


Figure 3.11: Measured data and theoretical photoreceptor excitations for a normal trichromat. (A) and (B), show the Farnsworth D-15 test results for a normal trichromat. (C) shows the normal trichromat's CAD test results displayed in CIE 1931 colour space. The normal trichromat's CAD thresholds lie within the grey ellipse that indicates the normal range of CAD thresholds expected for this subject's age. (D) shows the predicted photoreceptor excitations for each of the Farnsworth D-15 caps when illuminated with a lamp designed to approximate a D65 illuminant. These predictions were generated by measuring the spectral radiance of each of the 16 caps when illuminated with the D65 illuminant. Figure adapted from Evans et al. (2020).

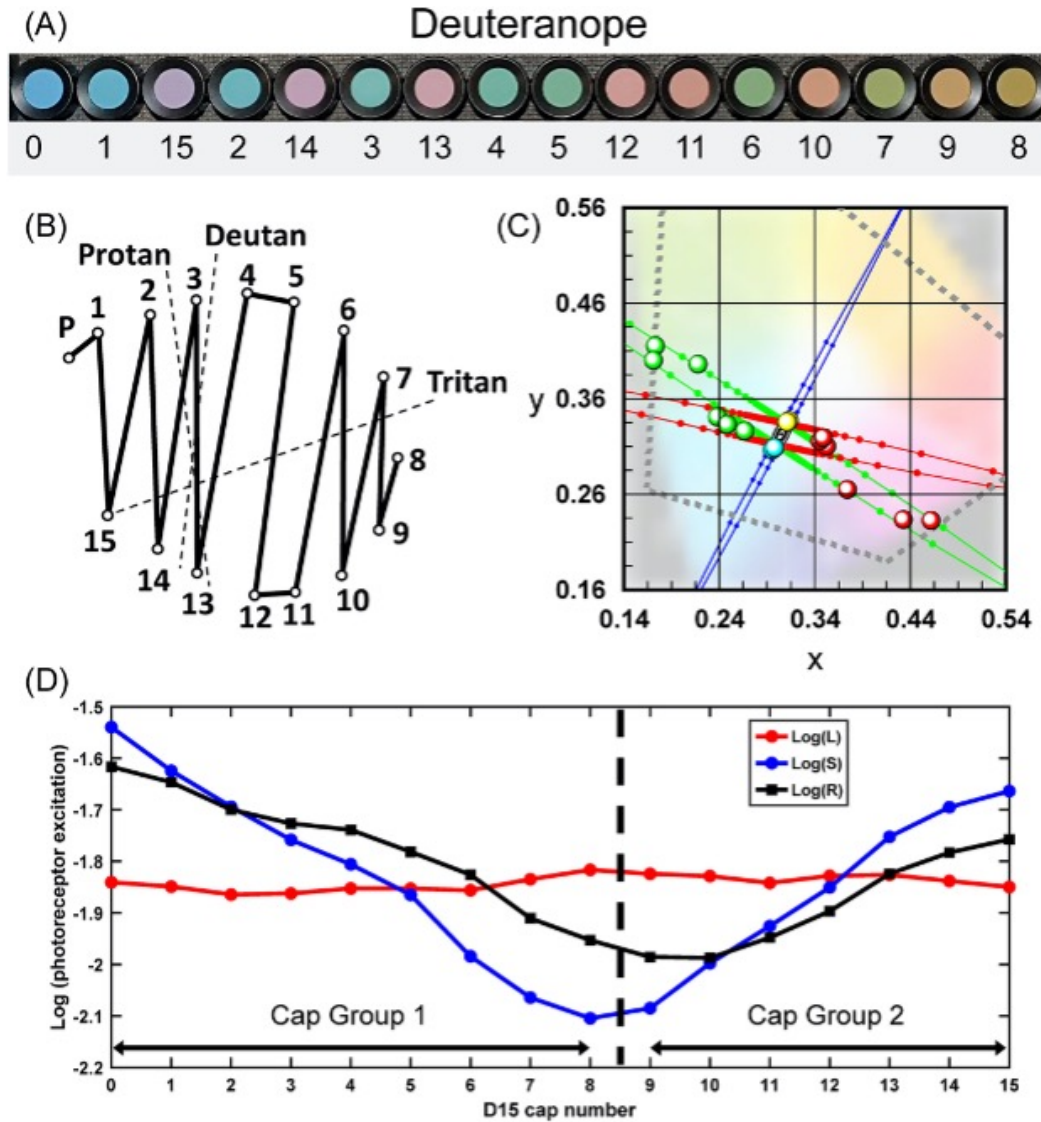


Figure 3.12: Measured data and theoretical photoreceptor excitations for a deuteranope. (A) and (B) show the recorded Farnsworth D-15 test results for an individual with deuteranopia. (C) shows the deuteranope's CAD test result. (D) shows the predicted photoreceptor excitations for each of the Farnsworth D-15 caps when viewed by an individual with deuteranopia. When separated in to two groups, the deuteranopes arranged all caps in the sequence recorded for normal trichromats, with no errors in each group. Figure adapted from Evans et al. (2020).

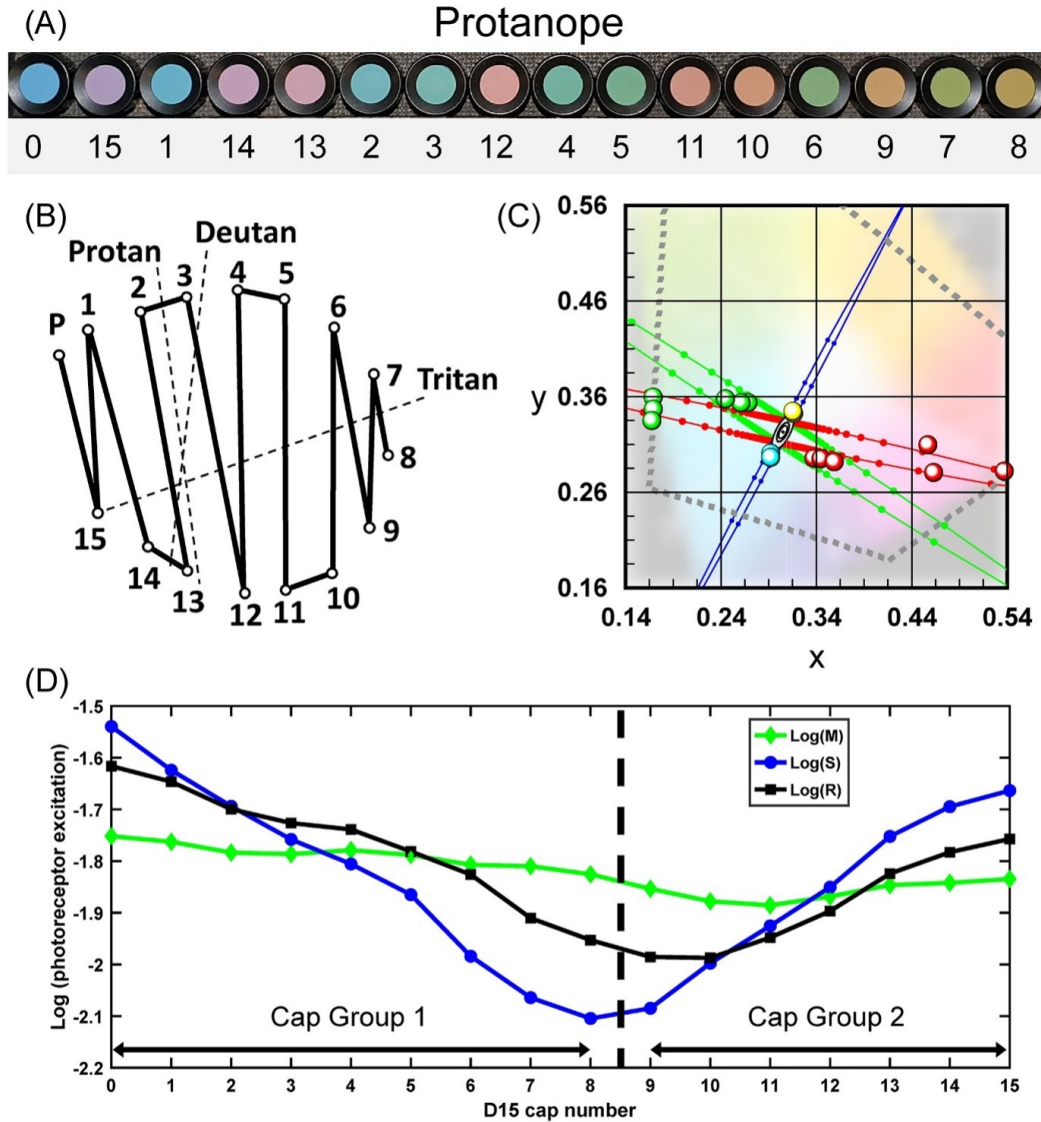


Figure 3.13: Measured data and computed photoreceptor excitations expected for a protanope. (A) and (B) show the recorded Farnsworth D-15 test results for an individual with protanopia. (C) shows the protanope's CAD test result. (D) shows the predicted photoreceptor excitations for each of the Farnsworth D-15 caps when viewed by an individual with protanopia. When separated into two groups, the protanopes arranged all caps in the sequence recorded for normal trichromats, with no errors in each group. Figure adapted from Evans et al. (2020).

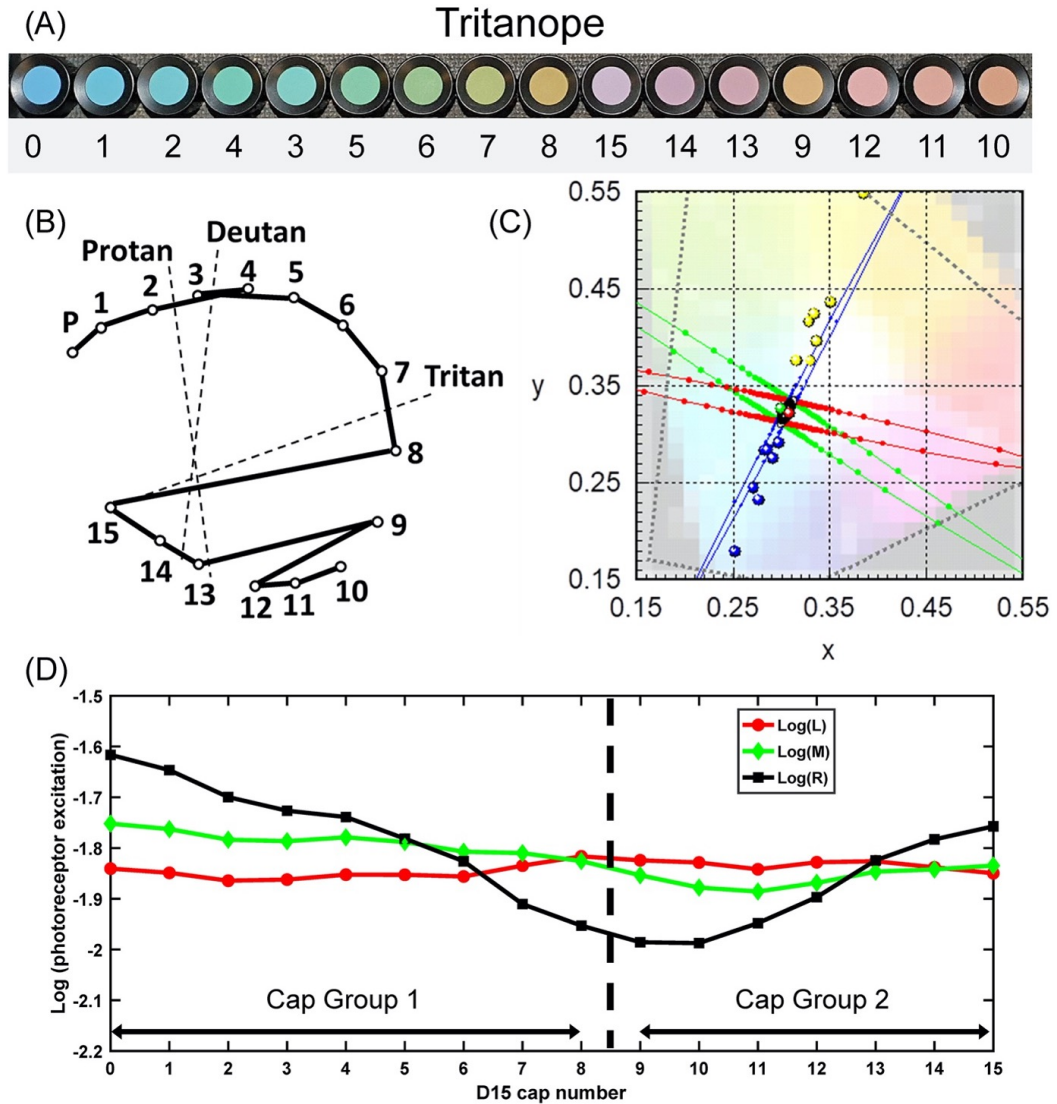


Figure 3.14: Measured data and theoretical photoreceptor excitations for a tritanope. (A) and (B) show the recorded Farnsworth D-15 test results for an individual with tritanopia. (C) shows the CAD results for the individual with tritanopia. (D) shows the predicted photoreceptor excitations for each of the Farnsworth D-15 caps when viewed by an individual with tritanopia. When separated into two groups, the tritanope is unable to arrange all caps in the sequence recorded for normal trichromats, making multiple errors in each group. Figure adapted from Evans et al. (2020).

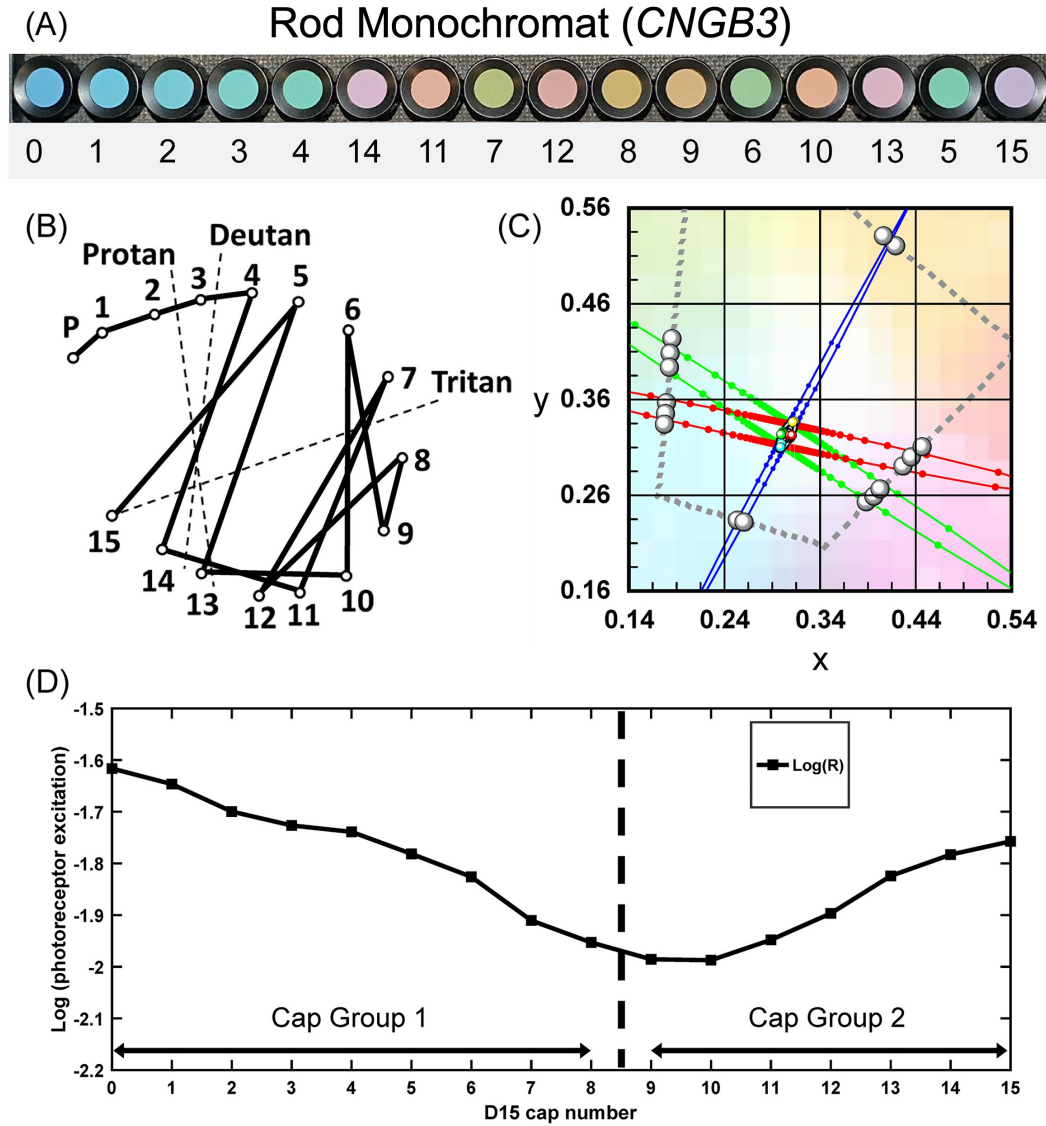


Figure 3.15: Measured data and theoretical photoreceptor excitations for an individual with *CNGB3*-associated rod monochromacy (*CNGB3*). (A) and (B) show the recorded Farnsworth D-15 test results for an individual with *CNGB3*-associated rod monochromacy. (C) shows the CAD results for the individual with *CNGB3*-associated rod monochromacy. (D) shows the predicted photoreceptor excitations for each of the Farnsworth D-15 caps when viewed by an individual with *CNGB3*-associated rod monochromacy. When separated in to two groups, the rod monochromat arranges the caps almost completely correctly, only making one minor transposition (cap 9 and 10 during the first test and interchanging the order of these two caps in repeated tests). Figure adapted from Evans et al. (2020).

3.3.4 Discussion

The expected changes in S-cone photoreceptor excitation when viewing the Farnsworth D-15 caps are shown in Figure 3.11D. As one moves across the caps employed in the Farnsworth D-15 one observes a large change in the predicted S-cone photoreceptor signal, accompanied by a slightly smaller change in the predicted rod-photoreceptor excitation. There is relatively low variation in the theoretical M- and L- cone photoreceptor excitation, by contrast, when one moves across the Farnsworth D-15 caps. The predicted photoreceptor excitations do not account for the relative number of photoreceptors in the eye, or post receptor processing. However, they do provide, at least on a basic level, an indication to the initial signals generated by the Farnsworth D-15 caps at the earliest stages of visual processing.

A limitation of this approach is the consideration of only 4 categories of photoreceptor in the eye. Intrinsically photosensitive retinal ganglion cells (ipRGCs) have recently been shown to play a role in form vision in the peripheral retina (Allen et al. 2019) and their impact upon non-image forming visual functions has also been investigated in primate retina (Patterson et al. 2020). Due to the sparse distribution and large receptive fields involved, ipRGCs are unlikely to notably contribute towards the successful completion of the Farnsworth D-15 task, given the relatively small angular size of each cap and the foveal location of the primary image location (Markwell et al. 2010, Schmidt et al. 2011, Cao et al. 2018). The spectral responsivity of ipRGCs overlaps significantly with that of rods. Rod signals are the greatest signal contributor to ipRGCs (Schroeder et al. 2018, Lee et al. 2019). The light levels involved in Farnsworth D15 tests are low photopic and therefore most unlikely to involve melanopsin mediated signals. The results and conclusions drawn from these findings remain unchanged even if rod signals associated with ipRGCs contribute to the perceived brightness difference that help with the Farnsworth D-15 task in rod monochromats. Further studies are required to fully establish the role ipRGCs may play in hue arrangement tasks, such as the Farnsworth D-15.

The results obtained from dichromats and a rod monochromat who also completed the two protocols demonstrate that protanopes, deuteranopes, and even rod monochromats are able to make use of the relatively large changes in S- cone and rod photoreceptor signals to complete the Farnsworth D-15 test. Protanopes and deuteranopes who make multiple isochromatic crossings under the standard Farnsworth D-15 protocol, made no errors when the caps were split into two groups. This is not surprising; the diminished or even absent RG colour signal differences that prevent typical protan and deutan confusions, particularly major crossings, are no longer possible and caps with equal S-cone excitation are not available within either of the two cap subgroups. This finding is in agreement with work by Montag & Boynton (1987) who found that rod photoreceptor signals can be used by protanopes and deuteranopes to help discriminate and categorise slightly larger surface colours ($\sim 4^\circ$).

The results in the tritanope (Figure 3.14) suggest strongly that normal RG colour signals are not sufficient to pass the Farnsworth D-15 test with no errors, since the tritanope makes multiple errors when presented with the full set of caps and also makes errors when presented separately with each of the two subgroups. The rod monochromat, on the other hand, can only rely on monotonic changes in rod signals which can be used effectively within each subgroup to arrange the caps in the correct sequence. The tritanope makes multiple errors within each subgroup, in spite of normal rod function. An interesting consequence of this finding is that either the presence of normal RG colour signals or simply the interaction between M, L and rod signals reduces the effectiveness of monotonic rod signal changes to carry out the test.

The results obtained in the rod monochromat merit further discussion. The results shown in Figure 3.15C reveal the complete absence of both RG and YB colour vision with thresholds limited only by the gamut of the visual display employed in the CAD test. Despite making typical confusions expected for a rod monochromat (Birch 2001) with the standard Farnsworth D-15 protocol, the rod monochromat was able to arrange the caps correctly when they were

separated into the two subgroups with only one minor transposition (mixing caps 9 and 10 in the first test). Given that these caps generate almost equal rod photoreceptor signals, as shown in Figure 3.15D, this result is expected. The results show that monotonic changes in rod photoreceptor signals, when the retina contains only functioning rods, can be used to complete the arrangement of Farnsworth D-15 caps without errors, but only when the ambiguity of equal rod signals is removed by separating the caps within the two subgroups. Systematic diseases and diseases of the retina can result in a reduction in luminance contrast and flicker sensitivity, in addition to loss of RG and/or YB colour vision (Rauscher et al. 2013, Bi et al. 2016). The Farnsworth D-15's inability to isolate colour signals and to quantify severity of RG and YB loss impacts its usefulness as a test for use in occupational settings. However, in clinical settings one may wish to distinguish those with acquired visual loss that can affect chromatic sensitivity but may not be specific to just colour vision from those with normal, healthy vision. A test that requires the use of multiple visual signals to complete, such as the Farnsworth D-15, may prove to be sensitive and effective in achieving this aim.

3.3.5 Conclusions

Neither RG nor YB colour signals in isolation are sufficient to pass the Farnsworth D-15 test with no errors. The results show that a subject can make use of monotonic changes in rod signals to arrange the caps in the correct sequence, but only in rod monochromats when rod signals do not interact with other cone signals.

3.4 Alternatives to single- and multi-test protocols

Despite the plethora of methods designed for assessing human colour vision, current colour assessment protocols that rely on commonly used methods of colour vision assessment fail, on an individual level, to achieve their intended

aims. Additionally, on a more general level, they fail to fully meet the needs of colour assessment protocols. The outcome of these protocols, their limitations, and the potential reasons as to why they are performing poorly have been described in sections 3.2 and 3.3. The critical limitation in any colour assessment protocol is determined by the tests used in the protocol. A protocol can complement and enhance a test's usability and applicability, but it cannot surpass the test it is imposed upon. This is a fundamental principle of human colour vision assessment.

There are several alternatives to the current protocols based upon conventional methods of colour vision assessment. The first of which, a protocol based on 'equivalent' CAD thresholds, has already been introduced. The advantage of a protocol based upon an 'equivalent' CAD threshold, particularly within occupational settings, is that it would enhance safety by ensuring that individuals with severe RG and/or YB colour vision deficiency are detected. Such a protocol would minimise variability by making use of the relatively small within-subject variability present in the CAD test and the CAD tests' ability to isolate colour signals and other cues (Barbur et al. 2020).

A statistically equivalent CAD threshold can be established for any occupational protocol. The implementation of an 'equivalent' CAD threshold protocol is simply an optimisation, and reduction in variability, of an existing protocol. In occupational settings there are two alternatives to an 'equivalent' CAD threshold: a specific limit, or limits, based upon the minimum level of chromatic sensitivity required to carry out the safety-critical colour-related tasks or the use of a practical occupational assessment that mimics the safety-critical colour-related tasks that individuals working in the occupational setting need to complete.

Specific limits based upon the minimum level of chromatic sensitivity required to complete a safety-critical colour-related task were established and implemented for commercial aviation in the UK and Europe. Detailed studies were carried out to find the safety-critical colour task, develop a simulation of the task, and establish deutan and protan specific pass/fail limits (Barbur et al. 2009, 2016). The limits established following these studies are currently in

use (≤ 12 CAD units for protans and ≤ 6 CAD units for deutans) and as with a protocol based upon ‘equivalent’ CAD thresholds, the current aviation protocol benefits from the reduced variability of the CAD test. Additionally, following the study and the subsequent introduction of the new limits 35% of colour deficient applicants now meet the aviation colour vision standard and are able to obtain a commercial pilot’s licence. An additional advantage of protocols based upon the CAD test is that they allow for a consideration of how human colour vision changes as a function of age, due to the normal age adjusted upper and lower limits for normal trichromats built into the test. This is particularly useful in maritime and coastguard agency and air traffic control environments when one wishes to only pass normal trichromats (Carter & Barbur 2015, Barbur et al. 2016).

Unfortunately, although a full study to investigate the safety-critical colour tasks is a holistically better approach, it is resource intensive, both in terms of the required cost and time investment. Normally such a study is carried out only if/when the governing body for the occupational setting decides that a study is required. Sadly, when looking at the history of colour vision assessment such changes typically only occur after a potentially avoidable incident occurs, often resulting in the loss of life and damage. The second alternative is incorporating the safety-critical colour-related tasks into a practical occupational assessment. This can be included as part of the occupational training and selection process, ensuring that applicants who successfully pass through the recruitment process demonstrate that they are able to complete any and all safety-critical colour-related tasks at the level required for their chosen occupational position. A simple example would be the introduction of chromatic targets and colour selection tasks in police firearms selection training, with the aim of ensuring that all applicants who pass the training have, in doing so, demonstrated that they have the required level of chromatic sensitivity to safely carry out their duties. A practical occupational based assessment seems intuitively appealing. However, as with a detailed study, this approach requires a thorough analysis of the safety-critical colour-related tasks, an understanding of the tasks that individuals with any type of colour

vision deficiency may find challenging, and the inclusion of comprehensive practical occupational testing as part of the recruitment process.

3.5 Implications

The stark reality of the current state of conventional colour vision assessment has been revealed by the studies described in this chapter. Current single- and multi-test protocols fail to achieve their aims and are unable to achieve high sensitivity and specificity or reliably quantify the severity of RG and YB loss. There are several reasons for this. Conventional methods of colour vision assessment do not fully isolate RG and/or YB colour signals. The tests and protocols do not eliminate or successfully mask additional cues that may be used to complete the tasks. Additionally, no conventional method of colour vision assessment investigated here attempts to take the changes in chromatic sensitivity as a result of healthy aging into consideration (Barbur & Rodriguez-Carmona 2015).

The CAD test offers clear advantages over conventional methods of colour vision assessment, such as the City University and Farnsworth D-15. As previously described in subsection 2.3.5.2, the CAD test isolates the use of RG and YB colour signals, has relatively small within subject variability, and unlike the anomaloscope the test does not need to be administered by a trained examiner. The normal age-adjusted limits are built into the test, allowing for the consideration of changes to RG and YB colour vision as a function of healthy aging. The test accurately classifies the type of colour vision deficiency and is able to detect acquired loss even when it is combined with congenital deficiency - a particularly valuable property for a test used in clinical environments. The primary disadvantages of using the CAD test in such protocols is that it requires expensive specialist equipment, including a fully calibrated display and the test takes a relatively long time ($\sim 10 - 15$ minutes) to complete. The later point is a key limitation, particularly if one wishes to screen a large population.

Alternatives to current protocols based on conventional methods of colour vision assessment are available. Statistically ‘equivalent’ CAD thresholds act to keep the number of applicants with colour vision deficiency who pass constant whilst offering reduced variability in the protocol outcome. For occupational settings, an improvement could be made through the use of bespoke studies to establish the colour-critical safety-related tasks and to establish the level of chromatic sensitivity required to safely carry out these tasks. Testing to determine whether applicants meet the required level of chromatic sensitivity can then be established under laboratory conditions or as part of practical training in the occupational environment.

As protocols consisting of an initial screening component followed by the CAD test begin to approach the limit of a colour vision protocol, the limitations imposed by the screening component become more pronounced. Current screening protocols based upon the Ishihara test struggle to effectively balance sensitivity and specificity. An additional disadvantage of such protocols is that they do not screen for YB loss, this is particularly significant in clinical settings where it is important to detect the earliest stages of acquired colour vision loss. An ideal ‘two-step’ protocol would employ a quick and efficient test that is able to screen for RG and YB colour vision deficiencies. Such a test would not only need to be easy to use, administer, and interpret, but it would also need to fully isolate RG and YB colour signals and minimise any additional cues.

In occupational and clinical settings an improved understanding of the limits of human colour vision assessment and technological advances in visual displays can overcome many of the limitations present in conventional colour vision assessment. This in turn will lead to safer and more equitable colour assessment protocols with clear benefits for occupations and clinicians alike (Barbur et al. 2020). The new ‘Colour Vision Screener’ produced by City Occupational Limited aims to efficiently screen for RG and YB loss and, when combined with the CAD test in a ‘two-step’ protocol, may result in a protocol that addresses many of the challenges described in section 2.4.1 and progresses further towards the potential upper limit of ‘full’ colour vision assessment.

Chapter 4

The development and validation of the Colour Vision Screener

The current state of conventional colour vision assessment, including limitations of several commonly used colour assessment protocols, were investigated in chapter 3. Despite many of the innovative techniques described in chapter 2 becoming more widely available, tests often remain too resource intensive for the screening of large populations, both in terms of the time taken and equipment required.

The focus of this chapter is the Colour Vision Screener (CVS), a new display-based test developed to provide a rapid means of screening for normal red/-green (RG) and yellow/blue (YB) colour vision to identify the presence of both congenital and acquired colour vision deficiencies. Studies designed and carried out to determine the parameters, optimise, and validate the new CVS test are discussed.

Some of the work described in this chapter has been published in the following paper:

- Barbur, J. L., Rodriguez-Carmona, M., & Evans, B. E. W. (2020), ‘Color vision assessment-3. An efficient, two-step, color assessment protocol’, *Color Research & Application* **46**(1), 33–45.

4.1 Introduction

Colour vision screening is the only facet of colour vision assessment that the majority of the general population ever encounter. Screening for normal trichromatic colour vision is carried out in a variety of environments, including occupational, educational and clinical settings, to identify any potential congenital or acquired CVD (Birch 2001, Goh et al. 2014). The wide range of settings in which one finds colour vision screening is, in no small part, due to the enhanced ability to interpret a visual scene, and subsequent benefit to visual performance, provided by colour vision (Regan et al. 2001, Martin et al. 2004). Given the importance of screening for normal trichromatic colour vision the impact, significance and value of colour screening tests cannot be understated.

The most common colour screening protocols in occupational health and primary healthcare are based on the use of Ishihara test plates (Margrain et al. 1996, Birch & Chisholm 2008, Stockman 2018). There are several advantages the Ishihara test plates provide in these screening protocols, namely, their portability, ease of administration, and ease of interpretation, generally only requiring examiners to simply count the number of errors made and apply a pre-determined cut-off. However, as mentioned in chapter 2, and shown experimentally in chapter 3, protocols that employ the Ishihara plates often fail normal trichromats and pass subjects with congenital and acquired colour vision deficiency. This variability in outcome is primarily as a result of the variation in both the number of errors accepted and the plates used in the protocol. A consequence of this variation is that many applicants who fail current protocols have adequate residual chromatic sensitivity whilst others who pass, often exhibit severe deficiency (Rodriguez-Carmona & Barbur 2017, Rodriguez-Carmona et al. 2020).

At the start of the 21st century, technical advances in display-based technologies combined with an increased understanding of the spatiotemporal properties of human vision led to the development, validation and commercial introduction of the Colour Assessment and Diagnosis (CAD) test (Birch

et al. 1992, Barbur et al. 1994, Rodriguez-Carmona et al. 2005, Rodriguez-Carmona 2006). As previously discussed, the CAD test is able to effectively detect both acquired and congenital colour vision deficiency, to accurately quantify the severity of colour vision loss, and to classify the type of any potential deficiency (Barbur & Rodriguez-Carmona 2017). The CAD test normally takes between 10 to 15 minutes to complete. A small percentage of subjects are offered an enhanced RG or YB CAD test, consisting of four standard RG or YB CAD tests. The enhanced CAD test is only offered to those within the uncertainty limits for their chosen occupational environment on the test or the upper normal limit for their age. The 10-15 minutes taken to complete a standard CAD test are not ideal, particularly if one wishes to screen a large population. However, out of the assortment of current methods for colour vision assessment, the CAD test comes closest to achieving the ideal requirements for a method of colour vision assessment.

An improvement to a protocol based solely on the CAD test, as introduced in chapter 3, is a multi-test protocol that uses both the CAD test and a screening test that is able to screen the population with high sensitivity and specificity. Such a protocol would efficiently identify and separate those who require full testing from individuals with normal trichromatic colour vision, optimising the time and resources required.

A single-test protocol using the Ishihara plates is the obvious candidate to be used as the first step in this theoretical protocol. The 38-plate edition of the Ishihara test has the highest sensitivity, but a poor specificity resulting in approximately 18% of normal trichromats failing the test. Assuming one wishes to test a large random selection of the population (50% male and female), just under 22% of the population (24.5% of males and 18.4% of females assuming a prevalence of congenital CVD of 8% in men and 0.5% in women) will fail the Ishihara plates and require full colour vision assessment. A perfect screener could reduce the percentage of individuals who require full colour vision assessment by a factor of ~ 3.4 (assuming a 2% prevalence of acquired colour vision deficiency). Ishihara protocols also have several other limitations including the inability to screen for YB deficiencies and the

poor correlation between the number of errors made and the severity of loss (Rodriguez-Carmona et al. 2012).

To attempt to overcome the issues present in current screening protocols, a Colour Vision Screener (CVS) test was developed by City Occupational Limited¹.

4.2 The Colour Vision Screener (CVS)

The CVS has been developed for use in primary healthcare and visually demanding occupations to avoid issues and limitations present in current colour vision screening protocols. The test aims to screen for congenital and acquired CVD with close to 100% sensitivity without sacrificing specificity by making use of the knowledge gained in isolating the use of colour signals in human vision and the use of visual displays (Barbur et al. 1994, Barbur & Connolly 2011, Rodriguez-Carmona & Barbur 2017).

One of the key discoveries that facilitated the development of the CVS was the establishment of normal CAD limits for healthy, young subjects (Rodriguez-Carmona et al. 2005). The upper and lower normal age adjusted limits are determined by the mean CAD threshold $\pm 2.5\sigma$, where σ is the standard deviation in CAD thresholds observed in 330 normal trichromats. Less than 1% of the normal population have CAD thresholds at the upper normal limit and approximately 7% of the least affected deuterans obtain CAD thresholds ≤ 2.35 CAD units (Rodriguez-Carmona & Barbur 2017, Barbur & Rodriguez-Carmona 2017).

Although the CVS is built upon the foundation formed by the CAD test there are several key differences between the CAD and CVS in terms of the implementation, design, and the theory associated with the design choices made.

¹City Occupational Limited also develops and markets the Advanced Vision and Optometric Tests (AVOT), including the CAD test

4.2.1 CVS implementation

The goal of the CVS development carried out by City Occupational Limited was to produce two versions of the CVS. A version designed to be carried out using calibrated displays accessible via a menu function in the CAD test and a freely downloadable standalone version to be used by the general public under the Windows operating system on computers with displays that support the sRGB colour standard. Although the fundamental test procedure is the same for both versions there are one or two key differences. These differences are primarily due to the distinction between a downloadable user interface in a standalone program and the CAD test interface, and chromaticity and luminance modifications to allow for the use of uncalibrated displays using the sRGB colour standard.

Studies carried out and described within this thesis use the first version of the CVS, designed for use on calibrated displays. For brevity, only a description of this first version of the CVS is included here.

4.2.2 CVS design and theory

The CVS is intended to be efficient and effective with practicality and usability at the heart of the test's design. The test uses the same background field and stimulus geometry as the CAD test. Although the preliminary version of the CVS requires a viewing distance of 0.8 metres, compared to the 1.4 metres used in the CAD test, the angular dimensions of the stimulus and background are the same in both tests. The viewing distance is reduced in the preliminary version of the CVS so as to make it easier to complete the test using a numeric keypad on a personal keyboard.

Similar to the CAD test the stimuli used in the CVS are buried in random luminance contrast modulation (or 'luminance contrast noise'). In the CVS the level of luminance contrast noise is set at 75%, an increase from the 45% used in the CAD test. The impact of the level of luminance contrast noise upon chromatic detection thresholds has been previously described in section 2.3.5.2. The colour signal strength of the CVS stimuli are fixed throughout the

test at a fraction of the signal strength associated with the normal upper CAD threshold for the subject's age. This contrasts with the staircase technique employed in the CAD test to determine the signal strength subjects require in order to be able to correctly identify 71% of the stimuli. The signal strength employed in the CVS is referred to in terms of the factor of standard deviations above the mean value for the subject's age. This means that if the CVS was using a signal strength equal to the normal upper limit for a subject's age, the signal strength would be equivalent to $+2.5\sigma$.

As the signal strength is proportional to normal 'age-adjusted' upper limit, two participants of different ages will be shown stimuli with different chromatic signal strengths. The variation in chromatic signal strength as a function of a subject's age is an essential component in any method of colour vision assessment that wishes to account for changes to colour vision as a result of 'normal healthy aging'.

Four narrow colour bands are used to isolate either the RG or the YB chromatic mechanisms. During its presentation the chromaticity of each CVS stimuli changes continuously within a given band (as shown graphically in Figure 4.1B). The preliminary version of the CVS (v2.5.9) employs 36 presentations of RG and YB isolating colours and 18 suprathreshold stimuli defined by both luminance and colour contrast for a total of 90 stimulus presentations. The suprathreshold stimuli are designed to be detected by almost all subjects, including those with congenital colour deficiency, and they are included to provide a measure of the subject's response reliability and identify malingering. In later versions of the CVS (v2.5.14, v2.6.1, and v2.8 used from section 4.4 onwards) the number of RG and YB stimuli were increased to 44 and the number of suprathreshold presentations increased to 22 for 110 total stimulus presentations (more information regarding the changes made to the preliminary version of the CVS can be found in section 4.4).

A short 'learning mode' is completed by all subjects prior to the main element of the CVS. The stimuli in the learning mode are defined by both luminance and colour contrast and, although the responses are not recorded, all stimuli shown in the learning mode need to be correctly identified before the main

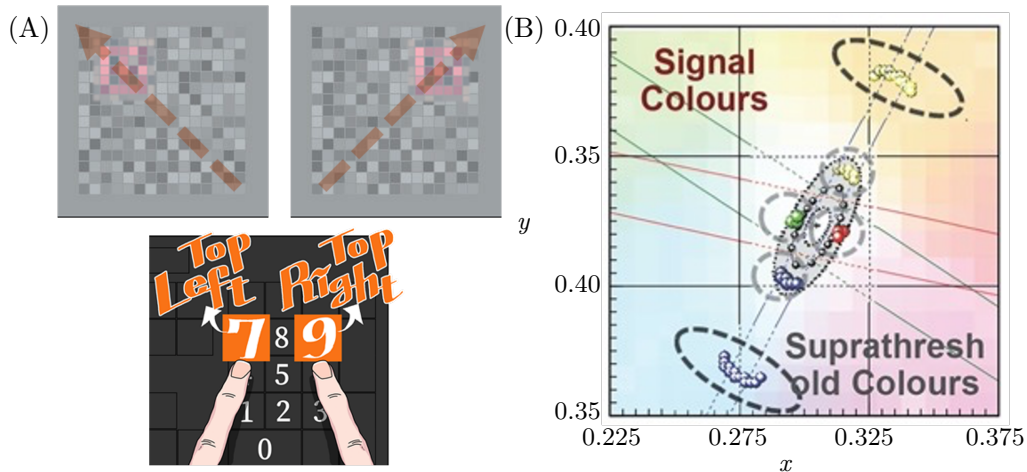


Figure 4.1: (A) An illustration of the motion directions associated with the subject’s task during the CVS and (B) the colours employed in the CVS test. During each presentation the chromaticity of the stimuli rotates smoothly over a range of angular directions selected to isolate the RG and YB chromatic mechanisms (adapted from Barbur et al. (2020)).

element of the test can be started. The learning mode is used to identify individuals with extremely poor vision who are unable to complete the CVS procedure. It also ensures that those who complete the CVS test have the ability to see motion and understand the test procedure.

The CVS test employs a simple two alternative forced choice (2AFC) procedure relative to the 2-down, 1-up, 4AFC used in the CAD test. The movement of the stimulus is the same as in the CAD test, although the directions of motion are reduced, from 4 to 2 possible directions, to match the transition from 4AFC to 2AFC. The input for the 2AFC procedure, the two keyboard buttons ‘7’ and ‘9’ used to identify the direction in which the colour-defined stimulus is moving, are shown in Figure 4.1A. The requirement for only two input buttons in a 2AFC task reduces the complexity and increases the ease of use of the test. This modification is particularly important if the test procedure is expected to be completed on private computers by members of the general public.

A two-alternative forced choice (2AFC) procedure can be used to establish both difference thresholds (or ‘just noticeable difference’) and absolute thresholds (Ulrich & Miller 2004). In the case of the CVS the chromatic contrast of

A) RG and YB			
No. correct	%	CVS classification	Confidence
36 - 34	100% - 94%	Normal RG/YB colour vision	High
33 - 29	92% - 81%		Good
28 - 25	78% - 69%		Poor
24 - 13	67% - 36%	RG/YB deficiency likely	High
12 - 0	33% - 0%		Poor

B) Suprathreshold			
No. correct	%	No. of errors	Response reliability
18	100%	0	Excellent
17	94%	1	Good
16 - 15	89% - 83%	2 - 3	Poor
14 - 0	78% - 0%	≥ 4	Unusable

Table 4.1: The preliminary CVS classification and confidence ratings as a function of A) the number of correct RG and YB responses and B) the number of suprathreshold stimuli correctly identified.

the stimulus is predetermined and a 2AFC procedure is used to establish if observers' chromatic thresholds are above or below the predetermined level.

A subject's performance on the CVS is determined by the proportion of correct responses, ranging from 0, in the event that all responses are incorrect, to 0.5, the chance probability, and 1.0 in the case that all the stimuli are identified correctly. This applies for all three outcome measures (RG, YB, and suprathreshold stimuli). In the preliminary version of the CVS the detection threshold (or the 'pass/fail limit') for RG and YB stimuli was set to a value of 0.69 to align with the 0.71 probability of correct response for a subject's CAD threshold. In later versions of the CVS, including v2.5.14, v2.6.1, and v2.8 used from section 4.4 onwards, the detection threshold is defined as a 0.75 probability of correct response for RG and YB stimuli.

In addition to the percentage of RG and YB stimuli correctly identified, the CVS also displays a 'classification' and 'confidence rating'. The percentage of suprathreshold stimuli correctly identified provide a measure of a subject's response reliability. The classification and confidence intervals for the preliminary version of the CVS are shown in Table 4.1.

During the CVS there are only two possible outcomes for each trial, the probability of success for each trial is constant (when the RG, YB and suprathreshold stimuli are considered separately), and all trials are independent of each other. Consequently, the variability in the outcome of the CVS test can be modelled using the binomial probability density function

$$f(x|n, p) = \left(\frac{n!}{x!(n-x)!} \right) p^x (1-p)^{n-x} \quad (4.1)$$

where x is the number of successes in n trials with a probability, p , of success in each trial.

In the preliminary version of the CVS, subjects need to achieve a minimum of 25 correct responses (out of 36) to pass with a poor confidence rating and be diagnosed as having normal RG or YB colour vision (Table 4.1A). For individuals with severe colour vision deficiency the expected probability of correct response is 0.5 and the probability of such individuals passing the CVS can be calculated using the binomial cumulative probability function. When 36 stimuli are used the probability of a subject with severe CVD passing is approximately 0.0144 (or 1.44%). Or in other words, approximately 1 in 69 subjects with severe colour vision deficiency, assuming they have 50% probability of responding correctly, are expected to correctly identify 25 or more stimuli in the preliminary version of the CVS. The probability of such a subject passing with a ‘good’ confidence interval is approximately 1.56×10^{-4} or 1 in 6399 subjects.

4.3 Preliminary study

Although the preliminary version of the CVS provides the option to change the signal strength, when released, the final versions of the CVS (both the downloadable and drop-down CAD test menu version) will use a fixed signal strength. Before the CVS can be validated and its outcome measured, all the test parameters, including the signal strength, must be determined. A preliminary study was designed and carried out at City University of London

with the principle aim of establishing the minimum CVS signal strength that would ensure that all normal trichromats pass the test.

4.3.1 Methods

Data collection was carried out at City, University of London over four and a half months with the assistance of a final year undergraduate Optometry student. Sixty-five participants were recruited via local advertisements. The inclusion criteria required that all participants were between the age of 18 and 75, had no history of ocular disease, had normal trichromatic colour vision as measured by the CAD test, and were healthy with no known conditions or medication that would affect their colour vision. All participants wore any refractive correction that they habitually used for tasks at the working distance for each test.

All subjects completed a CAD test followed multiple CVS tests. The CAD and the CVS tests were carried out at a viewing distance of 1.4 and 0.8 metres respectively. Each CVS test was carried out used a predetermined signal strength calculated relative to the mean CAD threshold for each subject's age. The predetermined signal strengths were $+1.5\sigma$, $+2.0\sigma$, $+2.5\sigma$, and $+3.0\sigma$ and the order of the first four runs was randomly varied between ascending and descending signal strength. The RG and YB CVS results were analysed to establish the percentage of normal trichromats who pass CVS and are diagnosed as having normal RG and YB chromatic sensitivity, at each signal strength.

Both the CAD and the CVS tests were carried out using a calibrated Eizo ColourEdge CS2420 monitor connected to a HP ProBook 650 G1 laptop running Windows 7 (64 bit with Service Pack 1). Version 2.5.10 of the CAD test was used with the preliminary version of the CVS. All the tests used were approved by the City University Research and Ethics Committee (ETH1819-0206). The study was conducted in compliance with the City, University of London research and ethical guidelines and followed the tenets of the Declaration of Helsinki.

4.3.2 Results

One participant was found to have an acquired colour vision deficiency and was excluded from the study. The remaining 64 participants were aged between 18 and 49 years old, with a median age [and interquartile range] of 19 [19 - 21] years (Figure 4.2). Sixty-four per cent of participants were female (N=41). The average time taken to complete the CVS was 123 ± 16 seconds (or 2 minutes and 3 seconds ± 16 seconds). The distribution of RG and YB CAD thresholds, and the age distribution, for all participants are shown in Figure 4.2.

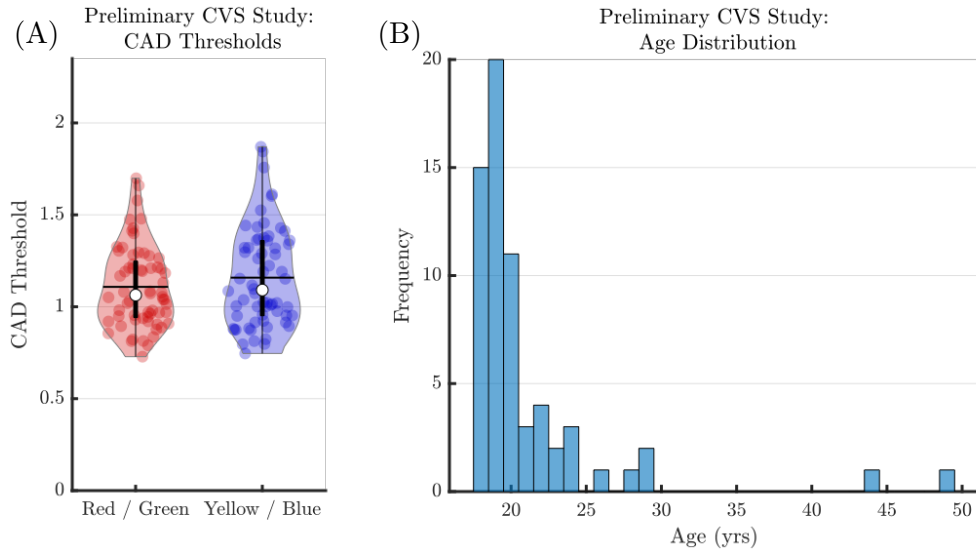


Figure 4.2: (A) The range of RG and YB CAD thresholds and (B) the age frequency distribution of 64 normal trichromats tested in preliminary CVS study.

The CVS results obtained at each signal strength are shown in Figure 4.3. All participants passed the preliminary version of the CVS when signal strengths equivalent to $+3.0\sigma$, $+2.5\sigma$ and $+2.0\sigma$ were used. One participant failed the YB portion of the CVS when a signal strength of $+1.5\sigma$ was used. This participant identified 100% of the suprathreshold stimuli during the same CVS test. Out of the 256 CVS tests carried out almost all ($\sim 98\%$) had a response reliability of ‘Excellent’ ($\sim 87\%$) or ‘Good’ ($\sim 11\%$). Five tests ($\sim 2\%$) had a ‘Poor’ response reliability and no tests obtained an ‘Unusable’ response reliability.

When a signal strength equivalent to $+3.0\sigma$ was used two participants ($\sim 3\%$) passed with a ‘poor’ confidence rating for either the YB or RG stimuli. At a signal strength equivalent to $+2.5\sigma$, one participant passed with a ‘poor’ confidence rating for the YB stimuli (although this was a not one of the two participants that obtained a ‘poor’ confidence rating for a signal strength of $+3.0\sigma$). At a signal strength equivalent to $+2.0\sigma$, all participants correctly identified at least 80% of the RG and YB stimuli.

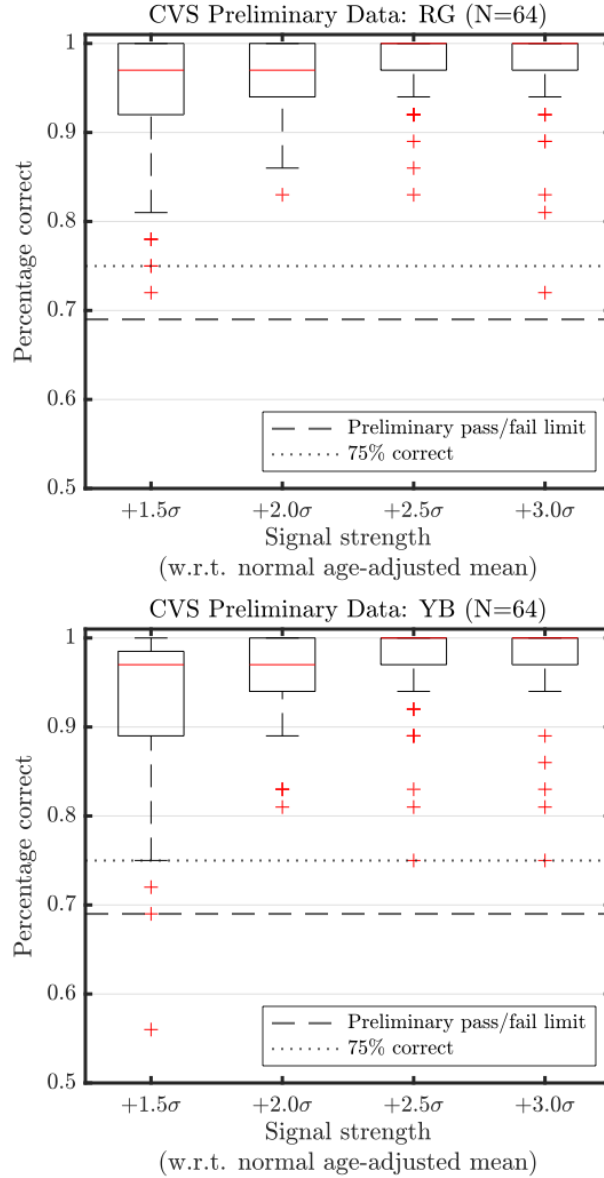


Figure 4.3: Preliminary CVS results showing the relationship between the RG and YB CVS signal strength and the proportion of RG and YB stimuli identified correctly by 64 normal trichromats. The signal strength is presented relative to the normal age-adjusted mean value, where the normal upper age-adjusted limit is equivalent to a signal strength of $+2.5\sigma$.

4.3.3 Discussion

The participants in the preliminary study were mostly young adults with 97% of participants between the age of 18 and 30. As shown in Figure 4.3 the median RG and YB CVS result decreased, along with the 25th percentile, as the signal strength reduced from $+3.0\sigma$ to $+1.5\sigma$. This reduction is expected for a decrease in signal strength and despite the general reduction in performance, all participants passed the RG section of the CVS at all signal strengths.

At a signal strength of $+1.5\sigma$ one participant failed the YB portion of the CVS. This participant completed the CVS measurements in descending order (finishing with the $+1.5\sigma$ measurement) and during their last test they correctly identified 100% of the suprathreshold stimuli with an ‘Excellent’ response reliability. The participant who failed the YB element of the CVS at a signal strength of $+1.5\sigma$ had a YB CAD threshold of 1.87 CAD units. This CAD threshold was the largest YB CAD threshold measured across all study participants and the result closest to the upper normal limit for a participant’s age (the age-adjusted normal upper YB CAD threshold limit for this participant was 1.89 CAD units). Despite being virtually at the upper normal limit for their age this participant passed the YB element of the CVS when a signal strength of $+2.0\sigma$ was used and correctly identified 81% of the stimuli. This indicates that even for individuals who have CAD test results at or exceptionally close to the normal upper limit for their age, a signal strength of $+2.0\sigma$ is sufficient to ensure a pass on the CVS.

Only one participant identified less than 75% of the RG or YB CVS stimuli when a signal strength $> +2\sigma$ was employed (72% correct at $+3.0\sigma$). Interestingly, the participant’s RG CVS result improved when the signal strength was reduced, correctly identifying 75%, 89%, and 92% of the RG stimuli at signal strengths of $+1.5\sigma$, $+2.0\sigma$, and $+2.5\sigma$, respectively. The low result at $+3.0\sigma$ may be attributable to variability in the participant’s response or lack of familiarity with the RG test stimuli. During the same CVS test the participant correctly identified 94% of YB stimuli and 100% of the suprathreshold stimuli suggesting that the participant was attentive. This participant completed

the CVS measurements in descending order, starting with a signal strength of $+3.0\sigma$, and obtained this RG result during their first CVS test. A lack of familiarity with the RG CVS stimuli could be addressed by adding an option to repeat the learning mode and/or by including on-screen instructions and images showing the CVS stimuli prior to the test being carried out.

A limitation of the preliminary study protocol is that it did not attempt to measure the sensitivity of the CVS in individuals with CVD. However doing so was beyond the scope of the preliminary study and the sensitivity of the CVS is investigated in later studies (see section 4.5). The small sample size ($N=64$) is also a limitation of the study. However, one must consider the trade-off between time spent collecting data for a preliminary study and the time spent collecting data for the main study. For any project with a fixed duration allocating more time and resources for a preliminary study will result in fewer resources for the main study or studies.

The preliminary results indicate that normal trichromats will pass the CVS with a ‘good’ confidence rating when a signal strength equivalent to the mean CAD threshold for a given age $+2.0\sigma$ is used. The average time taken to complete the CVS (123 ± 16 seconds) was comparable to the reported average time taken to complete the Ishihara pseudoisochromatic plate test (~ 138 seconds) (Murphy 2015). Therefore, it may be possible to improve the accuracy of the CVS by adding additional stimuli, whilst keeping the testing time comparable to the Ishihara pseudoisochromatic plate test.

4.4 Modifications made to the CVS following the preliminary study

Several changes were made to the CVS (v2.5.14 onwards) following analysis of the preliminary study data and informal feedback obtained from participants during the data collection. The primary changes were an increase in the number of presentations in the CVS and the default use of a fixed signal

strength just below the upper normal age-adjusted limited (equivalent to the mean CAD threshold for the age of the subject $+2.0\sigma$).

The number of RG and YB stimuli increased from 36 to 44 presentations and the number of suprathreshold presentations increased from 18 to 22. Additionally, the detection threshold, or pass/fail limit, was increased from 0.69 (25/34) to 0.75 (33/44). This change brings the CVS detection threshold in line with historical conventions for a 2AFC procedure (McKee et al. 1985). The increased number of presentations increases the accuracy and the statistical confidence of results obtained from the CVS. The confidence ratings were also modified following the changes to the number of presentations and the detection threshold (as shown in Table 4.2).

The probability of an individual with a 50% probability of responding correctly passing the current version of the CVS (v2.8) can be determined using the binomial probability function as being approximately 6.3×10^{-4} or 1 in 1587.

A ‘demonstration mode’ consisting of a series of detailed of on-screen instructions was added to all versions of the CVS. The introduction of a demonstration mode was implemented following informal feedback from a small number of participants in the preliminary study, who stated that they would benefit from on screen instructions. The graphical elements in the demonstration mode were created following a collaboration with Joshibi University, Tokyo. The learning mode was reduced to compensate for the increased test time and as a result of informal feedback from several participants who suggested that the learning mode was too time consuming. The learning mode was also modified to include the option to repeat the learning procedure, ensuring that any subjects who require additional ‘learning’ time have the option to do so. Version 2.5.14 of the CVS test was produced by City Occupational Limited following the completion of the preliminary study. The CVS version accessible via a menu function in the CAD test is referred to by the CAD test version number from v2.5.14 onwards. A freely downloadable version of the CVS was made available online for Windows computers connected to displays that support the sRGB colour standard. While the downloadable version of the

CVS is not described or used in the studies contained within this thesis, it is briefly touched upon in section 6.3 ‘Future directions’.

A) RG and YB			
No. correct	%	CVS classification	Confidence
44 - 41	100% - 93%	Normal RG/YB colour vision	High
40 - 33	91% - 75%		Good
32 - 27	73% - 61%	RG/YB deficiency likely	Good
26 - 0	59% - 0%		High
B) Suprathreshold			
No. correct	%	No. of errors	Response reliability
22	100%	0	Excellent
21	95%	1	Good
20 - 19	91% - 86%	2 - 3	Poor
18 - 0	82% - 0%	≥ 4	Unusable

Table 4.2: The final CVS classification and confidence ratings as a function of A) the number of correct RG and YB responses and B) the number of suprathreshold stimuli correctly identified.

Note: these confidence intervals were added to the CVS software in v2.8.

4.5 Main study

Several experiments were carried out to evaluate the CVS test following the completion of the preliminary study and the determination of the test parameters. Each experiment used the same calibrated Eizo ColourEdge CS2420 monitor connected to a HP ProBook 650 G1 laptop running Windows 7 (64 bit with Service Pack 1). As previously described in section 3.2, the Eizo CS2420 monitor was calibrated for luminance and the chromaticity of each primary colour using a Konica CS-2000A telespectroradiometer (manufactured by Konica Minolta Inc.) with a spectral range of 380 to 780 nm. The spectrum was sampled every nm with a bandwidth of just under 1 nm.

“Evaluation of a new colour test requires knowledge of its reliability and its validity”

(National Research Council 1981, Chapter 3, page 18)

The National Research Council (1981) define the reliability of a test as “whether the test measures the same property on each occasion” and the validity as “whether the test measures what it claims to measure”.

The main CVS study consists of three separate but linked experiments to:

- I. investigate the performance of the CVS in normal trichromats and least affected deuterans and protans
- II. determine the reliability of the CVS
- III. determine the validity of the CVS

Experiments I and II were carried out at City, University of London. Experiment III was carried out at City, University of London and at a number of external CAD centres. The description of the experimental work and data collected at City, University of London is described in section 4.5.3, and the multi-centre component is described in chapter 5. The study followed the tenets of the Declaration of Helsinki and, as with the preliminary study, all the tests used were approved by the City University Research and Ethics Committee (ETH1819-0206).

4.5.1 Experiment 1 - the performance of the CVS for normal trichromats, least-affected deuterans, and least-affected protans

Before investigating the validity and reliability of the CVS it is of considerable interest to establish an estimate for the sensitivity of the CVS. The maximum sensitivity of the CVS test will be determined by two factors: the probability of a correct response in the least affected (most sensitive) individuals with CVD and the prevalence of these least affected individuals in the CVD population.

As shown in Figure 4.4, and described in chapters 2, 3 and Barbur & Rodriguez-Carmona (2017, 2015), the severity of loss in congenital CVD can range from just above the normal age-adjusted upper limit to a complete absence of

colour vision. It is important to appreciate that the age-adjusted upper limits are a result of both the inter-subject variability (the variability between subjects) and the within-subject variability (the variability in a single subject's results). The within-subject variability (standard deviation observed in repeated CAD measurements in the same subject) has been reported to vary between approximately 5-15% of the subject's mean threshold, compared to the inter-subject variability in normal trichromats with standard deviations between approximately 21-25% of the subject's mean threshold (Barbur et al. 2020).

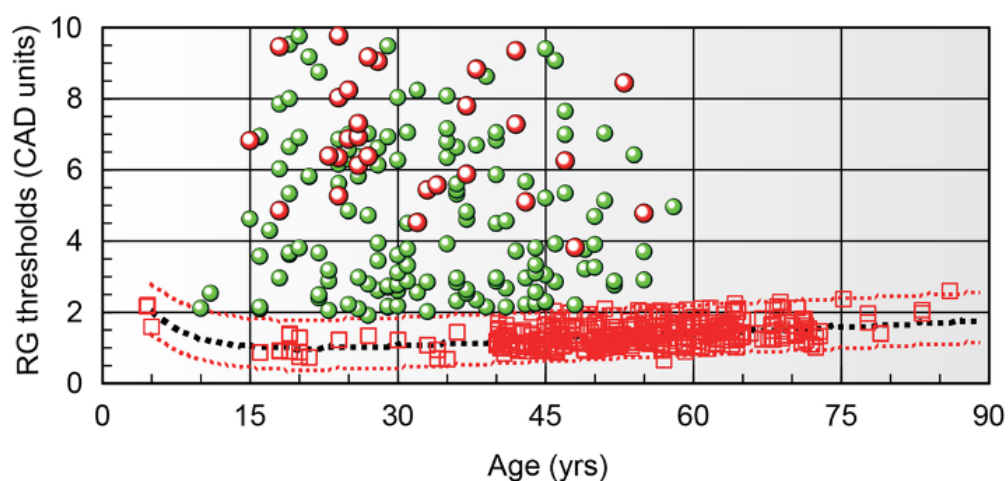


Figure 4.4: Changes in RG CAD thresholds as a function of age in normal trichromats (red squares), deuterans (green circles), and protans (red circles). The range of ‘normal’ RG CAD thresholds associated with each age are shown by the red dotted lines. A similar trend is observed for the change in the ‘normal’ YB thresholds as a function of age (Barbur & Rodriguez-Carmona 2015) (adapted from Barbur et al. (2020)).

A significant consequence of the measured within- and inter-subject CAD test variability is that RG CAD thresholds, measured in normal trichromats and least affected deuterans, which are close to the normal upper limit do not represent the mean CAD thresholds for these individuals. Rather, these measured CAD thresholds are likely obtained from individuals who have been observed at the limit of their within-subject variability. As such, these thresholds are equivalent to the mean thresholds plus/minus the within subject variability for normal trichromats and the least affected deuterans respectively.

The percentage of stimuli correctly identified in the CVS is related to the probability of a correct response by a simple chance probability correction

$$p_{corr} = \frac{S_{CVS} - p_{chance}}{1 - p_{chance}} \quad (4.2)$$

where p_{corr} is the probability of a correct response, S_{CVS} is the percentage of RG or YB stimuli correctly identified, and p_{chance} is the chance probability. Such a correction has previously been described by Blackwell (1952). In the CVS the chance probability is 0.5 and equation 4.2 simplifies to

$$p_{corr} = 2(S_{CVS} - 0.5) \quad (4.3)$$

RG and YB CVS responses are expected to follow a binomial distribution. While a probability of 0.5 corresponds to individuals who are unable to identify any of the RG or YB stimuli many individuals with CVD have some residual chromatic sensitivity. The highest probability of a correct response will be found in the least affected individuals with CVD.

The aim of this experiment was to establish the probability of a correct response for normal trichromats and least affected deuterans and protans as a function of the colour signal strength employed in the CVS test.

4.5.1.1 Methods

Six participants were recruited via local advertisements and the AVOT service at City, University of London. These participants included two normal trichromats, three least affected deuterans, and one least affected protan. The class and severity of participant's colour vision were determined using the CAD test. All participants were between the age of 19 and 41 and were healthy with no known conditions or medication that would affect their vision or colour vision. All participants wore any refractive correction that they habitually used for tasks at the working distance for each test.

Each participant completed v2.5.14 of the CAD and several CVS tests at varying colour signal strengths. The CVS data were corrected for the 0.5

chance probability using of a correct response using equation 4.3 and fitted to a Weibull function of the form

$$F_{Weibull}(x) = 1 - \exp\left(-\left[\frac{x}{a}\right]^b\right) \quad (4.4)$$

where x is the colour signal strength and a and b are constants typically associated with function's location and slope, or gradient, of the curve respectively (Prins 2012). The curves were fitted using the nonlinear least squares method and the Levenberg-Marquardt algorithm in the curve fitting toolbox (3.5.11) in Matlab version 2020a (MATLAB 2020). The fitted Weibull functions were used to estimate the probability of a correct response at the signal strength employed in the CVS ($+2\sigma$).

4.5.1.2 Results

One normal trichromat was female and all of the other participants were male. For the CVS measurements the response reliability was rated as 'Excellent' or 'Good' with all but one of the suprathreshold values being 100% (the other value was 94%). The RG CAD thresholds were 0.88 and 1.12 CAD units for the two normal trichromats, 2.13, 2.74, and 2.90 CAD units for the three least affected deuterans, and 6.55 CAD units for the least affected protan.

The Weibull curves fitted to the adjusted RG and YB CVS results for the normal trichromats and the least affected deuterans and protans are shown in Figure 4.5. A direct comparison between the probability of a correct RG response for a normal trichromat, a least affected deutan, and a least affected protan is shown in Figure 4.6. The goodness of fit statistics for the Weibull curves, including the sum of squares due to error (SSE), R-square, degrees of freedom adjusted R-square, and the root mean square error (RMSE), are shown in Table 4.4.

The probability of a correct response at the signal strength used in the CVS, based upon the Weibull fitted curves ranged from 1.0000 to 0.9947 for the normal trichromats, 0.3507 to 0.2538 for the least affected deuterans, and 0.0177 for the least affected protan. The corresponding theoretical probabilities of the

No.	CAD		Values from Weibull Fit			
	Classification	RG Threshold	p_{corr}	S_{RG}	$P \geq 33$	$P \leq 32$
1	Normal	0.88	1.0000	1.0000	1.0000	0.0000
2	Normal	1.12	0.9947	0.9973	1.0000	0.0000
3	Deutan	2.13	0.3507	0.6754	0.1861	0.8139
4	Deutan	2.74	0.3259	0.6630	0.1434	0.8566
5	Deutan	2.90	0.2538	0.6269	0.0598	0.9402
6	Protan	6.55	0.0177	0.5089	0.0009	0.9991

Table 4.3: RG CAD test results, the probability of a correct RG CVS response, and the predicted RG CVS outcome for the six participants tested in experiment 1.

No.	Fit	SSE	R-square	Adjusted R-square	RMSE
1	RG	0.1738	0.9078	0.8994	0.1257
2	RG	0.0036	0.9909	0.9891	0.0268
3	RG	0.1316	0.9120	0.9046	0.1047
4	RG	0.3612	0.8068	0.7919	0.1667
5	RG	0.2983	0.8085	0.7894	0.1727
6	RG	0.3638	0.8023	0.7892	0.1557
1	YB	0.1457	0.8614	0.8488	0.1151
2	YB	0.0149	0.9630	0.9556	0.0546
3	YB	0.0146	0.9863	0.9851	0.0349
4	YB	0.1219	0.8478	0.8361	0.0968
5	YB	0.0249	0.9658	0.9624	0.0499
6	YB	0.0016	0.9941	0.9938	0.0103

Table 4.4: Goodness of fit statistics for the Weibull functions fitted to CVS data collected in experiment 1.

least affected deutan and protan passing the CVS, based upon the probability of a correct response predicted by the Weibull fitted data, are 0.1861 and 0.0009 (18.61% and 0.09%) respectively.

The probability of a correct response for the signal strength used in the CVS, p_{corr} , the corresponding CVS outcome (% of stimuli correctly identified), S_{RG} , and the probability of passing and failing the CVS based upon the fitted Weibull function are shown for all participants in Table 4.3.

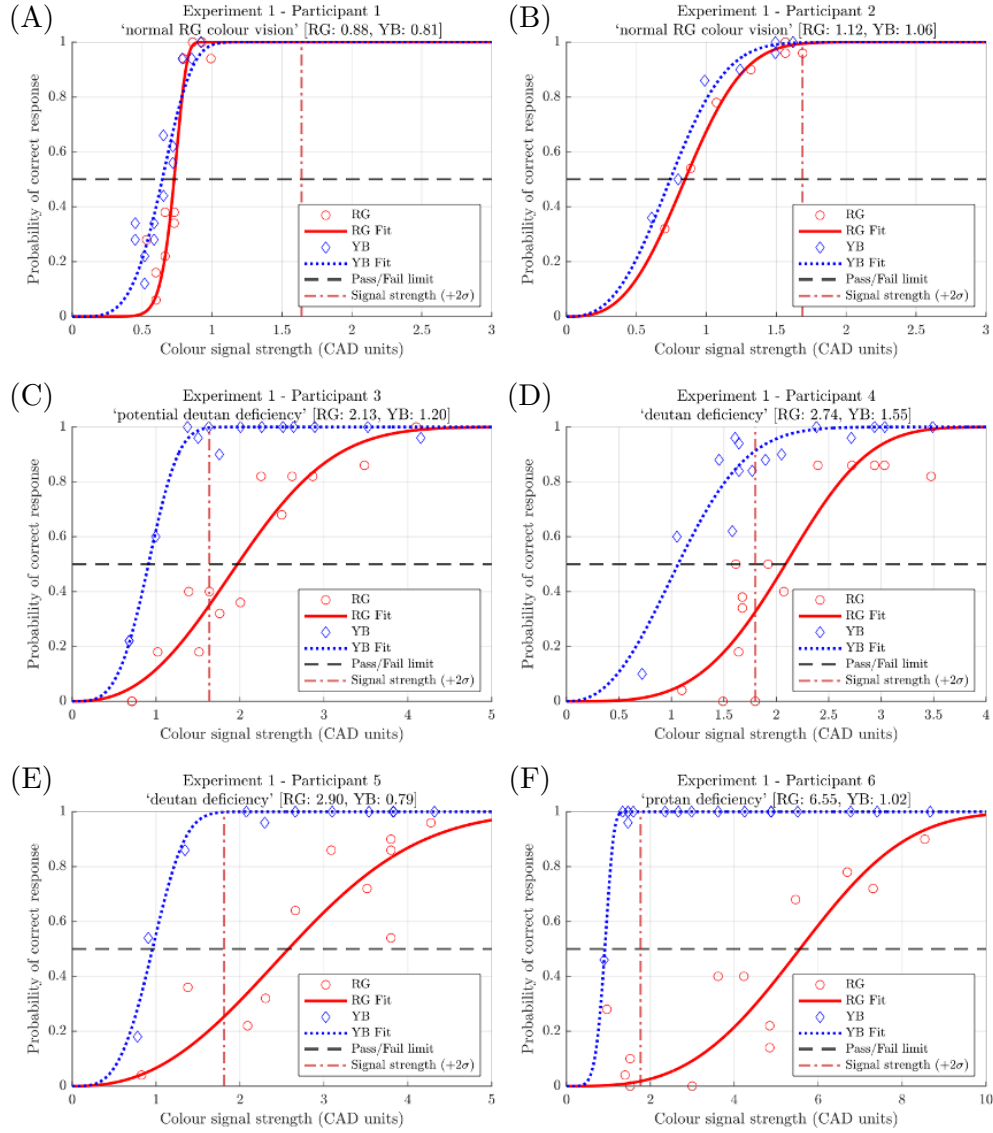


Figure 4.5: The Weibull function fitted to CVS data measured in two normal trichromats (A) and (B), three least affected deutans (C), (D), and (E), and a least affected protan (F). The CVS data shown has been converted from the percentage of RG/YB stimuli correctly identified to the probability of correct response using Equation 4.3. The dashed horizontal line in each plot shows the probability of correct response required to pass the CVS test (0.5 corresponding to correctly identifying 75% of the stimuli) and the vertical line represents the $+2\sigma$ signal strength used in the CVS. The signal employed in the CVS is slightly different for each participant (participants ranged from 19-41 years of age).

4.5.1.3 Discussion

All participants had normal YB colour vision as measured by the CAD test. The two normal trichromats had RG CAD thresholds of 0.88 and 1.12 CAD units, as shown in Table 4.3. Neither participant had a RG CAD thresh-

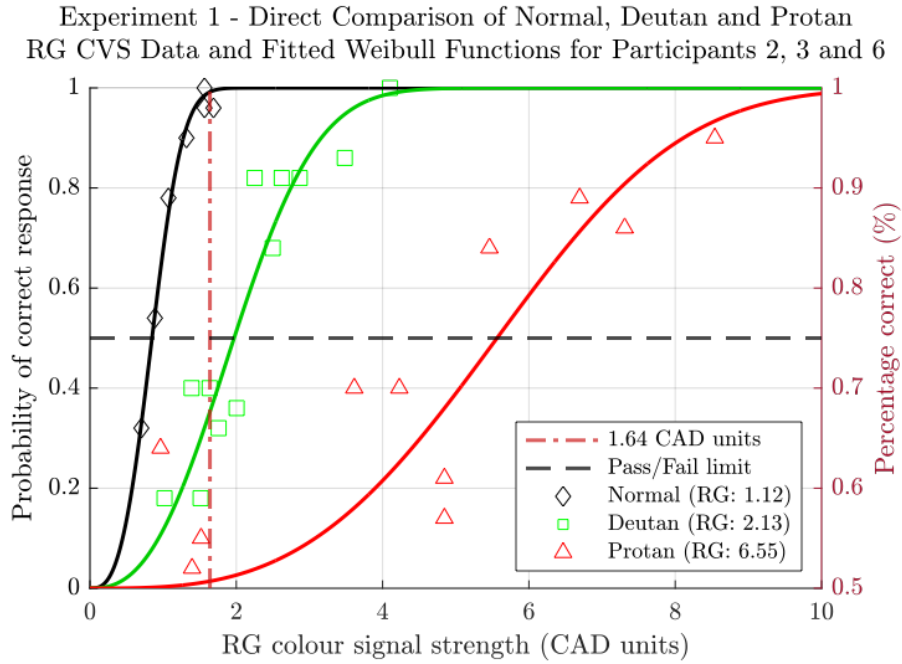


Figure 4.6: A direct comparison of the fitted Weibull function in a normal trichromat, a least affected deutan, and a least affected protan. The vertical dash-dotted line at 1.64 CAD units represents the signal strength used in the CVS for the least affected deutan. For clarity only the signal strength used for the least affected deutan is shown. The signal strengths used in the CVS is determined by the age of each participant (1.69 and 1.76 CAD units for the normal trichromat and least affected protan respectively).

old close to the normal upper limit for their age (1.79 and 1.85 CAD units respectively).

The three least affected deutan participants had RG CAD thresholds (and a normal upper age-adjusted limit) of 2.13 (1.79), 2.74 (1.96), and 2.90 (1.98) CAD units. The least affected protan participant had a RG CAD threshold of 6.55 and the normal upper RG CAD threshold for their age was 1.93 CAD units. In the deutan ($N=261$) and protan ($N=109$) populations described in section 3.2, $\sim 0.77\%$ of deutans had RG thresholds less than or equal to 2.13 CAD units, $\sim 6.51\%$ of deutans had RG thresholds less than or equal to 2.90 CAD units, and $\sim 8.26\%$ of protans had thresholds less than 6.55 CAD units.

The degrees of freedom adjusted R-square values for all the RG Weibull fits were relatively high, particularly for the two normal trichromats (0.8994 and 0.9891 for participants 1 and 2 respectively) and the least affected deutan (0.9046), with values ranging from 0.9891 to 0.7892. These, combined with

the other goodness of fit statistics shown in Table 4.4, indicate that Weibull fitted curves had a good quality of fit to the CVS experimental data.

The direct comparison of the fitted RG CVS Weibull functions for a normal trichromat, least affected deutan, and least affected protan shown in Figure 4.6 highlights the separation between the probability of a correct response in normal trichromats and least affected deutan and protans.

The probability of a correct response, and corresponding percentage of RG or YB stimuli correctly identified, can be used with the binomial cumulative distribution function to determine the probability of passing the CVS (correctly identifying over 32 of the 44 presentations). For both normal trichromats the predicted probability of a correct response at the signal strength employed in the CVS (1.0000 and 0.9947) and subsequent probability of passing the RG component of the CVS test (100% for both participants) were as expected, particularly as neither normal trichromat had a RG CAD threshold close to their upper age-adjusted limit.

The least affected deutan (RG CAD: 2.13 CAD units) had the highest probability of a correct response of all the CVD participants, 0.3507, for the signal strength employed in the CVS. This probability of a correct response corresponds to a 18.61% probability of passing the CVS. The two other deutan (RG CAD: 2.74 and 2.90 CAD units) had a probability of correct response of 0.3259 and 0.2538 corresponding to a 14.34% and 5.98% chance of passing the CVS respectively. The least affected protan (RG CAD: 6.55 CAD units) had a probability of a correct response at the signal strength employed in the CVS of 0.0177 corresponding to a 0.09% probability of passing the RG component of the CVS.

These results indicate that between 18.6% – 6.0% (estimated to make up between $\sim 0.8\%$ and $\sim 6.5\%$ of the deutan population) of the least affected deutan are likely to pass the CVS. Consequently only $\sim 0.53\%$ ($\sim 0.14\% + \sim 0.39\%$) of deutan are expected to pass the RG component of the CVS. A smaller percentage of the least affected protans, 0.09% (estimated to make up $\sim 8.26\%$ of the protan population) are likely to pass the CVS. Subsequently,

one would expect less than 0.008% of all protans to pass the RG component of the CVS.

One can compare these predictions to the observed outcome of the Ishihara test from section 3.2. The percentage of deutans expected to pass the CVS, $\sim 0.53\%$, is similar to the percentage of deutans who pass the most sensitive Ishihara protocol, $\sim 0.38\%$ (Table 3.1). The percentage of protans expected to pass the CVS, $\sim 0.008\%$, is far lower than the most sensitive Ishihara protocol, $\sim 0.917\%$ (Table 3.1).

A limitation of the methodology used to estimate the percentage of deutans and protans expected to pass the CVS is that it does not account for the percentage of deutans and protans with thresholds above 2.90 and 6.55 CAD units respectively, who have greater than a 0.5 probability of a correct response. The inclusion of these individuals is, however, unlikely to significantly increase the estimation for the percentage of deutans and protans who are likely to pass the CVS.

There are two possible expansions of this work and analysis that would directly address this limitation. The first expansion is experimental and would involve repeating these measurements in participants who have RG CAD thresholds ranging from 3 to 10 CAD units in deutans and from 5 to 10 CAD units in protans, preferably in ~ 0.5 CAD unit intervals. The same discrete interval analysis could then be carried out in combination with existing deutan and protan population statistics. The second approach would be to fit a sigmoid function of the type

$$z(U_{RG}) = 1 - \exp\left(-\left[\frac{a}{U_{RG}}\right]^b\right) \quad (4.5)$$

to the current experimental data, where U_{RG} is CVS signal strength in terms of the fraction of normal upper age-adjusted limit, $z(U_{RG})$ is the probability of passing the CVS as a function of the normal upper age-adjusted RG limit, and a and b are constants. These data could be used in combination with the population distribution functions for deutan and protan populations to estimate the overall percentage of deutans and protans who are expected

to pass the CVS. Mathematically this would be determined by solving the following definite integrals

$$p_{deutan} = \frac{\int_{U_{RGmin}}^{U_{RGmax}} [g(U_{RG})] [z(U_{RG})] dU_{RG}}{\int_{U_{RGmin}}^{U_{RGmax}} g(U_{RG}) dU_{RG}} \quad (4.6)$$

$$p_{protan} = \frac{\int_{U_{RGmin}}^{U_{RGmax}} [h(U_{RG})] [z(U_{RG})] dU_{RG}}{\int_{U_{RGmin}}^{U_{RGmax}} h(U_{RG}) dU_{RG}} \quad (4.7)$$

where p_{deutan} and p_{protan} are the percentage of protans and deutans who are expected to pass the RG component of the CVS, U_{RGmin} and U_{RGmax} are the minimum and maximum fraction of normal upper age-adjusted RG limit for the deutan/protan population, and $g(T_{RG})$ and $h(T_{RG})$ are the relative prevalence of deutans and protans as a function of RG CAD thresholds.

An additional limitation of the experimental data are the RG CAD thresholds of the normal trichromats and the least affected protan. The normal trichromats did not have RG CAD thresholds at, or close to, the normal upper limit for their age and as such the analysis of their results does not provide an indication of the performance of the CVS in the least sensitive normal trichromats. Similarly, the least affected protan is only in the lower $\sim 8.3\%$ of the protan population when ranked according to RG CAD thresholds (compared to the least affected deutan who had a RG CAD threshold in the lowest $\sim 0.8\%$ of all deutans), and does not accurately represent the least affected (most sensitive) protan observer. Both of these limitations are primarily due to the scarcity of individuals who have CAD thresholds at the limits of the within- and inter-subject variability for their respective class of colour vision and the associated challenges with recruiting such participants. Despite not being investigated here using the larger number of stimulus presentations in v2.5.14 the preliminary study indicated that normal trichromats close to their respective normal upper age-adjusted limits will pass the CVS.

An alternative approach to using the Weibull function in isolation would be to use the psychometric function in combination with the Weibull function. Doing so would remove the need to convert the CVS data from the percentage of stimuli correctly identified to the probability of a correct response. This modification would also introduce an additional term, the lapse rate (the probability of an incorrect response which is independent of stimulus intensity), to the fitted function. When using the Weibull function in isolation one assumes a fixed value of zero for the lapse rate which can introduce biases if lapses do occur (Wichmann & Hill 2001, Prins 2012). In this experiment the high response reliability across all CVS tests (all but one of the suprathreshold measurements were 100%) and the relatively short time taken to complete the CVS provide strong evidence to justify and support the assumption of a lapse rate of zero. The introduction of additional terms into the fitted function may also result in overfitting, although overfitting should be detected and accounted for in the degrees of freedom adjusted R-square statistic.

The results of this experiment show that the CVS test can achieve a high sensitivity when screening for congenital deutan and protan deficiencies. The test therefore has the potential to be used as the first stage in a ‘two-step’ protocol to efficiently screen for colour vision deficiencies.

4.5.2 Experiment 2 - the reliability of the CVS

The reliability or the extent to which the result of a test are consistent across multiple measurements is one of the two key test properties which must be determined before a test can be widely adopted (Blackwell 1952, Bland & Altman 1986).

The results obtained from a test with poor reliability will be highly variable. Unsurprisingly a test’s reliability is directly related to its practical utility and adoption. A poor reliability can make it difficult, or in some cases virtually impossible, for those who administer the test to determine whether a measurement is a valid measurement or a consequence of the test’s variability (Bartlett & Frost 2008, Noble et al. 2019).

The aim of this experiment was to investigate and establish the CVS test-retest reliability.

4.5.2.1 Methods

Ten participants were recruited via local advertisements and the AVOT clinic service at City, University of London. Participant's colour vision was quantified and classified by the CAD test. The participants included six normal trichromats, three deutan, and an individual with acquired YB loss. All participants were between the age of 20-44 and were healthy with no known conditions or medication that would affect their vision or colour vision. Data collection was carried out with the assistance of a final year undergraduate Optometry student and all participants wore any refractive correction that they habitually used for tasks at the working distance for each test.

Participants were invited to attend three sessions over the course of 3-4 weeks. During each visit participants completed eight CVS tests and participants also completed a full CAD test during their first and last session. Version 2.5.14 of the CAD and CVS tests were used throughout the experiment and the variation in the CVS test was evaluated to determine the test-retest reliability of the CVS.

4.5.2.2 Results

The average age of the participants tested in this experiment was 26.7 ± 7.0 years. The normal trichromats were 50% male and 50% female, all the deutan participants were male, and the participant with acquired YB loss, as diagnosed by the CAD test, was female. The CAD diagnoses were consistent across both CAD tests (in the first and the last session) for all participants.

Participants diagnosed as having a deutan deficiency using the CAD test had RG CAD thresholds of 7.13 & 6.18, 5.94 & 7.24, and 21.70 & 20.19 at their first & second visits respectively. The participant with acquired YB loss was found to have YB CAD thresholds of 1.95 and 1.82 at their first and final

CAD classification	Number of		CVS pass	
	Participants	CVS tests	R/G (%)	Y/B (%)
Normal	6	144	100.0	99.3
Deutan	3	72	0.0	100.0
Protan	0	0	0.0	0.0
Acquired YB Loss	1	24	100.0	100.0

Table 4.5: A summary of the CVS reliability data collected over 240 measurements.

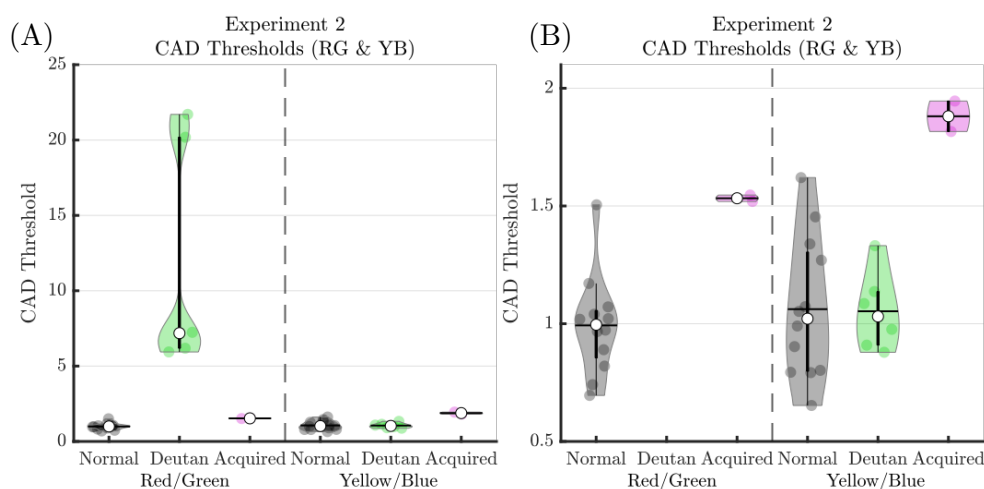


Figure 4.7: Distribution of RG and YB CAD thresholds for participants assessed in experiment 2. The data from the first and last CAD test have been included for all participants and (A) and (B) show the same data with a different vertical scaling.

session respectively. Figure 4.7 shows the distribution of RG and YB CAD thresholds for all participants.

Table 4.5 contains a summary of the CVS results collected across all sessions. Across all 240 CVS tests the suprathreshold outcome was either 100% correct (230/240 measurements $\sim 95.8\%$) or 95.5% correct (10/240 measurements $\sim 4.2\%$) corresponding to an ‘Excellent’ or ‘Good’ response reliability. The RG and YB outcome of the CVS tests completed at each visit are shown in Figure 4.8 and Figure 4.9 respectively. As shown in Figure 4.8B, one normal trichomat failed the YB component of the CVS during the first test of their last session.

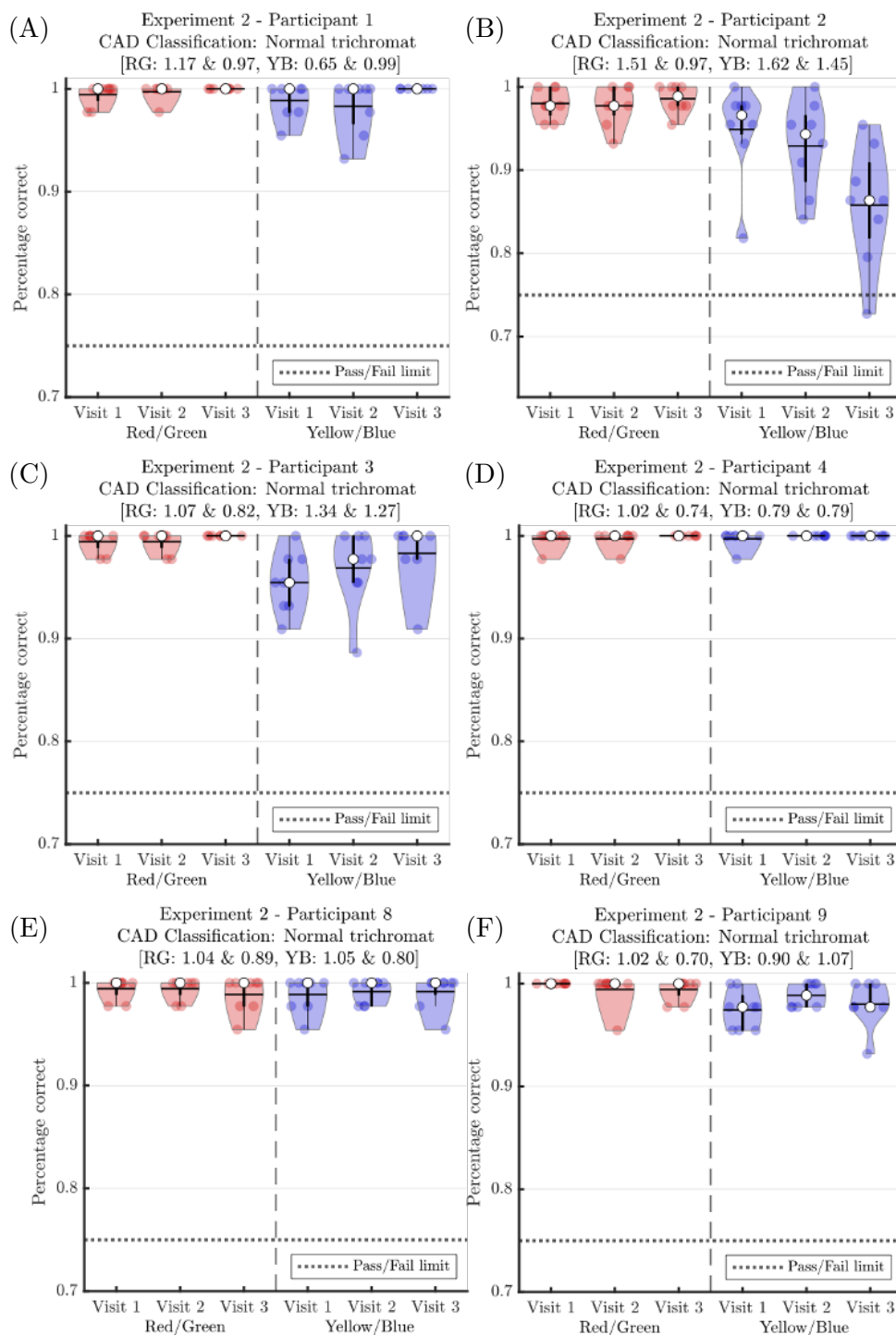


Figure 4.8: The RG and YB outcome from 24 CVS tests completed over three separate sessions, with 8 tests in each session, in six normal trichromats (A-F).

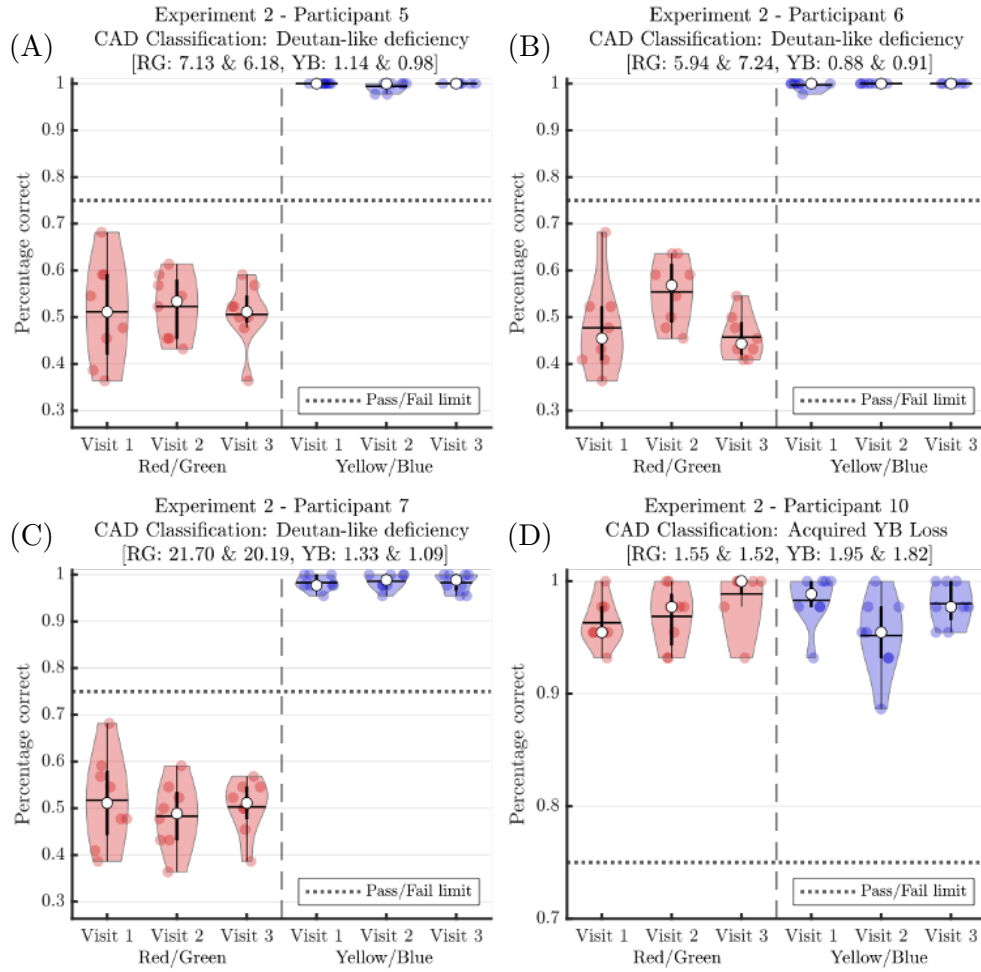


Figure 4.9: RG and YB outcome of 24 CVS tests completed over three separate sessions in three deutan (A-C) and (D) an individual with acquired YB loss, as diagnosed by the CAD test.

4.5.2.3 Discussion

All participants obtained the same CAD diagnosis at both their first and last visit. Only one participant (participant 2) exhibited a notable change in their CAD thresholds between the first and last session. This participant is discussed in more detail below. The six normal trichromats had CAD thresholds within the limits established for their age. Two of the three deutan who took part in the study had RG CAD thresholds below 8 CAD units (7.13 and 5.94 CAD units at their first session) with the other deutan having a higher RG CAD threshold (21.70 CAD units at their first session). The participant with acquired loss had a YB threshold slightly above the normal upper YB limit for their age in the first and last session (YB CAD thresholds: 1.95 and 1.82 compared to the normal upper age-adjusted YB limit of 1.75

CAD units). This participant was unaware of their YB, as determined by the CAD test, loss prior to being recruited.

The CVS results reveal a high test-retest reliability. For the RG diagnosis the CVS was found to have a 100% test-retest reliability in the ten participants across the 240 measurements. The YB diagnosis test-retest repeatability was 100% in deutan participants and $\sim 99.3\%$ in normal trichromats. The approach used in this experiment for determining reliability has previously been used by Mancuso et al. (2006) for another display based test, the CCT, using squirrel monkeys.

The results for the normal trichromat who failed the YB component of the CVS in one of the 24 tests are shown in Figure 4.8B (participant 2). This participant failed the YB component of the CVS during the first test in their third session. Interestingly this normal trichromat is one of only two participants tested who appear to show a trend in their CVS test results across the three sessions. Participant 2 showed a decrease in the mean and median YB CVS test outcome as a function of the session number. The trend is only present in the YB CVS data. The decrease in median YB CVS test outcome is also reflected in the YB CAD thresholds which increased from a threshold of 0.97 in the first session to 1.45 in the final session. The reasons for this reduction in YB sensitivity over the experiment duration are unknown. Unfortunately it was not possible to complete follow up testing at a later date to further investigate and follow up testing would have helped to potentially confirm the presence of an acquired colour vision deficiency that was progressing over time. However, the participant was an optometry student who reported normal general health and no visual problems. As with all participants, they were informed of their results at the end of each session. Additionally, as an optometry student the participant would be completing regular sight tests as part of their studies and they would be aware of the potential causes of acquired colour vision deficiency.

The other participant who showed a trend in their YB CVS outcomes was participant 3 who, as shown in Figure 4.8C, exhibited the opposite effect with a smaller magnitude. Their median and mean YB CVS outcome improved at

each subsequent session. This improvement may be due to a learning effect, although it's unclear as to why the improvement was only observed with the YB stimuli. There was no evidence that the participant had a higher level of concentration from their RG or suprathreshold CVS responses across the three sessions.

The YB CVS results from the participant with acquired YB loss show a 100% test-retest reliability – the results are entirely consistent across all 24 repeated measurements taken on three separate occasions. Despite the disagreement between the CAD diagnosis and the CVS outcome (discussed in more detail during the evaluation of the validity of the CVS in section 4.5.3), the CVS test produced the same outcome in all measurements.

Limitations of this experimental procedure include the relatively low number of participants. This was largely a design choice as the experiment was purposed to maximise the number of CVS tests at multiple visits, rather than the number of participants tested. Future work could address this by recruiting more normal trichromats, protans, individuals with tritanopia, and individuals with both RG and YB acquired loss. From the results of experiments 1 & 2, one would expect the reliability of the test to be similar in deuterans, protans and those with acquired loss, provided that the participants did not have RG or YB CAD thresholds close to the normal upper limit for their age.

Another limitation of the approach in this experiment is that it did not formally consider changes in the CVS outcome within each session. However, as shown in Figures 4.8 and 4.9 the variability within each session was relatively low. Variability in the CVS outcome as a function of measurement number within a specific session would likely be due to either fatigue, or a learning effect. Although fatigue can be difficult to identify and quantify, the high suprathreshold stimuli outcome across all measurements strongly suggests that this was not a significant factor. One would expect a learning effect to be visible for all stimuli - this was not the case and, excluding the small improvement in YB results by participant 3, this was not observed. A learning effect would also be minimised by the learning and demonstration modes employed within the CVS.

The results reveal the high test-retest reliability of the CVS. Despite the high reliability there was a consistent disagreement between the classification made by the CAD and CVS tests for one participant with acquired YB loss. Although this is not an indication of poor reliability, this discrepancy needs to be examined in more detail when evaluating the validity of the CVS.

4.5.3 Experiment 3 - the validity of the CVS

The only remaining property of the CVS that has yet to be investigated is the validity of the test. The validity was investigated both internally at City, University of London and through a multi-centre study using data collected at CAD testing centres located around the world. This chapter contain a description of the work carried out at City, University of London. The multi-centre study (including data collected at City, University of London) is described in chapter 5.

Given the aim of the CVS, to screen for congenital and acquired loss of colour vision, its validity can be determined by evaluating the outcome of the CVS relative to an established test. The aim of this experiment was to investigate the validity of the CVS using data collected at City, University of London.

4.5.3.1 Methods

Two hundred and eight participants were recruited via local advertisements and the AVOT clinic service at City, University of London. The inclusion criteria required that all participants were healthy and between the age of 18 and 75 with no history of ocular disease. All participants wore any refractive correction that they habitually used for tasks at the working distance for each test.

Each participant completed the CAD and CVS test. For participants who completed multiple tests, as part of either experiment I or experiment II, the first CAD and CVS test results were used. The results of the CVS were evaluated relative to the classification and severity of loss measured by the CAD test to determine the sensitivity and specificity of the CVS.

Data collection began using v2.5.14 of the CAD and CVS tests, however over the course of the study two updates, v2.6.1, and v2.8, were released. As mentioned in section 2.3, the version of the CAD test employed in the AVOT clinic is updated shortly after a new update from City Occupational Limited is made available. Consequently three versions of the CAD and CVS tests (v2.5.14, v2.6.1 and v2.8) were used over the course of the experiment.

The changes to the CAD test in v2.6.1 and v2.8 do not change the test procedure and consequently have no impact on the outcome of the CAD test. Version 2.5.14 and 2.6.1 of the CVS use a viewing distance of 0.8 metres while v2.8 uses the same viewing distance as the CAD test (1.4 metres). Although the viewing distance increases to 1.4 metres the angular dimensions of the CVS stimuli and background remained unchanged (as described in section 4.2).

4.5.3.2 Results

The 208 participants ranged from 18 to 71 years old, with a median age [and interquartile range] of 30 [24 - 35] years. The participants included 53 normal trichromats, 107 deutans, 42 protans, 5 congenital deutans with acquired YB loss, and 1 participant with an abnormal YB CAD threshold and a normal RG CAD threshold. The distribution of RG and YB CAD thresholds for the five CAD diagnosis categories are shown in Figure 4.10.

Approximately 94.2% of CVS tests had an ‘excellent’ suprathreshold response reliability (with 100% of suprathreshold stimuli correctly identified), ~3.9% had a ‘good’ response reliability (with ~95% of suprathreshold stimuli [21/22] correctly identified), and ~1.9% had a ‘poor’ response reliability (with ~91% of suprathreshold stimuli [20/22] correctly identified). There were no ‘unable’ CVS test results.

A summary of the CVS results are shown in Table 4.6. The range of RG, YB, and suprathreshold CVS results for each class of participant and the percentage of RG and YB stimuli correctly identified on the CVS as a function

CAD classification	Number assessed	CVS pass	
		R/G (%)	Y/B (%)
Normal	53	100.0	100.0
Deutan	107	0.0	100.0
Protan	42	0.0	100.0
Deutan w. Acq	5	0.0	0.0
Normal RG w. Acq	1	100.0	100.0

Table 4.6: CVS data collected from 208 participants as part of the CVS validity study at City, University of London.

of participant's RG and YB CAD thresholds are shown in Figures 4.11 and 4.12 respectively.

The specificity of the CVS was 1.00 with 100% of participants identified by the CAD test as having normal colour vision classified as having normal RG and YB colour vision by the CVS. The sensitivity of the CVS for deutan and protan was also 1.00 with all 107 deutan and 42 protan tested correctly identified as having normal YB chromatic sensitivity and a RG deficiency by the CVS. The 5 participants diagnosed as having acquired YB loss on top of a congenital deutan deficiency were correctly identified as having both RG and YB loss by the CVS. An individual with normal RG sensitivity and YB loss passed the RG and YB elements of the CVS and was not identified as having YB loss by the CVS.

4.5.3.3 Discussion

The results highlight the high specificity and sensitivity of the CVS test as a consequence of the high level of agreement between the CAD and CVS tests. For the population assessed in this study the specificity of the CVS was found to be 100%, with all normal trichromats passing the RG and YB components in the CVS. The sensitivity for those with congenital deutan and protan deficiencies was equally high, with all 149 participants failing the RG and passing the YB component of the CVS. All participants who were found to have acquired YB loss on top of a congenital deutan deficiency were detected by the CVS as having both RG and YB loss.

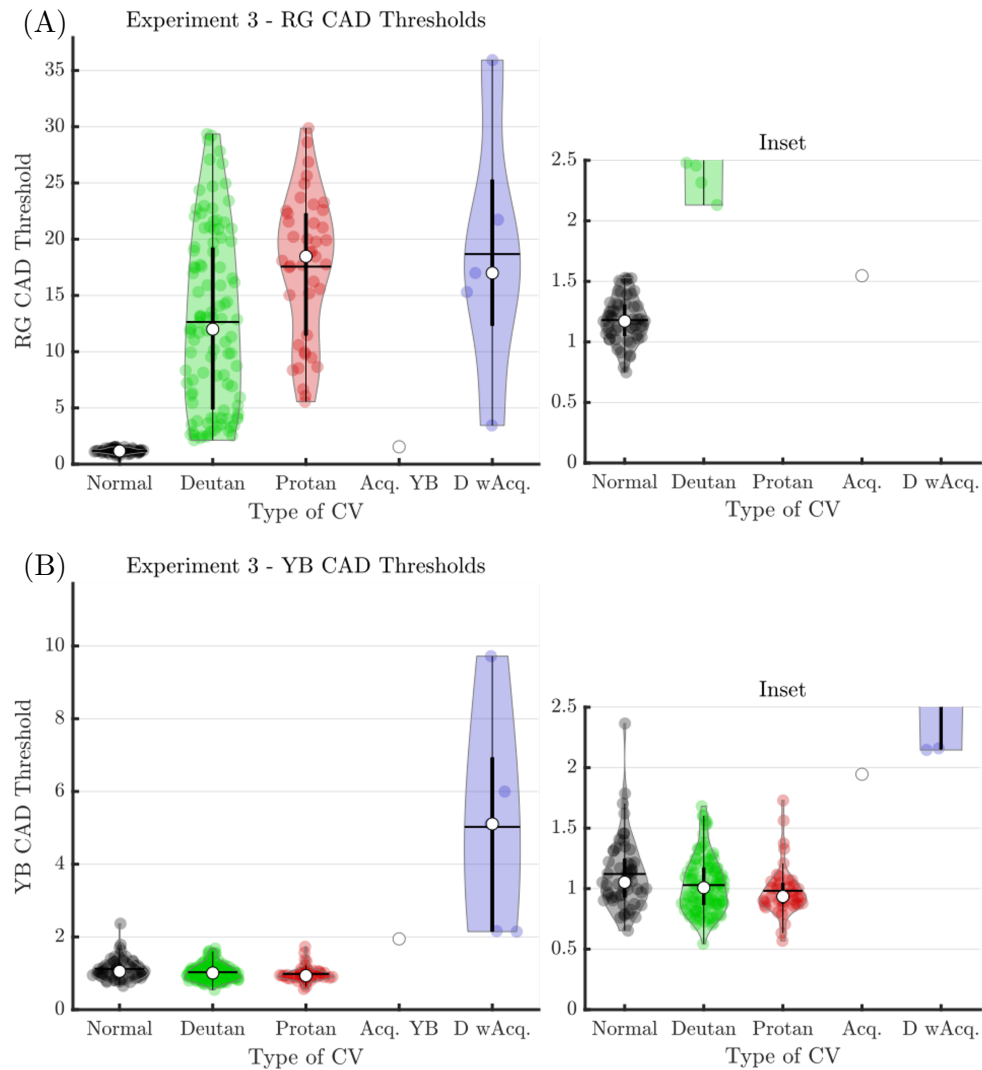


Figure 4.10: Distribution of (A) RG and (B) YB CAD thresholds for participants assessed in experiment 3. The largest YB CAD threshold for participants with YB CAD thresholds within the limit for their age (2.37 CAD units) was obtained by the oldest participant in this category, a 71-year-old. This YB CAD threshold is below the normal upper age-adjusted YB limit for a 71-year-old observer (2.57 CAD units).

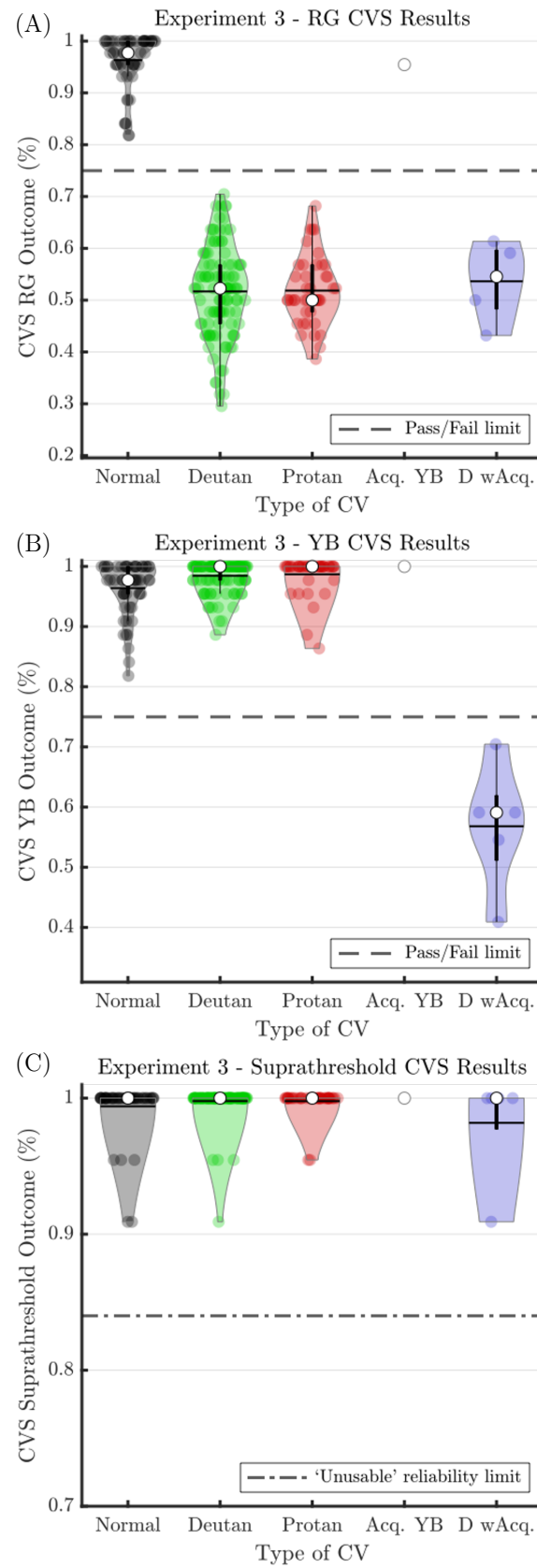


Figure 4.11: (A) RG, (B) YB, and (C) suprathreshold CVS outcomes for participants assessed in experiment 3.

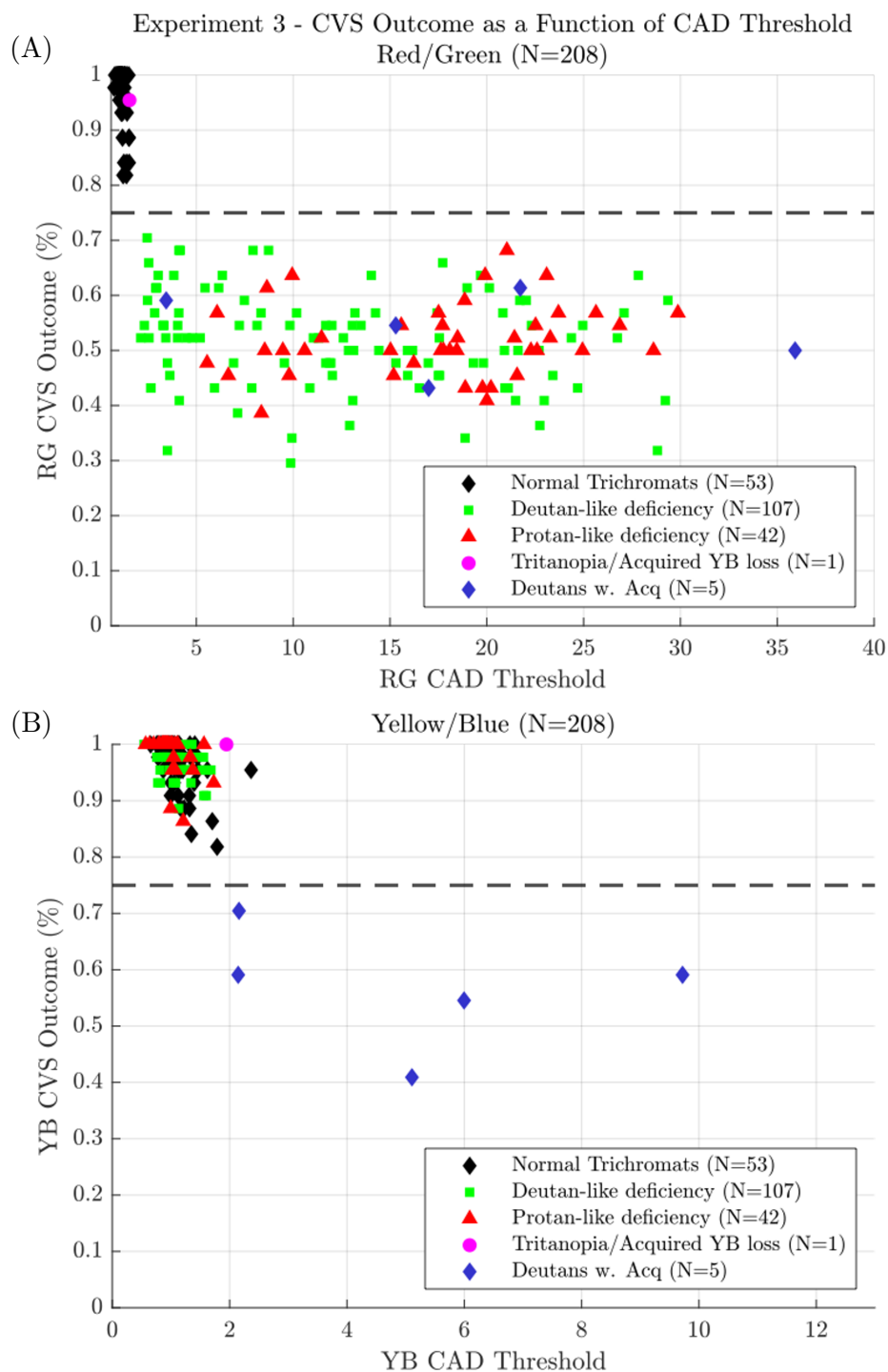


Figure 4.12: (A) RG and (B) YB CVS outcomes for the 208 participants assessed in experiment 3 compared to each participant's RG and YB CAD threshold.

One participant was found to have normal RG sensitivity and acquired YB loss with a YB CAD threshold just above the normal limit for their age (1.95 CAD units compared to the upper normal age-adjusted limit of 1.75 CAD units). This participant passed both the RG and YB components of the CVS, despite ‘failing’ the YB CAD test and being diagnosed as having acquired YB loss. This participant also took part in experiment 2 and had a CVS test-retest reliability of 100% across all three sessions – there was a consistent disagreement between the CAD and CVS tests. There are several potential reasons for the discrepancy between the two tests in this individual. However, given the high test-retest reliability, it is exceptionally unlikely that the disagreement is due to variability in the CVS diagnosis or the participant’s responses.

This finding may reflect the fact that the range of severity of acquired colour vision defects exists on a continuum with normal colour vision. Whereas there is a small but distinct separation between the least sensitive normal trichromats and the most sensitive deuterans due to the within- and inter-subject variability, no such separation exists for acquired colour vision loss. Indeed, for a normal trichromat with relatively low RG and YB CAD thresholds even after developing a form of acquired colour vision deficiency their CAD thresholds may still be within the normal limits for their age. An individual with an initial threshold of ~ 0.5 CAD units could suffer a threefold increase in their CAD threshold and still be within the normal limits expected for their age. The only way to detect and monitor such cases is to undertake regular eye examinations and carry out regular binocular and monocular colour vision testing (with a particular focus on monocular changes which are a unique characteristic of acquired colour vision deficiency).

An additional factor may be the level of luminance contrast (LC) noise used in the CAD and CVS tests. The higher LC noise in the CVS should result in a stronger masking of the luminance contrast signal, however the participant performed better using the higher level of LC noise employed in the CVS test. One would also expect such an effect to be present in other participants and this was not the case.

Coordination issues may also contribute to this observation. Given the additional response buttons used in the CAD test, a participant with poor motor skills may find a 4AFC procedure more challenging than a 2AFC procedure. This, however, should be minimised by the completion of the learning mode in the CAD test.

A limitation of the experimental procedure is that an anomaloscope was not used to confirm the colour vision classification made by the CAD test. This is a key limitation present across all of the experimental work described in this chapter. Although the agreement between the CAD classification and the anomaloscope is generally very high (as seen in section 3.2), future work could establish the agreement between the CVS test, the anomaloscope, and the CAD test.

An additional point regarding this limitation and the use of the CAD test as a ‘gold standard’ in these experiments, as opposed to a test such as the Nagel anomaloscope, was previously touched upon in section 2.4.2 ‘The ‘Gold standard’ of colour vision assessment’ - there are multiple ‘gold standards’ depending upon the dimension of colour vision one wishes to quantify.

No test is perfect and although regularly referred to as the ‘gold standard’ for distinguishing between protan and deutan deficiencies, and anomalous trichromatism and dichromatism, there is a well established history of cases of ‘extreme anomalous trichromacy’, as determined by the results of the anomaloscope (Walls 1959, Birch 1993). These cases are uncommon, estimated as 0.25% of the male population by Birch (2003), but they are worth noting specifically because they are cases in which there is no clear RG classification from the Rayleigh match made using the anomaloscope (the ‘gold standard’ for determining the type of RG deficiency). Previous work has also highlighted cases in which individuals are genetically found to be dichromats, yet are classified as anomalous trichromats using an anomaloscope (Barbur et al. 2008).

In this study the severity of RG loss, RG classification, severity of YB loss, and YB classification were all parameters which needed to be determined for each participant. Whilst the anomaloscope is able to classify the type of

RG loss, provided there is no YB loss, the test does not measure the other three properties of interest and is unsuitable to be used, at least in isolation. Although this could be overcome by completing additional conventional tests, one must be aware of the ethical obligations all researchers have - notably the respect for participants and the time they volunteer to take part in the study. There is a similar argument to support the investigation of the direct agreement between the CVS test and the Ishihara test in future studies, particularly as Dain (2004) stated that the Ishihara pseudoisochromatic plate test is considered to be the ‘gold standard’ in screening for congenital RG CVD. However, the test does not screen for YB loss and has several other limitations which have been described at the start of this chapter and in chapters 2 and 3. As such, to evaluate the CVS test the Ishihara plates would need to be used as part of a test battery, likely alongside the anomaloscope and a sensitive test for YB loss.

4.6 Summary

This chapter introduced a new method of colour vision assessment, the CVS test. A preliminary study with 64 normal trichromats led to the determination of suitable test parameters, including the chromatic signal strength. Following the determination of the test parameters, a study was carried out at City, University of London to ascertain the potential sensitivity and experimentally establish the reliability and validity of the CVS.

The preliminary study found that all normals tested with a preliminary version of the CVS passed the test when a signal strength equal to $+2.0\sigma$ was employed. Weibull functions fitted to experimental CVS data at various signal strengths and combined with the binomial probability density function, revealed the probability of two normal trichromats ($\sim 100\%$), three least affected deuterans ($\sim 18.6\%$, $\sim 14.3\%$, and $\sim 6.0\%$), and a least affected protan ($\sim 0.9\%$) passing the CVS. The results demonstrated that $\sim 0.53\%$ of deuterans and 0.09% of protans are expected to pass the CVS and indicate that the CVS is able to achieve a high sensitivity and specificity.

This experiment also highlights the relatively large impact that small changes in CAD thresholds close to the normal upper age-adjusted limits can have upon the probability of passing the CVS - particularly for least affected deuterans. A relatively small increase in the RG CAD threshold from 2.13 to 2.90 CAD units was associated with a ~ 3 times reduction in the probability of passing the CVS (from $\sim 18.6\%$ to $\sim 6.0\%$).

An investigation into the test-retest reliability of the CVS in six normal trichromats, two congenital deuterans, and one individual with acquired YB loss found that the CVS achieves a high test-retest reliability. The validity of the CVS was measured using a larger sample comprising of 208 participants. The specificity of the CVS was 1.0 for a sample of 53 normal trichromats. The CVS correctly identified 107 congenital deuterans, 42 congenital protans, and 5 participants with acquired loss on top of a congenital colour vision deficiency. The results of the CAD and CVS tests disagreed for one participant with slight acquired YB loss. This disagreement was consistent over 24 CVS tests and 2 CAD tests.

The findings described in this chapter are promising, indicating that the CVS has both the potential and ability to screen for RG and YB CVD with a high sensitivity and specificity. It is important, however, to appreciate that these measurements were all carried out by practised examiners who have a high degree of familiarity with both the CAD and CVS tests. For the CVS to be widely adopted the performance of the test in multiple external centres with a more diverse group of users needs to be evaluated.

Chapter 5

The Colour Vision Screener: A multi-centre study

Chapter 4 described experiments carried out at City, University of London to determine the test parameters, reliability and validity of the CVS test. Whilst the results collected at City, University of London were promising the generalisability of the work carried out, and consequently the translatability of the CVS, remains to be established.

The focus of this chapter is a multi-centre CVS validation study. The design, implementation, and outcome of a study carried out at a number of CAD centres is described.

5.1 Introduction

Multi-centre studies, in which research is conducted in multiple centres following the same protocol, can confer several advantages over single-centre studies including a larger sample size, more diverse population, and increased generalisability (Meinert & Tonascia 1986, Appel 2006, Lucas et al. 2007). While frequently used in clinical trials a collaborative multi-centre approach can be taken with almost all research (Cheng et al. 2017). However everything has a counterpoise and multi-centre studies introduce their own set of challenges requiring thorough design, clear and purposeful coordination, and strict adherence to any and all specified protocols (Das 2022).

The CAD test is sold commercially by City Occupational Limited. CAD systems, comprising of calibrated hardware and specific software, can be found around the world at clinical and occupational centres alike. The resulting international network of CAD centres provides an immensely valuable resource which can facilitate the examination of a large number of participants in a number of separate locations using consistent hardware and software. This allows for the widespread adoption and use of occupation specific CAD limits, something that would be exceptionally challenging without multiple testing centres. The international network of CAD centres also enables a multi-centre study methodology to be used to collect and analyse CAD (and CVS) data.

It should be noted that, as with any method, there are potential limitations and pitfalls associated with a multi-centre approach. Several of these are outlined by Seifirad & Alquran (2021) who, amongst other things, emphasise the importance of taking a pragmatic approach and ensuring that one is mindful of the applicability of large multi-centre study results to less heterogeneous populations. Interestingly such considerations are also at the heart of personalised medicine (Vogt et al. 2016, El-Alti et al. 2019). Within the context of multi-centre colour vision research, the primary consideration is related to the prevalence of CVD in separate populations.

Numerous studies undertaken to measure the prevalence of CVD in different populations, with varying levels of precision, have reported male RG preva-

lence's ranging from approximately 8% to 2% (Birch 2012). However, as expressed by Birch (2012) this range appears to be at odds with the Hardy-Weinberg equilibrium - in large populations that do not mate within a restricted group the prevalence of CVD should remain constant from one generation to the next (Edwards 2008, Abramovs et al. 2020). When evaluating the reported differences in populations across such studies it is critical to determine and acknowledge the effect of the screening method used, data collection protocol, sample size, and randomness of the sample on the reported prevalence (Birch 2012).

The potential impact of the screening protocol employed upon the observed prevalence of CVD has been highlighted in a recent study by Arnegard et al. (2022) in which 193 young Norwegian males were screened for congenital RG CVD. Screening was carried out with both the 24-plate edition of the Ishihara test (≥ 3 errors) and genetic testing using the Agena MassArray system. Genetic screening revealed a 10.4% prevalence of congenital RG CVD, yet the results of the Ishihara test only indicated a prevalence of 5.2% in the same sample. A similar discrepancy between the two screening methods was reported for a female sample. Although the male sample does not contain the number of participants recommended by Birch (2012) ($N \geq 5000$) these findings have to be considered when investigating the prevalence of congenital CVD using conventional methods of colour vision assessment. It should, however, be noted that several other studies have used the Ishihara test to identify the prevalence of RG CVD using larger samples (≥ 5000 male participants) in European caucasian populations and found prevalences $\sim 8\%$, including a study carried out in Norway with 9049 male participants (Waalder 1927, Birch 2012).

The sensitivity, the probability that a test will correctly identify individuals with CVD, and the specificity, the probability that a test will correctly identify individuals with normal trichromatic colour vision, were first introduced in section 2.1. In addition to the sensitivity and specificity there are two other statistical metrics which are useful to consider when evaluating the validity of a test (whether the test measures what it claims to measure). The positive

predictive value (PPV) is the proportion of individuals who fail the test that are correctly diagnosed as having a CVD. The negative predictive value (NPV) is the proportion of individuals who pass the test that are correctly diagnosed as not having a CVD (Altman 1991, Dekking et al. 2005, Trevethan 2017). A mathematical description of the sensitivity, specificity, PPV, and NPV are given in Appendix B (equations B.4, B.5, B.6, and B.7 respectively).

The primary aim of the multi-centre study described in this chapter was investigate the validity of the CVS test using data collected at several CAD centres. Secondary aims were to directly compare the outcome of the CVS to the Ishihara test, investigate the translatability of the CVS, determine how the test performs in a real-world setting, and to examine the difference in RG and YB prevalence across centres.

5.2 Methods

In late 2019 several external CAD centres were invited to take part in a multi-centre study to validate the new CVS test. Participation was voluntary and centres could choose to withdraw at any time. All centres were informed that those who opt to participate would receive the latest version of the CAD test which included the version of the CVS test designed for use on calibrated displays. Additionally, centres were informed that participating centres would be mentioned in any publication describing the results of the multi-centre study with the principal researchers at each participating centre being acknowledged as contributing authors.

Centres who agreed to participate were provided with a number of resources including an example research and ethical form, an example participant information sheet, and a simple contractual agreement stating the terms of the study. The centres who agreed to join the project and to upgrade the CAD test were also provided with the CVS from City Occupational Limited (v2.6.1). All centres agreed to follow all local research and ethical guidelines and to securely store all CAD and CVS data. Participating centres also agreed to follow the CVS testing protocol described in chapter 4, taking particular

care to ensure all tests were carried out at the correct viewing distance with a calibrated display. It should be noted that all CAD centres are provided with, and have access to, the LUMCAL calibration program produced by City Occupational Limited. Twelve centres initially agreed to take part in the study including a ‘City, University of London’ centre. Centres were instructed to remove all personal data from the CAD and CVS files to ensure the data transfer complied with general data protection regulations. The secure transfer of study data was facilitated via a City, University of London shared OneDrive folder or via email with both methods employing multi-factor authentication. The exclusion criteria were participants under 18 or over 75 years of age, participants who completed the CAD or CVS test monocularly, data which did not include the CVS results, measurements carried out with filters present, and measurements carried out using uncalibrated displays or a display mode which had not been calibrated. Following the application of the exclusion criteria, data from five centres were analysed, including the data collected at City, University of London described in section 4.5.3. Analysis was carried out on a centre-by-centre basis and by considering all data in a combined cohort. A description of the five centres is given in Table 5.1.

The CAD and CVS data were used to establish the prevalence of RG and YB deficiency at each centre, and to calculate the RG and YB CVS sensitivity and specificity. The RG and YB CVS positive and negative predictive values (PPV and NPV respectively) were calculated for a prevalence of 8% and 4.25% for RG deficiency and 2% for YB deficiency. The prevalence, sensitivity, specificity, PPV, and NPV were also calculated for a large sample of Ishihara data published by Rodriguez-Carmona, Evans & Barbur (2020) and compared to the multi-centre RG CVS outcome. Mathematical descriptions of the prevalence, sensitivity, specificity, PPV, and NPV are given in Appendix B. The CAD test data were used to determine the ‘true’ status (as defined in Appendix B) of each participant’s RG and YB colour vision.

The study followed the tenets of the Declaration of Helsinki. Data collection followed all local research and ethical guidance and was approved by local

No.	No. assessed	Centre name	Location
1	208	City, University of London	London, UK
2	8	Airport Medical Centre, Stuttgart Airport	Stuttgart, Germany
3	6	US Navy Refractive Surgery Center	San Diego, USA
4	28	University of Bergen	Bergen, Norway
5	31	ViOLa Visual Optics Lab, National Institute of Optics CNR	Florence, Italy

Table 5.1: CAD centres which contributed useable data to the multi-centre CVS study and the number of participants assessed at each centre.

research and ethical committees. All the tests used were approved by the City University Research and Ethics Committee (ETH1819-0206).

5.3 Results

The number of study participants assessed at each CAD centre is shown in Table 5.1. Participants were between 18 and 71 years old with a median age [and interquartile range (IQR)] of 30 [23 - 37] years. The median participant age ranged from 29 to 43 years across the five centres and the youngest and oldest participant were assessed at the same centre (Figure 5.1). The range of CAD thresholds measured across all centres is shown in Figure 5.2 and a centre-by-centre comparison of CAD thresholds is shown in Appendix C (Figure C.1).

The percentage of participants who passed the RG and YB components of the CVS for the entire cohort and at each centre are shown in Tables 5.2 and 5.3. The range of RG, YB, and suprathreshold CVS results for participants with each class of CVD are shown in Figure 5.3 (a centre-by-centre breakdown of RG and YB CVS data is shown in Appendix C [Figure C.2]). The percentage of RG and YB stimuli correctly identified on the CVS as a function of participant's RG and YB CAD thresholds are shown in Figure 5.4.

Ninety-eight point eight per cent of normal trichromats were correctly classified as having normal RG and YB sensitivity by the CVS. The only normal

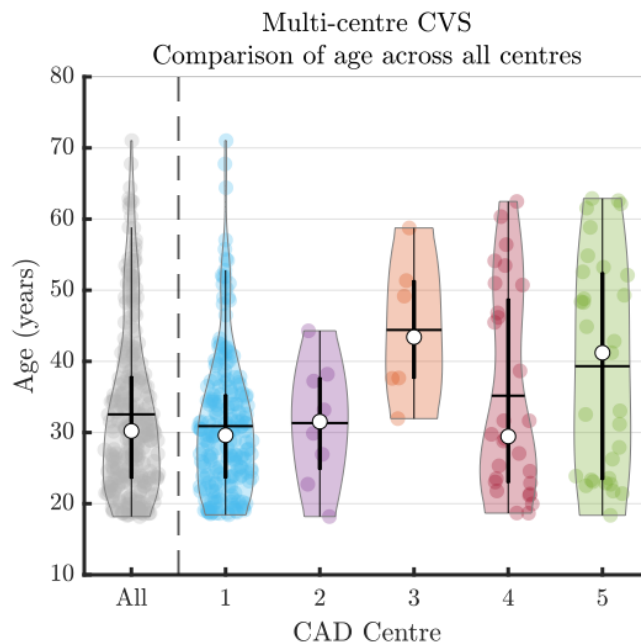


Figure 5.1: Age distribution of participants assessed at each CAD centre as part of the multi-centre CVS study.

trichromat incorrectly classified by the CVS was tested at centre 2. This participant had a RG and YB CAD thresholds of 1.77 and 0.93 CAD units and passed the YB (100% correct), but failed the RG component of the CVS ($\sim 73\%$ correct). All deuterans and protans were correctly identified by the CVS as having RG loss with normal YB sensitivity.

All six participants diagnosed as having RG loss consistent with a deutan deficiency and YB loss failed the YB component of the CVS. One participant in this subgroup passed the RG component of the CVS.

Two participants with acquired YB loss, out of five assessed, passed the RG and YB components of the CVS. One of these participants was assessed at centre 1, and has been described previously, the other was assessed at centre 5 (Table 5.3).

The prevalence, sensitivity, specificity, PPV, and NPV for the RG and YB component of the CVS are shown in Tables 5.4 and 5.5. The prevalence ranged from 0.129 - 0.893 for RG loss and 0.000 - 0.129 for YB loss across the five centres. When considered holistically the combined cohort had a prevalence of 0.690 for RG loss and 0.039 for YB loss. The sensitivity and specificity of the CVS was 0.995 and 0.989 for the RG stimuli and 0.818 and 1.000 for the

YB stimuli. For a 2% prevalence of YB loss the PPV of the CVS was 1.0000 and the NPV was 0.9844. For a 4.25% prevalence of RG loss the PPV and NPV of the CVS was 0.7935 and 0.9998.

The statistical outcome of four commonly used Ishihara protocols published by Rodriguez-Carmona, Evans & Barbur (2020) and the RG CVS data for the entire multi-centre study cohort are shown in Table 5.6. The sensitivity and specificity of the Ishihara test ranged from approximately 0.961 - 0.995 and 0.806 - 0.994 for the published protocols. For an 8% prevalence of RG loss the PPV of the CVS was 0.8827 and the PPV of the Ishihara test ranged from 0.3080 - 0.9360. The NPV of the CVS was 0.9995 and the NPV of the Ishihara test protocols ranged from 0.9966 - 0.9994.

CAD classification	Number assessed	CVS pass	
		R/G (%)	Y/B (%)
Normal	82	98.8	100.0
Deutan	133	0.0	100.0
Protan	55	0.0	100.0
Deutan w. Acq	6	16.7	0.0
Normal RG w. Acq	5	100.0	40.0

Table 5.2: Combined CVS data collected from 281 participants as part of the multi-centre CVS study at five CAD centres.

Type of CVD	Number assessed	CVS pass	
		R/G (%)	Y/B (%)
A) Centre 1			
Normal	53	100.0	100.0
Deutan	107	0.0	100.0
Protan	42	0.0	100.0
Deutan w. Acq	5	0.0	0.0
Normal RG w. Acq	1	100.0	100.0
B) Centre 2			
Normal	1	0.0	100.0
Deutan	6	0.0	100.0
Protan	1	0.0	100.0
C) Centre 3			
Normal	2	100.0	100.0
Deutan	2	0.0	100.0
Protan	2	0.0	100.0
D) Centre 4			
Normal	2	100.0	100.0
Deutan	16	0.0	100.0
Protan	9	0.0	100.0
Normal RG w. Acq	1	100.0	0.0
E) Centre 5			
Normal	24	100.0	100.0
Deutan	2	0.0	100.0
Protan	1	0.0	100.0
Deutan w. Acq	1	100.0	0.0
Normal RG w. Acq	3	100.0	33.3

Table 5.3: Individual centre data collected as part of the multi-centre CVS study at five CAD centres. Zero values, where no participants with a specific type of CVD were assessed, have been removed from the table.

Centre	N	Prev	Sens	Spfc	Prev: 0.0800		Prev: 0.0425	
					PPV	NPV	PPV	NPV
All	281	0.6904	0.9948	0.9885	0.8827	0.9995	0.7935	0.9998
1	208	0.7404	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	8	0.8750	1.0000	0.0000	0.0800	-	0.0425	-
3	6	0.6667	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	28	0.8929	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	31	0.1290	0.7500	1.0000	1.0000	0.9787	1.0000	0.9890

Table 5.4: The number of participants assessed (N), prevalence (Prev), sensitivity (Sens), and specificity (Spfc) for RG CVS data collected at the centres who took part in the multi-centre study. The positive and negative predictive values (PPV and NPV respectively) have been calculated for a prevalence of 0.08, or 8%, and 0.0425, or 4.25%.

Centre	N	Prev	Sens	Spfc	Prev: 0.0200	
					PPV	NPV
All	281	0.0391	0.8182	1.0000	1.0000	0.9844
1	208	0.0288	0.8333	1.0000	1.0000	0.9857
2	8	0.0000	-	1.0000	-	-
3	6	0.0000	-	1.0000	-	-
4	28	0.0357	1.0000	1.0000	1.0000	1.0000
5	31	0.1290	0.7500	1.0000	1.0000	0.9787

Table 5.5: The number of participants assessed (N), prevalence (Prev), sensitivity (Sens), and specificity (Spfc) for YB CVS data collected at the centres who took part in the multi-centre study. The positive and negative predictive values (PPV and NPV respectively) have been calculated for a prevalence of 0.02, or 2%.

Test and protocol	N	Prev	Sens	Spfc	Prev: 0.0800	
					PPV	NPV
CVS RG - All centres	281	0.6904	0.9948	0.9885	0.8827	0.9995
IH 38 pl. 0 err pl. 1-25	1827	0.8084	0.9946	0.8057	0.3080	0.9994
IH 38 pl. 0 err pl. 1-15	1827	0.8084	0.9932	0.8886	0.4367	0.9993
IH 38 pl. ≤ 2 err pl. 1-17	1827	0.8084	0.9716	0.9886	0.8808	0.9975
IH 38 pl. ≤ 4 err pl. 1-21	1827	0.8084	0.9607	0.9943	0.9360	0.9966

Table 5.6: The number of participants assessed (N), prevalence (Prev), sensitivity (Sens), and specificity (Spfc) for RG CVS data compared to statistics calculated from a large sample of published Ishihara (IH) data from Rodriguez-Carmona, Evans & Barbur (2020). The positive and negative predictive values (PPV and NPV respectively) have been calculated for a prevalence of 0.08, or 8%.

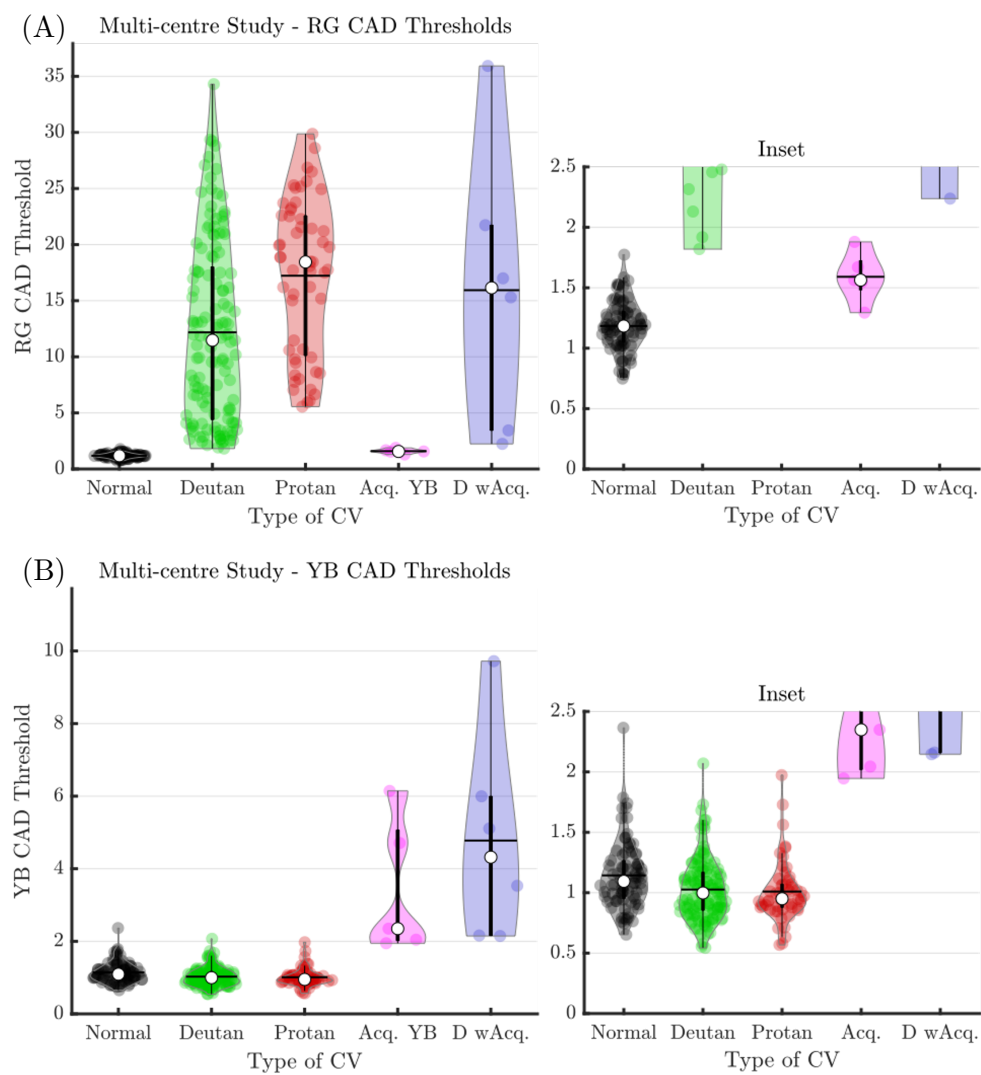


Figure 5.2: Distribution of (A) RG and (B) YB CAD thresholds for 281 participants assessed as part of the multi-centre CVS study.

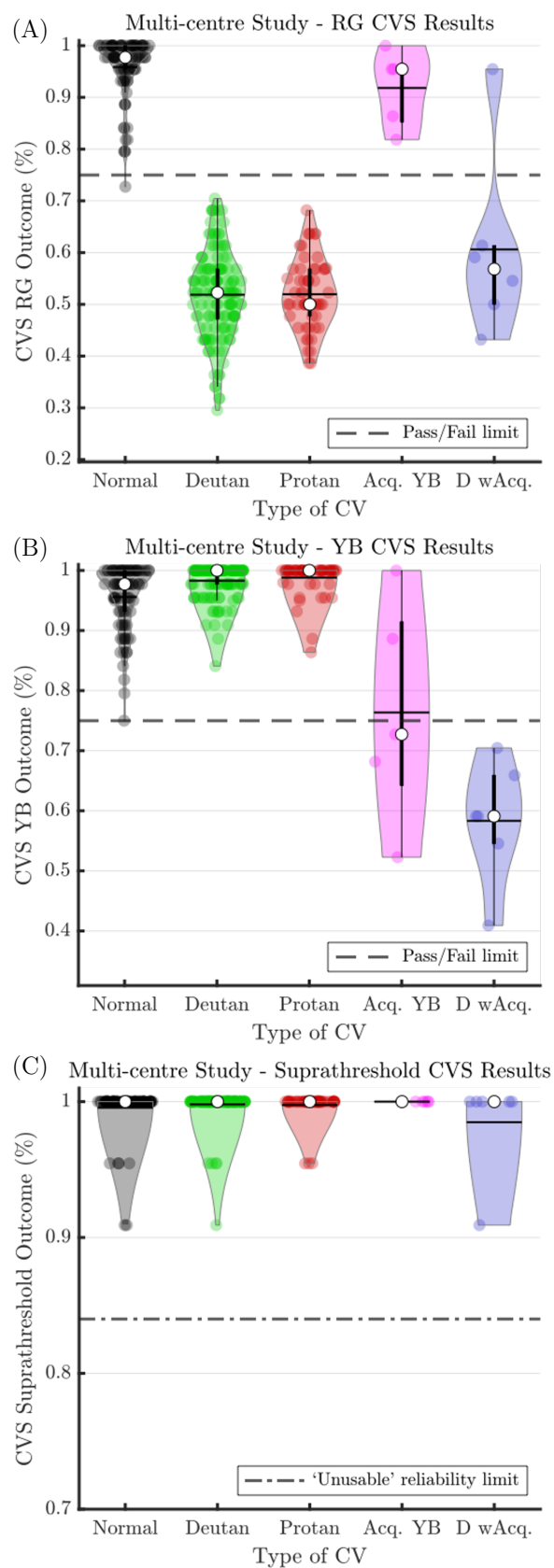


Figure 5.3: (A) RG, (B) YB, and (C) suprathreshold CVS outcomes for participants assessed as part of multi-centre CVS study.

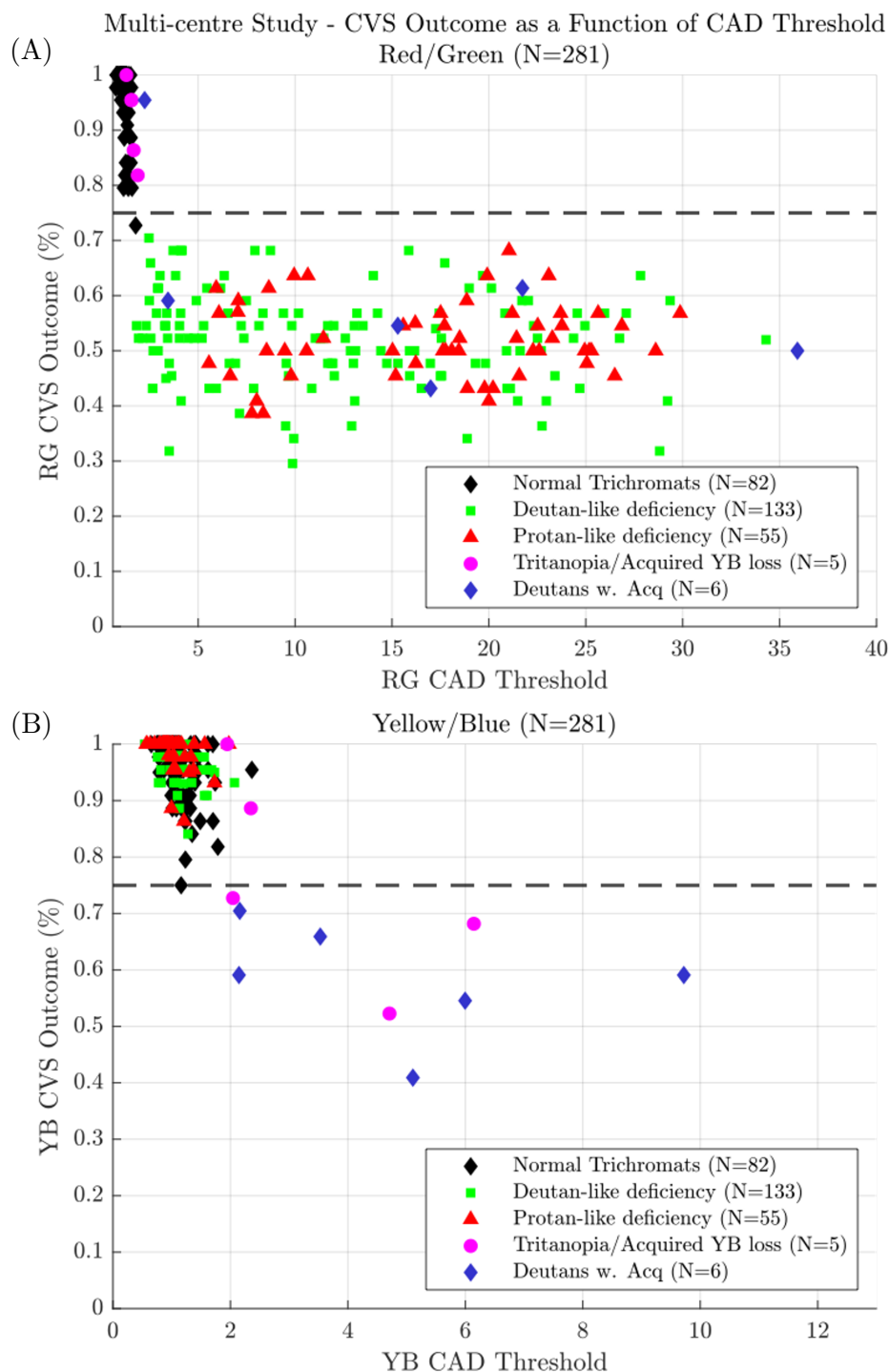


Figure 5.4: (A) RG and (B) YB CVS outcomes for the 281 participants assessed as part of multi-centre CVS study compared to each participant's RG and YB CAD threshold.

5.4 Discussion

The participants assessed were primarily young and of working age with a median [and IQR] age of 30 [23-37] years. The overall prevalence of CVD was 0.690 for RG CVD and 0.039 for YB CVD. It is important to note, however, that both the observed range of assessed participant's ages and the prevalence of CVD varied across the five centres. The observed inter-centre differences are primarily attributable to the number of participants assessed at each centre and each centre's social location.

As shown in Table 5.1, the number of participants assessed at each centre ranged from 6 to 208 with most participants ($\sim 74\%$) being assessed at centre 1 (participants assessed at centre 1 were described and discussed in section 4.5.3). As expected, the range of participant's ages was larger at centres that assessed a larger number of participants (centres 1, 4 and 5). The impact of the centre location is also visible in Figure 5.1. Centres based near or around universities are embedded in locations where many potential research participants are likely to be students (reflected in the age distribution for centres 1, 4, and 5). Whereas CAD testing carried out in a centre focused upon testing individuals undergoing refractive surgery, regardless of whether the research participants have or are planning to undergo refractive surgery, is likely to assess a sample of participants with a slightly higher median age. When evaluating centres in this way one must also acknowledge and recognise the low sample sizes at centres 2 and 3 compared to the other centres.

The prevalence of RG CVD ranged from 0.67-0.89 across centres 1-4. In centre 5 the prevalence of RG CVD was 0.13 (Table 5.4). One may be tempted to conclude that this is clear evidence of the difference in the prevalence of CVD across different populations. However, the reported difference in prevalence of RG CVD is likely to be more attributable to the number of participants seen and the recruitment procedure at each centre, which in turn are heavily influenced by the location and primary role of each centre. By virtue of their position CAD centres are primarily established, advertised, and used to assess those with known or suspected CVD and to determine whether

individuals have the level of chromatic discrimination required to work in a specific occupational setting. Consequently, the high prevalence of RG CVD in centres 1-4 is not surprising.

One must also acknowledge the limitation of attempting to interpret and attach meaning to the prevalence from a small sample size. This is particularly relevant for centres 2 and 3 where less than 10 participants were assessed, but also applies to the calculated prevalence for centres 4 and 5. For a small sample the prevalence is a descriptive metric indicating the proportion of participants with CVD and has limited value within a wider context. As such, the prevalence in centres 2-5, at least when considered in isolation, cannot be used to infer meaningful information about the prevalence in a larger population.

The recruitment process at each centre, and the impact of the COVID-19 pandemic just a few months after centres were invited to participate in the study also contribute to the observed inter-centre variation. Participants who are more vulnerable members of society or those caring for vulnerable members of society are less likely to take part in research if they are actively minimising their social exposure. From a centre perspective, the additional challenges and responsibilities of research centres to ensure the safety and wellbeing of research participants will have an unavoidable impact upon the ease of participant recruitment. These challenges were experienced by all centres, but the exact nature of the difficulties varied over the course of the study due to international restrictions and lockdowns over the course of the study.

Fortunately, the age-adjusted nature of the classification made by the CAD test ensures that any systematic inter-centre differences in the median age of participants are taken into account by CAD test when the class of CVD is determined.

As shown in Figure 5.2, the YB CAD thresholds for those diagnosed as having normal RG sensitivity with acquired YB loss ranged from just above the normal limit, 1.94 to 6.15 YB CAD units. The CAD thresholds for those diagnosed as having a deutan deficiency and acquired YB loss ranged from 2.24 to 35.92 RG CAD units and 2.15 to 9.72 YB CAD units. Participants

diagnosed as having a congenital deutan or protan deficiency had YB CAD thresholds within the normal limit for their age and RG CAD thresholds outside the normal limit for their age, and participants diagnosed as having normal trichromatic colour vision obtained RG and YB CAD thresholds within the normal limits for their age.

The outcome of the CVS is shown in Tables 5.2 and 5.3. The normal trichromat who failed the RG component of the CVS at centre 2 obtained a RG CAD threshold of 1.77 and correctly identified 73% of the RG stimuli. This participant's RG CAD threshold is just below 1.79, the normal upper RG CAD threshold limit for their age, and their RG CVS outcome was close to the CVS pass/fail limit of 75%. This participant's CVS result is an expected consequence of the statistical nature of the CVS test. As discussed in section 4.5.1, the predicted probability of correct response on the CVS test, given by the binomial distribution, has a degree of overlap with the CVS pass/fail limit of 75% for participants with CAD thresholds close to the upper age-adjusted limit.

In addition to the normal trichromat tested at centre 2, one deutan with acquired loss passed the RG component of the CVS (assessed at centre 5), and two participants with acquired YB loss and normal RG chromatic sensitivity (assessed at centres 1 and 5) were incorrectly classified by the CVS. The deutan with acquired loss was diagnosed by the CAD test as having a 'potential deutan deficiency with acquired loss' with RG and YB CAD thresholds (and an upper age-adjusted limit) of 2.24 (2.11) and 3.53 (2.19) respectively. Although this participant passed the RG component of the CVS their YB loss was detected and in a screening scenario their YB CVS result would have identified them for further testing.

The first of the two participants with acquired YB loss who passed the YB component of the CVS was assessed at centre 1 and has been described previously (sections 4.5.2 and 4.5.3). The second participant was assessed at centre 5 and obtained a YB CAD threshold of 2.35 (normal upper age-adjusted limit of 1.81) CAD units and correctly identified 89% of the YB CVS stimuli.

It's important to appreciate that, although these participants with acquired loss are incorrectly diagnosed by the CVS, these participants have CAD thresholds close to the normal upper limit for their age (Figure 5.4). While any incorrect diagnosis is undesirable, there is significant value in the likelihood of an incorrect diagnosis being directly proportional to the severity of chromatic sensitivity loss, as is the case in the CVS.

As previously described in section 4.5.3, the continuous nature of acquired colour vision deficiency, and subsequent lack of separation between the least affected individuals with acquired loss and those with normal trichromatic colour vision, makes the screening of mild acquired loss particularly challenging. Only a personalised individual approach to colour vision assessment, making use of a highly sensitive test, monocular changes, and directly comparing the results to previous results for the same individual, can be used to detect the earliest signs of acquired loss for those who naturally have high chromatic sensitivity.

Data from the entire cohort shows that RG component of the CVS is able to achieve a high sensitivity without the cost of a low specificity as often seen in other tests, such as the Ishihara pseudoisochromatic plates. This study found that the RG component of the CVS had a sensitivity of 0.9948. This sensitivity is equivalent than the most sensitive Ishihara protocol evaluated in this study, 0.9946. This Ishihara sensitivity is, however, achieved at the cost of specificity - the corresponding specificity for this Ishihara protocol was 0.8057, far lower than the RG CVS specificity of 0.9885. Although less sensitive Ishihara protocols are able to achieve a higher specificity, 0.9886 and 0.9943, as shown in Table 5.6 these protocols have a far worse sensitivity of 0.9716 and 0.9607 respectively.

The predicted PPV, the proportion of individuals who fail the CVS and are correctly diagnosed as having a CVD, and NPV, the proportion of individuals who pass the CVS who are correctly diagnosed as not having a CVD, are high for the RG component of the CVS. Both the PPV and NPV were calculated for a RG CVD prevalence of 8% and 4.25%, corresponding to the expected prevalence in a male population and a 50/50 male/female popula-

tion. For a RG CVD prevalence of 8% the PPV of the RG CVS is 0.8827 and is higher than all but one of the Ishihara protocols, the least sensitive protocol, which had a PPV of 0.9360. For the same prevalence of 8% the NPV of the CVS, 0.9995, was higher than NPV of all the Ishihara protocols which were evaluated as part of this study, 0.9966-0.9994.

The specificity of the YB component of the CVS was 1.00, with 100% of those with normal YB chromatic sensitivity being correctly identified by the CVS, and the sensitivity of the YB component of the CVS was 0.8182. For an estimated YB CVD prevalence of 2% the PPV was 1.00 and the NPV was 0.9844. The high NPV, given the sensitivity of 0.8182, is due to the low assumed prevalence of YB CVD (2%). The difference in the CVS outcome as a consequence of the continuous nature of acquired CVD, and in contrast to the discrete separation present in congenital CVD is clearly visible in Figure 5.3A and Figure 5.3B.

These results show that the CVS achieves a better outcome than the Ishihara test if one wishes to screen for RG CVD. This is significant in occupational settings where achieving a high sensitivity for RG loss is desirable to detect applicants working in environments where RG loss results in ‘unsafe’ working conditions. It is also worth noting that the CVS can screen for YB loss with a high sensitivity and 100% specificity. The Ishihara test can not do this.

The limitations present in the validation experiment at City, University of London, the use of the CAD test to determine the ‘true status’ of participants’ colour vision, the higher level of LC noise used in the CVS, and the different levels of coordination required to complete a 4AFC task compared to the 2AFC CVS task are all present in the multi-centre study. The impact of these limitations is broadly consistent across all centres and each limitation has been described previously in section 4.5.3.3.

A potential limitation specific to the multi-centre methodology is the protocol used at all CAD centres because one of the primary causes of error in multi-centre studies can be due to a lack of protocol adherence across centres (Meinert & Tonascia 1986, Cheng et al. 2017, Das 2022). This limitation was minimised by displaying clear and concise user instructions as part of the

CAD and CVS software every time CAD or CVS testing is carried out. These instructions describe the test procedure centres should follow. Additionally, the CVS provides on-screen instructions to all participants prior to starting the test, at the end of the learning mode, and at the end of the final CVS test, ensuring that the instructions provided to participants remained constant across all centres. The protocol adherence was not externally validated, instead it was left to each centres' discretion. To quantify protocol adherence one could conduct random 'centre checks' or employ actors as standardised patients to ensure centres are following the specified protocol. Neither of these methods were employed in this study nor possible to implement during the COVID-19 pandemic.

An additional limitation, as previously mentioned, is the unequal recruitment at each centre and the subsequent impact upon the collected data - particularly with regards to the prevalence of CVD in different populations. As described in section 5.1, previously reported differences in the prevalence of congenital CVD have been attributed to the screening method employed and the sampling technique used (Birch 2012, Arnegard et al. 2022). Part of evaluating the validity of the CVS involves testing a large number of participants with CVD. The increased prevalence of CVD in the centre samples, compared to the expected prevalence (at least for RG CVD), is not therefore a disadvantage. Without using a sample with a higher prevalence than expected in the normal population one would need to assess approximately 1663 male participants or 3130 participants with a 50/50 male to female ratio, assuming a prevalence of 8% and 4.25% respectively, to assess 133 deutan participants. Another potential limitation is that the statistical analysis did not include receiver operator curves (ROC). However, as directly addressed by Altman (1991) the primary value of ROC analysis is in comparing two or more pass/fail values and *“For a single test it [the ROC method] does not add anything to a table”* (Altman 1991, p. 418). Given that the pass/fail limit employed in the CVS was determined following a preliminary study and the calculation of the sensitivity, specificity, PPV, and NPV provide clear metrics by which the outcome of the CVS can be compared to other tests, such as the Ishihara test,

ROC analysis is of limited benefit. ROC analysis could, however, be utilised in future if one wished to change the CVS pass/fail limit to further optimise the outcome of the CVS following the collection of a larger sample of CVS data.

Despite these limitations the multi-centre study described in this chapter was carried out on a wider population of both users and participants and consequently achieves a greater validity than the single centre validity study described in section 4.5.3.3 (Meinert & Tonascia 1986, Lucas et al. 2007). The results of the study, as a consequence of the multi-centre methodology, also have a larger potential for translation to clinical practice and policy by demonstrating that the CVS achieves a high sensitivity and specificity at several separate testing locations with a range of users (Das 2022).

The experimental work described in this chapter reveals the high sensitivity and specificity of the new CVS test using a robust multi-centre methodology. Across multiple centres, with a diverse group of users, the CVS reliably detects congenital RG CVD with a sensitivity and specificity of 0.99. For YB loss the specificity is 1.00 and the sensitivity of the CVS is slightly lower (0.82), due in no small part to the continuous nature of acquired CVD. All those incorrectly diagnosed by the CVS are minimally affected with chromatic thresholds close to the normal upper limit for their age. This property of the CVS is advantageous particularly when compared to other screening tests which, in addition to having stimuli which can be memorised, do not have such a strong statistical relationship between the test outcome and participants severity of loss.

Chapter 6

Conclusion and outlook

6.1 Summary of findings

The original research presented in this thesis demonstrates the creation and interpretation of new knowledge extending the forefront of colour vision research. Data from chapters 3 and 4 have been published in peer reviewed journals and two additional manuscripts are planned for work described in chapters 2, 4, and 5.

The first two chapters of this thesis critically analyse and synthesise the current body of knowledge regarding human colour vision and colour vision assessment. The optical, retinal and neural systems which contribute towards the perception of colour are introduced in chapter 1. These are followed in chapter 2 with the theoretical context for the studies described in the experimental chapters, 3-5, along with a description of the challenges of ‘full’ colour vision assessment and the properties of an ideal colour vision test.

A description of experimental work conceived, designed, implemented, and analysed to create new knowledge is primarily contained within chapters 3, 4, and 5. The current state of conventional colour vision assessment is laid bare in chapter 3 as the limitations of single- and multi-test protocols are revealed. An investigation into the Farnsworth D-15 test emphasised the importance of isolating colour signals, and reinforced conclusions drawn from the single- and multi-test protocol investigation. Critically it showed that many single-

and multi-test colour assessment protocols in current use fail to achieve their aims. These protocols are unable to achieve high sensitivity and specificity or reliably quantify the severity of RG and YB loss. This work established the value of, and need for, a sensitive screening test for RG and YB CVD.

The Colour Vision Screener (CVS), a recently developed method to address these issues, is the primary focus of chapters 4 and 5. The early development of the test and a preliminary study carried out to determine the test parameters are described in chapter 4. Experiments carried out at City, University of London established the performance of the CVS in the most sensitive deuterans and protans and determined the reliability of the test. A multi-centre study was coordinated to examine the validity of the CVS. Data collected at City, University of London and a number of external CAD centres quantified the performance of the CVS relative to the CAD test. The results indicate that only a small percentage of those with congenital colour vision deficiency are expected to pass the CVS ($\sim 0.5\%$ of deuterans and $< 0.1\%$ of protans), the CVS test has very high test-retest reliability, sensitivity (RG: 0.99, YB: 0.82), and specificity (RG: 0.99, YB: 1.00).

6.2 Reflection

Reflection, also known as reflective thinking or reflective thought, is a cognitive process in which an individual critically assesses their experiences (Dewey 1933, Rogers 2001). As with many other forms of contemplative practice, the exact process of academic reflection varies (Mortari 2015, Bonnardel et al. 2018). Some suggest processes containing discrete steps or stages while others describe reflection as a more continuous process (Dewey 1933, Schön 1992, Ryan 2011, Mortari 2015). Regardless of the minutia of the technique or process one follows, reflection is a powerful contemplative practice to enhance one's understanding and learning through a critical evaluation of one's actions. Personal reflections made following the work completed and included in this thesis are included here.

I understand that the most critical element of a project is the design process which occurs at the start of a project. Throughout my studies the importance and value of having a clear plan, structure, and methodology from the outset has been strongly reinforced. There are many instances over the course of the project where the early time and resource investment spent designing and planning the study paid dividends. Inevitably some elements of any study will change, either through the emergence of new opportunities or due to unforeseen circumstances. However, these modifications are far easier to address and accommodate through the adaptation of an existing comprehensive plan. Notable examples of this occurred during the exploratory Farnsworth D-15 study in chapter 3, the coordination of the multi-centre study in chapter 5, and dealing with the consequences of the COVID-19 pandemic during the CVS data collection in chapters 4 & 5. The experimental Farnsworth D-15 study was largely possible due to the clear study design with built in review points. It was through one of these review points that the opportunity to complete an exploratory study was discovered.

Reflecting upon the multi-centre study design in particular has shown me how the use of structured and planned communication windows and the nurturing of good inter-personal relationships, to notify all centres with updates regarding the study progress, are invaluable additions present from the design phase of the study. The clear protocol made it far easier to identify the disruption caused by the COVID-19 pandemic and, in turn, incorporate steps to minimise the impact upon the study.

An interesting personal note is that this process, particularly after writing up and publishing some of this work, has reinforced the priority of the questions ‘How am I going to publish this?’/‘How am I going to write this up?’. These questions are now, relative to when I started, far more at the forefront of my mind.

My experience of critical peer review as both an author and reviewer has provided insight into study design and the rigour and forethought needed to ask and address the most pertinent research questions. These experiences have given me a greater appreciation and understanding of the need to hone the

study question, design the study methodology to best answer the question, and ultimately provide a new finding which advances the forefront of a discipline and merits publication. Raising my awareness of this process has begun to permeate my decisions over the course of a project. I am truly glad to have a both a better appreciation and deeper understanding of peer review and publication and its importance for a scientific career.

6.3 Future directions

The work described within this thesis built upon the foundation laid by previous researchers and has opened avenues for future exploration and study. Three of the most alluring future directions enabled by this research are described here.

6.3.1 Continuing validation of the CVS at external centres

The multi-centre validity study described in chapters 4 and 5 investigated the performance of the CVS using data from a number of external CAD centres. Following the publication of the results of the multi-centre study it is likely that the CVS will be used in a ‘two-step’ protocol along with the CAD test at all CAD centres using the latest version of the AVOT software. The continued evaluation of the CVS and continued validation of the test using even larger samples is an exciting avenue to explore.

The primary future exploration would be a direct continuation of the multi-centre validation with larger sample sizes. The results could be analysed using the same techniques described in chapter 5 and would provide increased statistical robustness.

It would also be of considerable interest to use the CVS to investigate the prevalence of RG and YB CVD in large ($N \geq 1000$) random populations. This work could be carried out by existing CAD centres or by a centre with access to the CAD and CVS tests. When carried out in more than one population,

provided the samples were collected in the manner detailed by Birch (2012), such a study provides to the opportunity to obtain statistics regarding the prevalence of RG and YB CVD in different populations.

In the same vein, the agreement between the CVS and other forms of colour vision assessment, particularly the Ishihara plates, AO-HRR plates, and Nagel anomaloscope are of interest and important to investigate. However, in any such study care must be taken to ensure that the ‘true state’ of participants’ colour vision is determined as accurately as possible. This care must also be applied to the interpretation of the results to ensure the findings are reported in context. For instance, some have reported the high agreement between the City University test and the Farnsworth D-15, yet this is of limited value as both tests have poor sensitivity and pass those with severe RG loss.

6.3.2 Investigation into performance of the CVS on uncalibrated visual displays

As introduced in section 4.2, a downloadable version of the CVS has been designed and developed for use on uncalibrated displays running the sRGB colour mode. The mechanics of the test are very similar to the version designed for use on calibrated displays described and used for the experimental work in this thesis. Changes between the downloadable version and version designed for use on calibrated displays are primarily related to the assumed luminance and chromaticity of the visual display.

A project to evaluate the performance of the downloadable version of the CVS upon uncalibrated visual displays running the sRGB colour mode is a natural continuation of the work within this thesis. Such a study would ideally assemble a series of commercially available displays, measure the photometric properties of each display when set to the sRGB colour mode, assess participants using the downloadable CVS, and evaluate the results.

This study would enable one to comment upon the variation observed from the supposedly standard sRGB colour mode, the identification of any potential relationship between the properties of a monitor (i.e. the deviation

from the sRGB standard and chromaticities produced), and importantly the suitability of the CVS for personalised testing on home computers. The use of the CVS on home computers has the potential to greatly expand colour vision assessment, potentially enabling patients to use the test at home to detect the earliest changes in visual diseases, monitor the progression of vision changes caused by systemic diseases, and expand colour vision assessment within visually demanding occupations and schools (Barbur et al. 2020).

6.3.3 Usability and validity of the CVS in children

An investigation into the performance of the CVS on uncalibrated monitors would in turn facilitate an expansion of colour vision assessment within schools (Barbur et al. 2020). The outcome and performance of the CVS in younger participants is an active research area ripe for future exploration.

Notable recent work looking at paediatric colour vision assessment has been carried out by the Sussex Colour Group. Part of this work included the development and validation of an iPad based colour vision test, ‘ColourSpot’, for children as young as 4 (Tang et al. 2022).

Testing younger participants introduces additional challenges and would be a true test of the usability of the CVS - particularly in terms of the instructions provided during the test and the use of the buttons in the 2AFC choice task. Tang et al. (2022) have demonstrated the value of gamification for younger participants. It is an open question as to whether the CVS test procedure will prove effective in maintaining children’s attention and response reliability, but the test is very similar, at least from a participant’s perspective, to a video game.

This avenue also links to future population studies and if the CVS proves to be an efficient test with younger participants it would enable the test to be used for population sampling studies in schools.

6.4 Conclusion

This research is timely and answers and addresses several key questions in the field of colour vision assessment. Investigating the current outcome of commonly used methods of conventional colour vision assessment is of paramount importance to those who work in and rely upon occupations where the use of colour is safety critical. The introduction, development, and validation of a new method of colour vision assessment, the CVS test, enables clinicians and occupational healthcare professionals to efficiently screen for congenital and acquired colour vision deficiencies. Such a screener acts to improve the safety of occupational environments, and has the potential to improve patient care, allowing for the early detection of some retinal and neural diseases as well as enabling a personalised approach to colour vision assessment. The benefits to society and the impact of the original research presented and described within this thesis will, I hope, be felt in years to come.

Appendix A

The statistically equivalent CAD threshold

Participants who pass and fail a given colour assessment protocol will have a range of CAD thresholds. For every possible RG CAD threshold, T_{RG} , there will be a number of participants who fail the protocol with thresholds lower than T_{RG} and a number of participants who pass the protocol with thresholds above T_{RG} . The range of statistically equivalent RG CAD thresholds is defined as the range of thresholds for which these two values (the number of participants who fail with thresholds lower than T_{RG} and the number of participants who pass with thresholds above T_{RG}) are equal.

The statistically equivalent RG CAD threshold for a given colour assessment protocol is the mean of this range, given by

$$SET_{RG} = \frac{\min(SETR_{RG}) + \max(SETR_{RG})}{2} \quad (A.1)$$

where SET_{RG} is the statistically equivalent RG CAD threshold and $SETR_{RG}$ is the range of statistically equivalent RG CAD thresholds.

An extract from the MATLAB script used to determine the statistically equivalent CAD threshold (Chpt3.RankPlotFigures.m) is shown in Figure A.1. This algorithm was used for all classes of CVD across all protocols.

```

35 deutan_cadRG = subj_cadRG(subj_simpleTypeOfCVD(:) == 2);
36
37 [~,placeholderForDeutanRGSorted]=sort(deutan_cadRG,'ascend');
38 [~,deutan_cadRGRanking]=sort(placeholderForDeutanRGSorted);
39
40 deutan_cadRGPrcolPass = deutan_cadRG(deutan_PrcolPassFail(:) ==
    1);
41 deutan_cadRGPrcolFail = deutan_cadRG(deutan_PrcolPassFail(:) ==
    0);
42
43 deutan_cadRGRankingPrcolPass = deutan_cadRGRanking(
    deutan_PrcolPassFail(:) == 1);
44 deutan_cadRGRankingPrcolFail = deutan_cadRGRanking(
    deutan_PrcolPassFail(:) == 0);
45
46 %Want to calculate statistically equivalent RG CAD threshold
47 deutan_statEquivRGCAD(:) = -1; count = 0;
48 for x = min(deutan_cadRG):0.01:max(deutan_cadRG)
49     if(sum(deutan_cadRGPrcolPass(:) > x) == sum(
        deutan_cadRGPrcolFail(:) < x))
50         count=count+1;
51         deutan_statEquivRGCAD(count) = x;
52     elseif x > 1 %check to see if there is an instant jump
53         if(sum(deutan_cadRGPrcolPass(:)>x)) > (sum(
            deutan_cadRGPrcolFail(:)<x)) && ...
54             (sum(deutan_cadRGPrcolPass(:)>(x+0.01))) < (sum(
                deutan_cadRGPrcolFail(:)<(x+0.01)))
55             count=count+1;
56             deutan_statEquivRGCAD(count) = x;
57         end
58     end
59 end
60 deutan_statEquivRGCADLower = min(deutan_statEquivRGCAD(1:count));
61 deutan_statEquivRGCADUpper = max(deutan_statEquivRGCAD(1:count));
62 deutan_statEquivRGCADMean = mean(deutan_statEquivRGCAD(1:count));

```

Figure A.1: Extract from Chpt3_RankPlotFigures.m.

Appendix B

Mathematical description of statistical terms

The **true** state of an individuals colour vision, t , and the diagnostic **outcome** of a given colour vision test, o , can be denoted as

$$t \in \{z, \bar{z}\} \tag{B.1}$$

$$o \in \{p, \bar{p}\} \tag{B.2}$$

where z is ‘normal’ colour vision, $\bar{z}$ is a colour vision deficiency (or ‘abnormal’ colour vision), p is a pass (a test diagnosis of ‘normal’ colour vision) and $\bar{p}$ is a fail (a diagnosis of CVD or ‘abnormal’ colour vision) (Rodriguez-Carmona 2006, Appendix C). Subsequently the prevalence of a colour vision deficiency is

$$\text{prevalence} = P(t = \bar{z}) \tag{B.3}$$

and the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of a test are defined as

$$\text{sensitivity} = P(o = \bar{p} \mid t = \bar{z}) \tag{B.4}$$

$$\text{specificity} = P(o = p \mid t = z) \tag{B.5}$$

$$\text{positive predictive value} = P(t = \bar{z} \mid o = \bar{p}) \tag{B.6}$$

$$\text{negative predictive value} = P(t = z \mid o = p) \tag{B.7}$$

Practical calculations

Definitions described by Altman (1991) for the ‘general representation of a diagnostic test based on a binary indicator’ adapted for colour vision assessment are shown in Tables B.1 and B.2.

	True state of an individuals colour vision (t)			
		‘Normal’ (z)	CVD ($\bar{z}$)	Σ
Diagnostic outcome of a colour vision test (o)	Pass (p)	d	c	$c + d$
	Fail ($\bar{p}$)	b	a	$a + b$
	Σ	$b + d$	$a + c$	n

Table B.1: General representation of a diagnostic colour vision test adapted from Altman (1991)

Diagnostic outcome (o)	True state (t)	Name
Fail	CVD	True positive (a)
Fail	‘Normal’	False positive (b)
Pass	CVD	False negative (c)
Pass	‘Normal’	True negative (d)

Table B.2: Names given to cells in Table B.1 (adapted from Altman (1991))

As described in Dekking et al. (2005) it is known that

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, \text{ if } P(B) \neq 0 \quad (\text{B.8})$$

As such, equations B.4-B.7 can be re-written and expressed in terms of the values shown in Tables B.1 and B.2

$$\text{sensitivity} = \frac{P(o = \bar{p} \cap t = \bar{z})}{P(t = \bar{z})} = \frac{a}{(a + c)} \quad (\text{B.9})$$

$$\text{specificity} = \frac{P(o = p \cap t = z)}{P(t = z)} = \frac{d}{(b + d)} \quad (\text{B.10})$$

$$\text{positive predictive value} = \frac{P(t = \bar{z} \cap o = \bar{p})}{P(o = \bar{p})} = \frac{a}{(a + b)} \quad (\text{B.11})$$

$$\text{negative predictive value} = \frac{P(t = z \cap o = p)}{P(o = p)} = \frac{d}{(c + d)} \quad (\text{B.12})$$

Altman (1991) also describes how one can rewrite equations B.6 and B.7 to calculate the positive and negative predictive values for different prevalences using Baynes' theorem

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)}, \quad P(B) \neq 0 \quad (\text{B.13})$$

By substituting equations B.6 and B.7 into Baynes' theorem (equation B.13) one finds that the positive and negative predictive values can be described as

$$PPV = P(t = \bar{z} | o = \bar{p}) = \frac{P(o = \bar{p} | t = \bar{z})P(t = \bar{z})}{P(o = \bar{p})} \quad (\text{B.14})$$

$$NPV = P(t = z | o = p) = \frac{P(o = p | t = z)P(t = z)}{P(o = p)} \quad (\text{B.15})$$

where $P(o = \bar{p})$ is the probability that the outcome of the colour vision test is a fail and $P(o = p)$ is the probability that the outcome of the colour vision test is a pass.

The probability that the outcome of a colour vision test is a fail is equal to the sum of the probability that the test is a fail given the subject has CVD multiplied by the prevalence of CVD and the probability that the test is a fail given the subject has normal colour vision multiplied by prevalence of normal colour vision.

Similarly, the probability that the outcome of a colour vision test is a pass is equal to the sum of the probability that the test is a pass given the subject has CVD multiplied by the prevalence of CVD and the probability that the test is a pass given the subject has normal colour vision multiplied by prevalence of normal colour vision.

Mathematically

$$P(o = \bar{p}) = [P(o = \bar{p} | t = \bar{z}) \times P(t = \bar{z})] + [P(o = \bar{p} | t = z) \times P(t = z)] \quad (\text{B.16})$$

$$P(o = p) = [P(o = p | t = \bar{z}) \times P(t = \bar{z})] + [P(o = p | t = z) \times P(t = z)] \quad (\text{B.17})$$

Substituting equation B.16 into B.14 and equation B.17 into B.15

$$PPV = \frac{P(o = \bar{p} | t = \bar{z})P(t = \bar{z})}{[P(o = \bar{p} | t = \bar{z}) \times P(t = \bar{z})] + [(P(o = \bar{p} | t = z) \times P(t = z))]} \quad (\text{B.18})$$

$$NPV = \frac{P(o = p | t = z)P(t = z)}{[P(o = p | t = \bar{z}) \times P(t = \bar{z})] + [(P(o = p | t = z) \times P(t = z))]} \quad (\text{B.19})$$

The positive and negative predictive values are now expressed in terms of the prevalence, sensitivity, and specificity and one can substitute equations B.3, B.4, and B.5 into equations B.18 and B.19

$$PPV = \frac{\text{sensitivity} \times \text{prevalence}}{(\text{sensitivity} \times \text{prevalence}) + [(1 - \text{specificity}) \times (1 - \text{prevalence})]} \quad (\text{B.20})$$

$$NPV = \frac{\text{specificity} \times (1 - \text{prevalence})}{[(1 - \text{sensitivity}) \times \text{prevalence}] + [\text{specificity} \times (1 - \text{prevalence})]} \quad (\text{B.21})$$

Appendix C

Supplementary figures

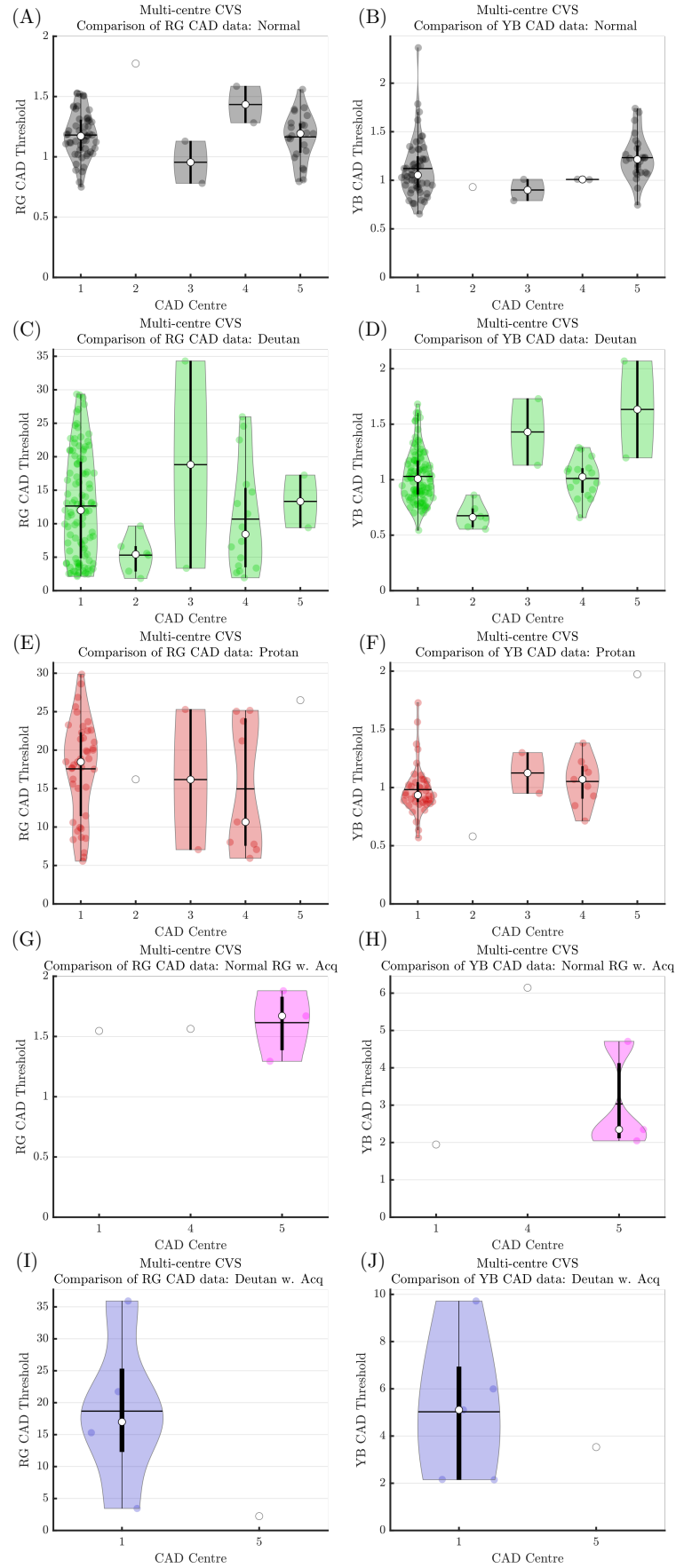


Figure C.1: Comparison of RG and YB CAD thresholds obtained at each CAD centre for participants assessed as part of the multi-centre CVS study.

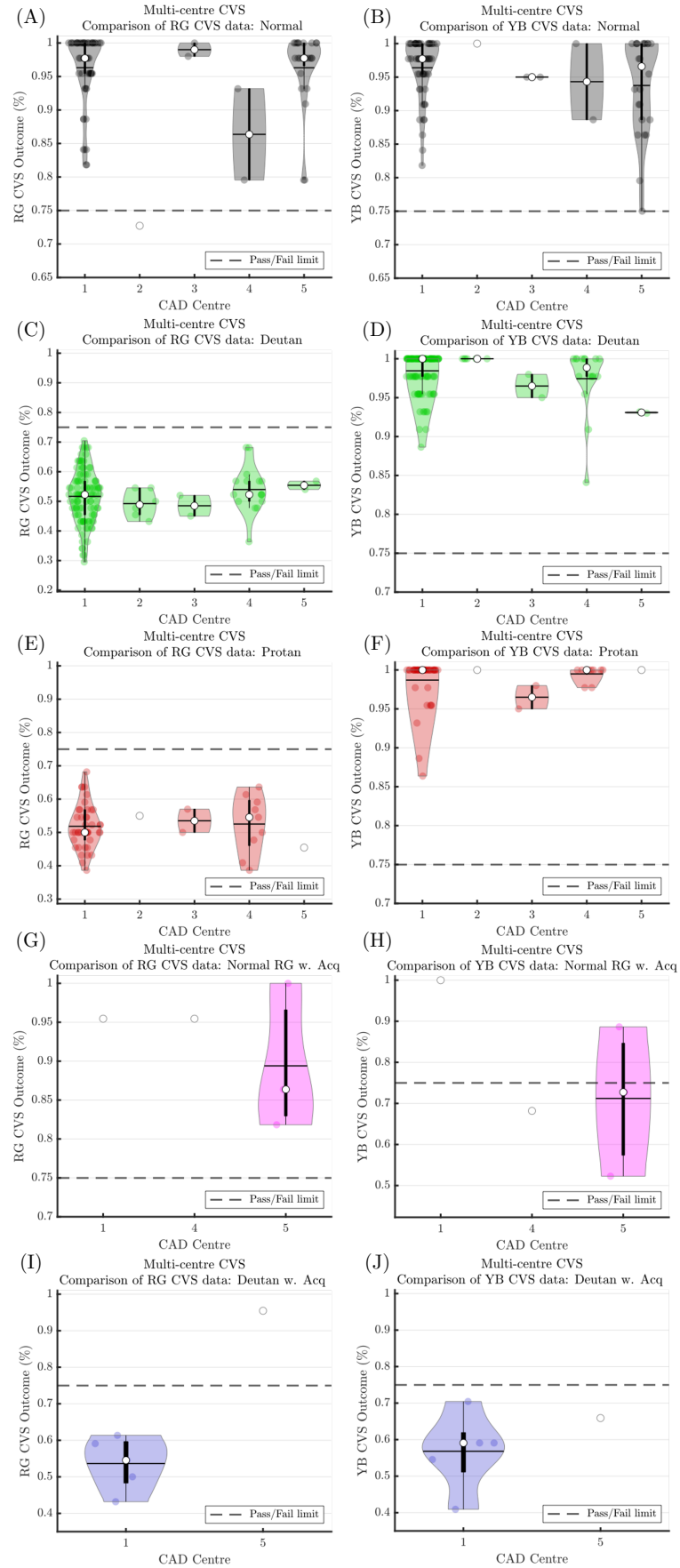


Figure C.2: Comparison of RG and YB CVS outcomes obtained at each CAD centre for participants assessed as part of the multi-centre CVS study.

References

- Aboshiha, J., Dubis, A. M., Carroll, J., Hardcastle, A. J. & Michaelides, M. (2016), ‘The cone dysfunction syndromes’, *British Journal of Ophthalmology* **100**(1), 115–121.
Available from: <https://bjoo.bmj.com/content/100/1/115>
- Abramovs, N., Brass, A. & Tassabehji, M. (2020), ‘Hardy-Weinberg Equilibrium in the Large Scale Genomic Sequencing Era’, *Frontiers in Genetics* **11**, 1–11.
Available from: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00210>
- Alfaro, A., Bernabeu, A., Agulló, C., Parra, J. & Fernández, E. (2015), ‘Hearing colors: an example of brain plasticity’, *Frontiers in Systems Neuroscience* **9**.
Available from: <https://www.frontiersin.org/articles/10.3389/fnsys.2015.00056>
- Allen, A. E., Martial, F. P. & Lucas, R. J. (2019), ‘Form vision from melanopsin in humans’, *Nature Communications* **10**(1).
Available from: <https://rdcu.be/cVtYf>
- Almustanyir, A., Hovis, J. & Glaholt, M. G. (2020), ‘Predicting the Farnsworth–Munsell D15 and Holmes–Wright-A lantern outcomes with computer-based color vision tests’, *J. Opt. Soc. Am. A* **37**(4), A1–A10.
Available from: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-37-4-A1>
- Almustanyir, A., Hovis, J. K., Alduhayan, R., Alhassan, M., Bokhary, K. & Alabdulkader, B. (2022), ‘Evaluation of the Third edition of the City University Colour Vision Test’, *Ophthalmic and Physiological Optics* **42**(1), 123–132.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/opo.12915>
- Altman, D. (1991), *Practical Statistics for Medical Research*, Chapman & Hall.
Available from: <https://books.google.it/books?id=v-walRnRxWQC>
- Andreani, A., Rosi, M. C., Guidi, R., Jafrancesco, D., Farini, A., Belcari, A. & Sacchetti, P. (2021), ‘Colour Preference of the Deer Ked *Lipoptena fortisetosa* (Diptera: Hippoboscidae)’, *Insects* **12**(9), 845.
Available from: <https://www.mdpi.com/2075-4450/12/9/845>
- Appel, L. J. (2006), ‘A primer on the design, conduct, and interpretation of clinical trials’, *Clinical Journal of the American Society of Nephrology*

- 1(6), 1360–1367.
Available from: <https://cjasn.asnjournals.org/content/1/6/1360>
- Arnegard, S., Baraas, R. C., Neitz, J., Hagen, L. A. & Neitz, M. (2022), 'Limitation of standard pseudoisochromatic plates in identifying colour vision deficiencies when compared with genetic testing', *Acta Ophthalmologica* **100**(7), 805–812.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/aos.15103>
- Artigas, J. M., Felipe, A., Navea, A., Fandiño, A. & Artigas, C. (2012), 'Spectral Transmission of the Human Crystalline Lens in Adult and Elderly Persons: Color and Total Transmission of Visible Light', *Investigative Ophthalmology & Visual Science* **53**(7), 4076–4084.
Available from: <https://doi.org/10.1167/iovs.12-9471>
- Atchison, D. A., Pedersen, C. A., Dain, S. J. & Wood, J. M. (2003), 'Traffic Signal Color Recognition Is a Problem for Both Protan and Deutan Color-Vision Deficients', *Human Factors: The Journal of the Human Factors and Ergonomics Society* **45**(3), 495–503.
Available from: <http://journals.sagepub.com/doi/10.1518/hfes.45.3.495.27247>
- Barbur, J. L. (2004), 'Double-blindsight' revealed through the processing of color and luminance contrast defined motion signals, in 'The roots of visual awareness: a festschrift in honour of Alan Cowey', Vol. 144 of *Progress in Brain Research*, Elsevier, pp. 243–259.
Available from: <https://www.sciencedirect.com/science/article/pii/S0079612303144172>
- Barbur, J. L., Birch, J. & Harlow, A. J. (1992), Threshold and Suprathreshold Responses to Chromatic Stimuli Using Psychophysical and Pupillometric Methods, in 'Noninvasive Assessment of the Visual System', Optica Publishing Group, p. SuC5.
Available from: <http://opg.optica.org/abstract.cfm?URI=NAVS-1992-SuC5>
- Barbur, J. L. & Connolly, D. M. (2011), 'Effects of hypoxia on color vision with emphasis on the mesopic range', *Expert Review of Ophthalmology* **6**(4), 409–420.
Available from: <https://doi.org/10.1586/eop.11.32>
- Barbur, J. L., Harlow, A. J. & Plant, G. T. (1994), 'Insights into the Different Exploits of Colour in the Visual Cortex', *Proceedings of the Royal Society B: Biological Sciences* **258**(1353), 327–334.
Available from: <http://rspb.royalsocietypublishing.org/content/258/1353/327.abstract>
- Barbur, J. L., Harlow, A. J. & Sahraie, A. (1992), 'Pupillary responses to stimulus structure, colour and movement', *Ophthalmic and Physiological Optics* **12**(2), 137–141.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.1992.tb00276.x>
- Barbur, J. L., Harlow, J. A., Plant, G. & Williams, C. (1993), Colour discrimination measurements in patients with cerebral achromatopsia, in

- 'Noninvasive Assessment of the Visual System', Optica Publishing Group, p. NTuB.3.
Available from: <http://opg.optica.org/abstract.cfm?URI=NAVS-1993-NTuB.3>
- Barbur, J. L., Holliday, I. E. & Ruddock, K. H. (1981), 'The spatial and temporal organisation of motion perception units in human vision.', *Acta psychologica* **48**(1-3), 35–47.
Available from: <https://psycnet.apa.org/record/1982-11230-001>
- Barbur, J. L., Konstantakopoulou, E., Rodriguez-Carmona, M., Harlow, J. A., Robson, A. G. & Moreland, J. D. (2010), 'The Macular Assessment Profile test – a new VDU-based technique for measuring the spatial distribution of the macular pigment, lens density and rapid flicker sensitivity', *Ophthalmic and Physiological Optics* **30**(5), 470–483.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.2010.00748.x>
- Barbur, J. L. & Rodriguez-Carmona, M. (2015), Color vision changes in normal aging, in A. J. Elliot, M. D. Fairchild & A. Franklin, eds, 'Handbook of Color Psychology', Cambridge Handbooks in Psychology, Cambridge University Press, chapter 8, p. 180–196.
Available from: <http://openaccess.city.ac.uk/12513/>
- Barbur, J. L. & Rodriguez-Carmona, M. (2017), 'Colour vision requirements in visually demanding occupations', *British Medical Bulletin* **122**(1), 51–77.
Available from: <https://doi.org/10.1093/bmb/ldx007>
- Barbur, J. L., Rodriguez-Carmona, M. & Evans, B. E. W. (2020), 'Color vision assessment-3. an efficient, two-step, color assessment protocol', *Color Research & Application* **46**(1), 33–45.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/col.22599>
- Barbur, J. L., Rodriguez-Carmona, M., Harlow, J., Mancuso, K., Neitz, J. & Neitz, M. (2008), 'A study of unusual Rayleigh matches in deutan deficiency', *Visual Neuroscience* **25**(3), 507–516.
Available from: <https://www.cambridge.org/core/journals/visual-neuroscience/article/abs/study-of-unusual-rayleigh-matches-in-deutan-deficiency/A9916127C3D1924500D9C373E4412263>
- Barbur, J. L., Rodriguez-Carmona, M., Hickey, J., Evans, S. & Chorley, A. (2016), 'Analysis of European colour vision certification requirements for air traffic control officers', *Civil Aviation Authority CAP* **1429**.
Available from: <https://publicapps.caa.co.uk/docs/33/CAP1429OCT16.pdf>
- Barbur, J. L. & Ruddock, K. H. (1980a), 'Spatial characteristics of movement detection mechanisms in human vision. I. Achromatic vision.', *Biological cybernetics* **37**(2), 77–92.
Available from: <https://link.springer.com/article/10.1007/BF00364247>
- Barbur, J. L. & Ruddock, K. H. (1980b), 'Spatial characteristics of movement detection mechanisms in human vision. II. Chromatic stimuli.', *Biological*

- cybernetics* **37**(2), 93–98.
Available from: <https://link.springer.com/article/10.1007/BF00364248>
- Barbur, J. L., Ruddock, K. H. & Waterfield, V. A. (1980), ‘Human visual responses in the absence of the geniculo-calcarine projection’, *Brain* **103**(4), 905–928.
Available from: <https://doi.org/10.1093/brain/103.4.905>
- Barbur, J. L. & Spang, K. (2008), ‘Colour constancy and conscious perception of changes of illuminant’, *Neuropsychologia* **46**(3), 853–863. Consciousness and Perception: Insights and Hindsight.
Available from: <https://www.sciencedirect.com/science/article/pii/S0028393207004101>
- Barbur, J., Rodriguez-Carmona, M., Evans, S. & Milburn, N. (2009), ‘Minimum Color Vision Requirements for Professional Flight Crew, Part 3: Recommendations for New Color Vision Standards’.
Available from: <https://libraryonline.erau.edu/online-full-text/faa-aviation-medicine-reports/AM09-11.pdf>
- Bargary, G. & Mitchell, K. J. (2008), ‘Synaesthesia and cortical connectivity’, *Trends in Neurosciences* **31**(7), 335–342.
Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0166223608001343>
- Barlow, H. (1982), ‘What causes trichromacy? A theoretical analysis using comb-filtered spectra’, *Vision Research* **22**(6), 635–643.
Available from: <https://www.sciencedirect.com/science/article/pii/0042698982900992>
- Barrionuevo, P. A., Filgueira, C. P. & Cao, D. (2022), ‘Is melanopsin activation affecting large field color-matching functions?’, *J. Opt. Soc. Am. A* **39**(6), 1104–1110.
Available from: <https://opg.optica.org/josaa/abstract.cfm?URI=josaa-39-6-1104>
- Bartlett, J. W. & Frost, C. (2008), ‘Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables’, *Ultrasound in Obstetrics & Gynecology* **31**(4), 466–475.
Available from: <https://obgyn.onlinelibrary.wiley.com/doi/abs/10.1002/uog.5256>
- Bechtold, B. (2016), ‘Violin Plots for Matlab’, <https://github.com/bastibe/Violinplot-Matlab>.
- Belcher, S. J., Greenshields, K. W. & Wright, W. D. (1958), ‘Colour vision survey’, *British Journal of Ophthalmology* **42**, 355 – 359.
Available from: <https://bjo.bmj.com/content/bjophthalmol/42/6/355.full.pdf>
- Bi, W., Gillespie-Gallery, H., Binns, A. & Barbur, J. L. (2016), ‘Flicker sensitivity in normal aging—monocular tests of retinal function at photopic and mesopic light levels’, *Investigative Ophthalmology and Visual Science* **57**(2), 387–395.
Available from: <https://doi.org/10.1167/iovs.15-16481>

- Bickerton, T. H. (1900), 'Remarks on the Holmgren Wool Test: Is it Adequate for the Detection of Colour Blindness?', *British Medical Journal* **1**(2046), 621–622.
Available from: <https://www.bmj.com/lookup/doi/10.1136/bmj.1.2046.621>
- Biersdorf, W. R. & Armington, J. C. (1957), 'Response of the human eye to sudden changes in the wavelength of stimulation', *J. Opt. Soc. Am.* **47**(3), 208–215.
Available from: <http://opg.optica.org/abstract.cfm?URI=josa-47-3-208>
- Birch, J. (1975), 'New diagnostic pseudoisochromatic plates for dichromats based on subtractive color matches', *Optometry and Vision Science* **52**(6), 398–404.
Available from: https://journals.lww.com/optvissci/Abstract/1975/06000/NEW_DIAGNOSTIC_PSEUDOISCHROMATIC_PLATES_FOR.4.aspx
- Birch, J. (1989), 'Use of the Farnsworth—Munsell 100-Hue test in the examination of congenital colour vision defects', *Ophthalmic and Physiological Optics* **9**(2), 156–162.
Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1475-1313.1989.tb00836.x>
- Birch, J. (1993), Classification of anomalous trichromatism with the nagel anomaloscope, in B. Drum, ed., 'Colour Vision Deficiencies XI: Proceedings of the eleventh Symposium of the International Research Group on Colour Vision Deficiencies, held in Sydney, Australia 21–23 June 1991 including the joint IRGCVD-AIC Meeting on Mechanisms of Colour Vision 24 June 1991', Springer Netherlands, Dordrecht, pp. 19–24.
Available from: https://doi.org/10.1007/978-94-011-1856-9_3
- Birch, J. (1997a), 'Clinical use of the City University Test (2nd Edition)', *Ophthalmic and Physiological Optics* **17**(6), 466–472.
Available from: <http://doi.wiley.com/10.1046/j.1475-1313.1997.9700029x.x>
- Birch, J. (1997b), 'Efficiency of the Ishihara test for identifying red-green colour deficiency', *Ophthalmic and Physiological Optics* **17**(5), 403–408.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1475-1313.1997.97000227.x>
- Birch, J. (1999), 'Performance of red-green color deficient subjects on the Holmes-Wright lantern (Type A) in photopic viewing.', *Aviation, space, and environmental medicine* **70**(9), 897–901.
- Birch, J. (2001), *Diagnosis of defective colour vision*, 2nd edn, Butterworth-Heinemann.
- Birch, J. (2003), Extreme Anomalous Trichromatism, in 'Normal and Defective Colour Vision', Oxford University Press, pp. 364–369.
Available from: https://www.researchgate.net/publication/284783625_Extreme_Anomalous_Trichromatism

- Birch, J. (2008a), 'Failure of concordance of the Farnsworth D15 test and the Nagel anomaloscope matching range in anomalous trichromatism', *Visual Neuroscience* **25**(3), 451–453.
Available from: shorturl.at/bfP19
- Birch, J. (2008b), 'Pass rates for the Farnsworth D15 colour vision test', *Ophthalmic and Physiological Optics* **28**(3), 259–264.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.2008.00566.x>
- Birch, J. (2008c), 'Performance of colour-deficient people on the Holmes-Wright lantern (type A): Consistency of occupational colour vision standards in aviation', *Ophthalmic and Physiological Optics* **28**(3), 253–258.
Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1475-1313.2008.00558.x>
- Birch, J. (2010), 'Identification of red-green colour deficiency: Sensitivity of the Ishihara and American Optical Company (Hard, Rand and Rittler) pseudo-isochromatic plates to identify slight anomalous trichromatism', *Ophthalmic and Physiological Optics* **30**(5), 667–671.
Available from: <https://onlinelibrary.wiley.com/doi/10.1111/j.1475-1313.2010.00770.x>
- Birch, J. (2012), 'Worldwide prevalence of red-green color deficiency', *J. Opt. Soc. Am. A* **29**(3), 313–320.
Available from: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-29-3-313>
- Birch, J., Barbur, J. L. & Harlow, A. J. (1992), 'New method based on random luminance masking for measuring isochromatic zones using high resolution colour displays', *Ophthalmic and Physiological Optics* **12**(2), 133–136.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.1992.tb00275.x>
- Birch, J. & Chisholm, C. M. (2008), 'Occupational colour vision requirements for police officers', *Ophthalmic and Physiological Optics* **28**(6), 524–531.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.2008.00605.x>
- Birch, J. & McKeever, L. M. (1993), 'Survey of the accuracy of new pseudoisochromatic plates', *Ophthalmic and Physiological Optics* **13**(1), 35–40.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.1993.tb00423.x>
- Birch, J. & Patel, N. (1995), Design and use of the holmgren wool test, in 'Colour Vision Deficiencies XII: Proceedings of the twelfth Symposium of the International Research Group on Colour Vision Deficiencies, held in Tübingen, Germany July 18–22, 1993', Springer Netherlands, Dordrecht, pp. 495–500.
Available from: https://doi.org/10.1007/978-94-011-0507-1_59
- Birch, J. & Platts, C. E. (1993), 'Colour vision screening in children: an evaluation of three pseudoisochromatic tests', *Ophthalmic and Physiological Optics* **13**(4), 344–349.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.1993.tb00489.x>

- Birch, J. & Rodriguez-Carmona, M. (2014), ‘Occupational color vision standards: new prospects’, *J. Opt. Soc. Am. A* **31**(4), A55–A59.
Available from: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-31-4-A55>
- Blackwell, H. R. (1952), ‘Studies of psychophysical methods for measuring visual thresholds*†’, *J. Opt. Soc. Am.* **42**(9), 606–616.
Available from: <http://opg.optica.org/abstract.cfm?URI=josa-42-9-606>
- Bland, J. & Altman, D. (1986), ‘Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement’, *Lancet* **327**, 307–310.
Available from: <http://www.sciencedirect.com/science/article/pii/S0140673686908378>
- Bonnardel, V. (1998), ‘Demonstration of a new stimulator for human colour vision testing’, *The Journal of Physiology* **506**(31), 13P.
Available from: https://www.researchgate.net/publication/295539119_Demonstration_of_a_new_stimulator_for_human_colour_vision_testing
- Bonnardel, V. (2021), Colour, human experience and cyborgism, in ‘Proceedings of the International Colour Association (AIC) Conference 2021’, pp. 1397–1401.
Available from: https://aic-color.org/resources/Documents/Proceedings_AIC2021_r10.pdf
- Bonnardel, V., Barlow, H. B. & Ruderman, D. L. (1997), Fast determination of the spectral modulation sensitivity function: a comparison between trichromats and deuteranopes, in C. R. Cavonius, ed., ‘Colour Vision Deficiencies XIII: Proceedings of the thirteenth Symposium of the International Research Group on Colour Vision Deficiencies, held in Pau, France July 27–30, 1995’, Springer Netherlands, Dordrecht, pp. 415–424.
Available from: https://doi.org/10.1007/978-94-011-5408-6_47
- Bonnardel, V., Bellemare, H. & Mollon, J. D. (1996), ‘Measurements of Human Sensitivity to Comb-filtered Spectra’, *Vision Research* **36**(17), 2713–2720.
Available from: <https://www.sciencedirect.com/science/article/pii/0042698996000144>
- Bonnardel, V., Biddington, T., May, B., Jones, R. & Roffey, S. (2018), ‘Toward the Implementation of Contemplative Practices in Higher Education’, *Journal of Perspectives in Applied Academic Practice* **6**(3), 3–13.
Available from: <https://jpaap.ac.uk/JPAAP/article/view/362>
- Bonnardel, V., Martinoya, C. & Varela, F. (1991), ‘Method for measuring human sensitivity to periodic spectral energy modulations’, *Comptes rendus de l’Académie des sciences. Série III, Sciences de la vie* **312**(13), 695–700.
Available from: https://www.researchgate.net/publication/21243233_Method_for_measuring_sensitivity_to_periodic_spectral_energy_modulations_in_human
- Bonnardel, V. & Valero, E. M. (2001), ‘Study of colour discrimination with comb-filtered spectra’, *Vision Research* **41**(5), 541–548.
Available from: <https://www.sciencedirect.com/science/article/pii/S0042698900002789>

- Bonnardel, V. & Varela, F. J. (1991), 'A frequency view of colour: measuring the human sensitivity to square-wave spectral power distributions', *Proceedings of the Royal Society of London. Series B: Biological Sciences* **245**(1314), 165–171.
Available from: <https://royalsocietypublishing.org/doi/abs/10.1098/rspb.1991.0104>
- Bonnardel, V. & Varela, F. J. (2003), 'Color vision in the comb frequency domain.', *Biological research* **36** **1**, 119–134.
Available from: <https://doi.org/10.4067/S0716-97602003000100011>
- Brang, D. & Ramachandran, V. S. (2011), 'Survival of the Synesthesia Gene: Why Do People Hear Colors and Taste Words?', *PLOS Biology* **9**(11), e1001205.
Available from: <https://dx.plos.org/10.1371/journal.pbio.1001205>
- Brindley, G. S. (1970), 'Sensations produced by electrical stimulation of the occipital poles of the cerebral hemispheres, and their use in constructing visual prostheses.', *Annals of the Royal College of Surgeons of England* **47**(2), 106–108.
Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2387784/>
- Brindley, G. S. & Lewin, W. S. (1968), 'The sensations produced by electrical stimulation of the visual cortex', *The Journal of Physiology* **196**(2), 479–493.
Available from: <https://onlinelibrary.wiley.com/doi/10.1113/jphysiol.1968.sp008519>
- Calkins, D. J., Tsukamoto, Y. & Sterling, P. (1998), 'Microcircuitry and mosaic of a blue–yellow ganglion cell in the primate retina', *Journal of Neuroscience* **18**(9), 3373–3385.
Available from: <https://www.jneurosci.org/content/18/9/3373>
- Cao, D., Chang, A. & Gai, S. (2018), 'Evidence for an impact of melanopsin activation on unique white perception', *J. Opt. Soc. Am. A* **35**(4), B287–B291.
Available from: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-35-4-B287>
- 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.
Available from: <https://doi.org/10.1167/2.8.1>
- Carter, T. & Barbur, J. (2015), Colour vision assessment for maritime navigational lookout : Review for UK Maritime and Coastguard Agency (MCA), Technical Report July 2015, Norwegian Centre for Maritime Medicine, Bergen and Applied Vision Research Centre, City University London.
Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/527470/FINAL_report_Colour_vision_testing.pdf
- Cheng, A., Kessler, D., Mackinnon, R., Chang, T. P., Nadkarni, V. M., Hunt, E. A., Duval-Arnould, J., Lin, Y., Pusic, M. & Auerbach, M. (2017), 'Conducting multicenter research in healthcare simulation: Lessons learned from

- the INSPIRE network', *Advances in Simulation* **2**(1), 6.
Available from: <https://doi.org/10.1186/s41077-017-0039-0>
- Chorley, A. C. (2015), 'A Modified Protocol for Color Vision Screening Using Ishihara', *Aerospace Medicine and Human Performance* **86**(8), 747–751.
Available from: <http://openurl.ingenta.com/content/xref?genre=article&issn=2375-6314&volume=86&issue=8&spage=747>
- Christ, T. J. & Herriges, M. (2018), Screening Tests, in B. B. Frey, ed., 'The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation', SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks, California 91320, pp. 1485–1487.
Available from: <https://dx.doi.org/10.4135/9781506326139.n616>
- Cole, B. L. (2004), 'The handicap of abnormal colour vision', *Clinical and Experimental Optometry* **87**(4-5), 258–275.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1444-0938.2004.tb05056.x>
- Cole, B. L. (2007), 'Assessment of inherited colour vision defects in clinical practice', *Clinical and Experimental Optometry* **90**(3), 157–175.
Available from: <https://doi.org/10.1111%2Fj.1444-0938.2007.00135.x>
- Cole, B. L. & Maddocks, J. D. (1998), 'Can clinical colour vision tests be used to predict the results of the Farnsworth lantern test?', *Vision Research* **38**(21), 3483–3485.
Available from: <https://www.sciencedirect.com/science/article/pii/S0042698998001199>
- Cranwell, M. B., Pearce, B., Loveridge, C. & Hurlbert, A. C. (2015), 'Performance on the Farnsworth-Munsell 100-Hue Test Is Significantly Related to Nonverbal IQ', *Investigative Ophthalmology and Visual Science* **56**(5), 3171–3178.
Available from: <https://doi.org/10.1167/iovs.14-16094>
- Crognale, M. A., Switkes, E., Rabin, J., Schneck, M. E., Hægerström-Portnoy, G. & Adams, A. J. (1993), 'Application of the spatiochromatic visual evoked potential to detection of congenital and acquired color-vision deficiencies', *J. Opt. Soc. Am. A* **10**(8), 1818–1825.
Available from: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-10-8-1818>
- 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', *Journal of Comparative Neurology* **312**(4), 610–624.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cne.903120411>
- Curcio, C. A., Sloan, K. R., Kalina, R. E. & Hendrickson, A. E. (1990), 'Human photoreceptor topography', *Journal of Comparative Neurology* **292**(4), 497–523.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cne.902920402>

- Dacey, D. M. (2000), 'Parallel pathways for spectral coding in primate retina', *Annual Review of Neuroscience* **23**(1), 743–775. PMID: 10845080.
Available from: <https://doi.org/10.1146/annurev.neuro.23.1.743>
- Dacey, D. M. & Lee, B. B. (1994), 'The 'blue-on' opponent pathway in primate retina originates from a distinct bistratified ganglion cell type', *Nature* **367**(6465), 731–735.
Available from: <https://doi.org/10.1038/367731a0>
- Dain, S. J. (2004), 'Clinical colour vision tests', *Clinical and Experimental Optometry* **87**(4-5), 276–293.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1444-0938.2004.tb05057.x>
- Dain, S. J. & Adams, A. J. (1990), 'Comparison of the standard and Adams desaturated D-15 tests with congenital colour vision deficiencies', *Ophthalmic and Physiological Optics* **10**(1), 40–45.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.1990.tb01105.x>
- Dain, S. J., Atchison, D. A. & Hovis, J. K. (2019), 'Limitations and Precautions in the Use of the Farnsworth-Munsell Dichotomous D-15 Test', *Optometry and Vision Science* **96**(9), 695–705.
Available from: https://journals.lww.com/optvissci/Abstract/2019/09000/Limitations_and_Precautions_in_the_Use_of_the.10.aspx
- Dain, S. J., Casolin, A., Long, J. & Hilmi, M. R. (2015), 'Color Vision and the Railways: Part 1. The Railway LED Lantern Test', *Optometry and Vision Science* **92**(2), 138–146.
Available from: <https://journals.lww.com/00006324-201502000-00005>
- Dain, S. J. & Ling, B. Y. (2009), 'Cognitive Abilities of Children on a Gray Seriation Test', *Optometry and Vision Science* **86**(6), E701–E707.
Available from: <https://journals.lww.com/00006324-200906000-00029>
- Dain, S. J., Wood, J. M. & Atchison, D. A. (2009), 'Sunglasses, Traffic Signals, and Color Vision Deficiencies', *Optometry and Vision Science* **86**(4), E296–E305.
Available from: <https://journals.lww.com/00006324-200904000-00003>
- Dalton, J. (1798), 'Extraordinary facts relating to the vision of colours: with observations', *Memoirs of the Literary and Philosophical Society of Manchester* **5**, 28–45.
Available from: <https://www.biodiversitylibrary.org/part/308502>
- Danilova, M., Chan, C. & Mollon, J. (2013), 'Can spatial resolution reveal individual differences in the L:M cone ratio?', *Vision Research* **78**, 26–38.
Available from: <https://www.sciencedirect.com/science/article/pii/S0042698912003963>
- Das, M. K. (2022), 'Multicenter Studies: Relevance, Design and Implementation', *Indian Pediatrics* **59**(7), 571–579.

- Available from: <https://link.springer.com/article/10.1007/s13312-022-2561-y>
- Davidoff, C., Neitz, M. & Neitz, J. (2016), 'Genetic Testing as a New Standard for Clinical Diagnosis of Color Vision Deficiencies', *Translational Vision Science & Technology* **5**(5), 2–2.
Available from: <https://doi.org/10.1167/tvst.5.5.2>
- Davies, A., Pierce, W., Stephenson, A. & Myers, C. (1934), *The Selection of Colour Workers: Being a Research Into the Practical Methods of Measuring the Ability to Discriminate Fine Shades of Colour*, Sir I. Pitman & Sons, Limited.
Available from: https://books.google.co.uk/books?id=54G_wQEACAAJ
- 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.
Available from: <https://onlinelibrary.wiley.com/doi/10.1113/jphysiol.1975.sp011086>
- Dekking, F., Kraaikamp, C., Lopuhaä, H. & Meester, L. (2005), *A Modern Introduction to Probability and Statistics: Understanding Why and How*, Springer Texts in Statistics, Springer.
Available from: <https://books.google.co.uk/books?id=XLUMIlombgQC>
- DeMarco, P., Pokorny, J. & Smith, V. C. (1992), 'Full-spectrum cone sensitivity functions for x-chromosome-linked anomalous trichromats', *J. Opt. Soc. Am. A* **9**(9), 1465–1476.
Available from: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-9-9-1465>
- Derrington, A. M., Krauskopf, J. & Lennie, P. (1984), 'Chromatic mechanisms in lateral geniculate nucleus of macaque.', *The Journal of Physiology* **357**(1), 241–265.
Available from: <https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1984.sp015499>
- Dewey, J. (1933), *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*, College S, D.C. Heath.
- Do, M. T. H. & Yau, K.-W. (2010), 'Intrinsically photosensitive retinal ganglion cells', *Physiological Reviews* **90**(4), 1547–1581. PMID: 20959623.
Available from: <https://doi.org/10.1152/physrev.00013.2010>
- Dodt, E., Copenhaver, R. M. & Gunkel, R. D. (1958), 'Photopischer Dominator und Farbkomponenten im menschlichen Elektroretinogramm', *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere* **267**(5), 497–507.
Available from: <https://doi.org/10.1007/BF00361736>
- Edwards, A. W. F. (2008), 'G. H. Hardy (1908) and Hardy–Weinberg Equilibrium', *Genetics* **179**(3), 1143–1150.
Available from: <https://doi.org/10.1534/genetics.104.92940>
- El-Alti, L., Sandman, L. & Munthe, C. (2019), 'Person Centered Care and Personalized Medicine: Irreconcilable Opposites or Potential Companions?', *Health Care Analysis* **27**(1), 45–59.

- Available from: <http://link.springer.com/10.1007/s10728-017-0347-5>
- Estévez, O. & Spekreijse, H. (1982), 'The "silent substitution" method in visual research', *Vision Research* **22**(6), 681–691.
Available from: <https://www.sciencedirect.com/science/article/pii/0042698982901043>
- Estévez, O., Spekreijse, H., Van Dalen, J. T. & Verduyn Lunel, H. F. (1983), 'The Oscar Color Vision Test: Theory and Evaluation (Objective Screening of Color Anomalies and Reductions)', *Optometry and Vision Science* **60**(11), 892–901.
Available from: <https://doi.org/10.1097/00006324-198311000-00004>
- Estévez, O. & Spekreijse, H. (1974), 'A spectral compensation method for determining the flicker characteristics of the human colour mechanisms', *Vision Research* **14**(9), 823–830.
Available from: <https://www.sciencedirect.com/science/article/pii/0042698974901473>
- Evans, A. (2003), *Color Is in the Eye of the Beholder: A Guide to Color Vision Deficiency and Colorblindness*, CVD Publishing.
Available from: <https://books.google.it/books?id=5hG3Z8EshyQC>
- Evans, B. E. W., Rodriguez-Carmona, M. & Barbur, J. L. (2020), 'Color vision assessment-1: Visual signals that affect the results of the Farnsworth D-15 test', *Color Research & Application* **46**(1), 7–20.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/col.22596>
- Fanlo Zarazaga, A., Gutiérrez Vázquez, J. & Pueyo Royo, V. (2019), 'Review of the main colour vision clinical assessment tests', *Archivos de la Sociedad Española de Oftalmología (English Edition)* **94**(1), 25–32.
Available from: <https://www.sciencedirect.com/science/article/pii/S2173579418301828>
- Farini, A., Baldanzi, E., Raffaelli, M. & Russo, F. (2019), 'Evolution of colours in football shirts through colorimetric measurements: Fiorentina's case.', *Cultura e Scienza del Colore - Color Culture and Science* **11**(01), 41–46.
Available from: <https://jcolore.gruppodelcolore.it/ojs/index.php/CCSJ/article/view/CCSJ.110105>
- Farnsworth, D. (1943), 'The Farnsworth-Munsell 100-Hue and Dichotomous Tests for Color Vision*', *J. Opt. Soc. Am.* **33**(10), 568–578.
Available from: <http://opg.optica.org/abstract.cfm?URI=josa-33-10-568>
- Farnsworth, D. (1947), *Farnsworth dichotomous test for color blindness.*, Psychological Corporation, San Antonio, Texas.
- Farnsworth, D. (1957), 'Testing for color deficiency in industry.', *A.M.A. archives of industrial health* **16**(2), 100–3.
- Feitosa-Santana, C., Paramei, G. V., Nishi, M., Gualtieri, M., Costa, M. F. & Ventura, D. F. (2010), 'Color vision impairment in type 2 diabetes assessed by the D-15d test and the Cambridge Colour Test', *Ophthalmic and*

- Physiological Optics* **30**(5), 717–723.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.2010.00776.x>
- Field, G. D., Gauthier, J. L., Sher, A., Greschner, M., Machado, T. A., Jepson, L. H., Shlens, J., Gunning, D. E., Mathieson, K., Dabrowski, W., Paninski, L., Litke, A. M. & Chichilnisky, E. J. (2010), ‘Functional connectivity in the retina at the resolution of photoreceptors’, *Nature* **467**(7316), 673–677.
Available from: <http://dx.doi.org/10.1038/nature09424>
- Fletcher, R. (1972), ‘Methods of examination of macular colour vision. A modified D-15 test.’, *Modern problems in ophthalmology* **11**, 22–24.
- Fletcher, R. (1978), ‘Recent experiences with the City spot test.’, *Modern problems in ophthalmology* **19**, 142–143.
- Forbes, A., Burleigh, S. & Neyland, M. (1955), ‘Electric responses to color shift in frog and turtle retina’, *Journal of Neurophysiology* **18**(6), 517–535.
PMID: 13272038.
Available from: <https://doi.org/10.1152/jn.1955.18.6.517>
- Foster, D. H. (2010), Chromatic Function of the Cones, in D. A. Dartt, ed., ‘Encyclopedia of the Eye’, Academic Press, Oxford, pp. 266–274.
Available from: <https://www.sciencedirect.com/science/article/pii/B9780123742032002323>
- Foster, D. H. (2011), ‘Color constancy’, *Vision Research* **51**(7), 674–700. Vision Research 50th Anniversary Issue: Part 1.
Available from: <https://www.sciencedirect.com/science/article/pii/S0042698910004402>
- Fraunfelder, F. & Fraunfelder, F. (2020), *Drug-Induced Ocular Side Effects*, 8th edn, Elsevier Health Sciences.
Available from: <https://doi.org/10.1016/C2018-0-00445-4>
- French, A. N., Rose, K. A., Cornell, E. D. & Thompson, K. (2008), ‘The evolution of colour vision testing’, *Australian Orthoptic Journal* **40**, 7.
Available from: <https://www.aojournal.com.au/static/uploads/files/aoj20084009-wfvkfydgyhy.pdf>
- Gegenfurtner, K. R. (2003), ‘Cortical mechanisms of colour vision’, *Nature Reviews Neuroscience* **4**(7), 563–572.
Available from: <http://www.nature.com/articles/nrn1138>
- Gegenfurtner, K. R. & Rieger, J. (2000), ‘Sensory and cognitive contributions of color to the recognition of natural scenes’, *Current Biology* **10**(13), 805–808.
Available from: <https://www.sciencedirect.com/science/article/pii/S0960982200005637>
- Gegenfurtner, K., Sharpe, L. & Boycott, B. (2001), *Color Vision: From Genes to Perception*, Cambridge University Press.
Available from: <https://books.google.com/books?id=4zQMQLLVkFYC>

- Gerling, J., Meigen, T. & Bach, M. (1997), 'Shift of Equiluminance in Congenital Color Vision Deficiencies: Pattern-ERG, VEP and Psychophysical Findings', *Vision Research* **37**(6), 821–826.
Available from: <https://www.sciencedirect.com/science/article/pii/S0042698996001617>
- Ghose, S., Shrey, D., Venkatesh, P., Parmar, T. & Sharma, S. (2014), 'A simple modification of the Farnsworth-Munsell 100-Hue test for much faster assessment of color vision', *Indian Journal of Ophthalmology* **62**(6), 721.
Available from: <https://journals.lww.com/10.4103/0301-4738.129778>
- Goh, S. S. M., Chan, V. X. B. & Tan, N. C. (2014), 'Colour Vision Deficiency: Is it a Handicap? A Narrative Review of its Impact on Medical & Dental Education and Practice', *Proceedings of Singapore Healthcare* **23**(2), 149–157.
Available from: <https://doi.org/10.1177/201010581402300209>
- Goodman, D. M. & Livingston, E. H. (2013), 'Screening Tests', *JAMA* **309**(11), 1185–1185.
Available from: <https://doi.org/10.1001/jama.2012.144651>
- Grünert, U. & Martin, P. R. (2020), 'Cell types and cell circuits in human and non-human primate retina', *Progress in Retinal and Eye Research* **78**, 100844.
Available from: <https://www.sciencedirect.com/science/article/pii/S1350946220300161>
- Guo, L., Choi, S., Bikkannavar, P. & Cordeiro, M. F. (2022), 'Microglia: Key Players in Retinal Ageing and Neurodegeneration', *Frontiers in Cellular Neuroscience* **16**.
Available from: <https://www.frontiersin.org/articles/10.3389/fncel.2022.804782>
- Hardy, L. H., Rand, G. & Rittler, M. C. (1945), 'Tests for the Detection and Analysis of Color-Blindness. I. The Ishihara Test: An Evaluation', *J. Opt. Soc. Am.* **35**(4), 268–275.
Available from: <http://opg.optica.org/abstract.cfm?URI=josa-35-4-268>
- Hardy, L. H., Rand, G. & Rittler, M. C. (1954), 'The H-R-R Polychromatic Plates: I. A Test for the Detection, Classification, and Estimation of the Degree of Defective Color Vision', *A.M.A. Archives of Ophthalmology* **51**(2), 216–228.
Available from: <https://doi.org/10.1001/archopht.1954.00920040218009>
- Harris, R. W. & Cole, B. L. (2005), 'Diagnosing protan heterozygosity using the Medmont C-100 colour vision test', *Clinical and Experimental Optometry* **88**(4), 240–247.
Available from: <https://doi.org/10.1111/j.1444-0938.2005.tb06702.x>
- Hasrod, N. & Rubin, A. (2015), 'Colour vision: A review of the Cambridge Colour Test and other colour testing methods', *African Vision and Eye Health* **74**(1), 7.

- Available from: <https://avehjournal.org/index.php/aveh/article/view/23>
- Helve, J. H. (1972), 'A comparative study of several diagnostic tests of colour vision used for measuring types and degrees of congenital red-green defects.', *Acta Ophthalmol. Suppl.* **115**, 3–64.
- Hickey, J. D. (2015), Studies of normal and deficient colour vision with relevance to occupational environments, PhD thesis, City University London. [Unpublished Doctoral thesis].
Available from: <https://openaccess.city.ac.uk/id/eprint/15009/>
- Hintze, J. L. & Nelson, R. D. (1998), 'Violin plots: A box plot-density trace synergism', *The American Statistician* **52**(2), 181–184.
Available from: <https://www.tandfonline.com/doi/abs/10.1080/00031305.1998.10480559>
- Hofer, H., Carroll, J., Neitz, J., Neitz, M. & Williams, D. R. (2005), 'Organization of the Human Trichromatic Cone Mosaic', *Journal of Neuroscience* **25**(42), 9669–9679.
Available from: <https://www.jneurosci.org/content/25/42/9669>
- Holmes, J. G. & Wright, W. D. (1982), 'A new colour-perception lantern', *Color Research & Application* **7**(2), 82–88.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/col.5080070204>
- Holmgren, F. (1877), *Color-blindness in its relation to accidents by rail and sea*, Washington : Government Printing Office.
Available from: <http://resource.nlm.nih.gov/101703557>
- Honson, V. J. & Dain, S. J. (1987), 'Analysis of the Mark II Edition of The City University Colour Vision Test', *Optometry and Vision Science* **64**(4).
Available from: https://journals.lww.com/optvissci/Fulltext/1987/04000/Analysis_of_the_Mark_II_Edition_of_The_City.9.aspx
- Hovis, J. K. & Almustanyir, A. (2017), Assessment of the Next Generation of Colour Vision Tests for Pilots and Aircrew, Technical Report DRD C-RDDC-20 017-C191, Defence Research and Development Canada.
- Hovis, J. K. & Oliphant, D. (1998), 'Validity of the Holmes-Wright lantern as a color vision test for the rail industry', *Vision Research* **38**(21), 3487–3491.
Available from: <https://www.sciencedirect.com/science/article/pii/S0042698998000546>
- Hupé, J.-M. & Dojat, M. (2015), 'A critical review of the neuroimaging literature on synesthesia', *Frontiers in Human Neuroscience* **9**.
Available from: <https://www.frontiersin.org/articles/10.3389/fnhum.2015.00103>
- Hurvich, L. M. (1972), Color vision deficiencies, in D. Jameson & L. M. Hurvich, eds, 'Visual Psychophysics', Springer Berlin Heidelberg, Berlin, Heidelberg, chapter 23, pp. 582–624.
Available from: https://doi.org/10.1007/978-3-642-88658-4_23

- Iregren, A., Andersson, M. & Nylén, P. (2002), 'Color Vision and Occupational Chemical Exposures: I. An Overview of Tests and Effects', *Neuro-Toxicology* **23**(6), 719–733.
Available from: <https://www.sciencedirect.com/science/article/pii/S0161813X02000888>
- Jacobs, G. H. & Neitz, J. (1993), ERG flicker photometric evaluation of spectral sensitivity in protanopes and protanomalous trichromats, in B. Drum, ed., 'Colour Vision Deficiencies XI', Springer Netherlands, Dordrecht, pp. 25–31.
Available from: https://doi.org/10.1007/978-94-011-1856-9_4
- Jacobs, G. H., Neitz, J. & Krogh, K. (1996), 'Electroretinogram flicker photometry and its applications', *J. Opt. Soc. Am. A* **13**(3), 641–648.
Available from: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-13-3-641>
- Jägle, H., Pirzer, M. & Sharpe, L. T. (2005), 'The Nagel anomaloscope: its calibration and recommendations for diagnosis and research', *Graefe's Archive for Clinical and Experimental Ophthalmology* **243**(1), 26–32.
Available from: <http://link.springer.com/10.1007/s00417-004-0893-z>
- Jägle, H., Sadowski, B., Kremers, J., Scholl, H. P., Leo-Kottler, B. & Sharpe, L. T. (2003), 'Detecting color vision in a malingerer', *Documenta Ophthalmologica* **106**(2), 121–128.
Available from: <https://rdcu.be/cVFpc>
- Jennings, B. J. & Barbur, J. L. (2010), 'Colour detection thresholds as a function of chromatic adaptation and light level', *Ophthalmic and Physiological Optics* **30**(5), 560–567.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.2010.00773.x>
- Jiang, X. & Mahroo, O. A. (2022), 'Human retinal dark adaptation tracked *in vivo* with the electroretinogram: insights into processes underlying recovery of cone- and rod-mediated vision', *The Journal of Physiology* **600**(21), 4603–4621.
Available from: <https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/JP283105>
- Jolly, J. K., Simunovic, M. P., Dubis, A. M., Josan, A. S., Robson, A. G., Bellini, M. P., Bloch, E., Georgiadis, O., da Cruz, L., Bridge, H. & MacLaren, R. E. (2021), 'Structural and Functional Characteristics of Color Vision Changes in Choroideremia', *Frontiers in Neuroscience* **15**.
Available from: <https://www.frontiersin.org/articles/10.3389/fnins.2021.729807>
- Jordan, G. & Mollon, J. D. (1993), 'The Nagel anomaloscope and seasonal variation of colour vision', *Nature* **363**(6429), 546–549.
Available from: <http://www.nature.com/articles/363546a0>
- 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: Proceedings of the thirteenth Symposium of the International Research Group on Colour

- Vision Deficiencies, held in Pau, France July 27–30, 1995', Springer Netherlands, Dordrecht, pp. 385–392.
Available from: https://doi.org/10.1007/978-94-011-5408-6_42
- Jurasevska, K., Ozolinsh, M., Fomins, S., Gutmane, A., Zutere, B., Pausus, A. & Karitans, V. (2014), 'Color-discrimination threshold determination using pseudoisochromatic test plates', *Frontiers in Psychology* **5**(November), 1–7.
Available from: <http://journal.frontiersin.org/article/10.3389/fpsyg.2014.01376/abstract>
- Kaiser, P. K. & Boynton, R. M. (1996), *Human Color Vision*, 2nd edn, Optical Society of America.
Available from: <https://books.google.co.uk/books?id=c1NqAAAAAMAAJ>
- Kandel, E. R., Koester, J. D., Mack, S. H. & Siegelbaum, S. A. (2021), *Principles of Neural Science, Sixth Edition*, McGraw-Hill Education.
Available from: <https://books.google.co.uk/books?id=IYoEEAAAQBAJ>
- Karampatakis, V., Almaliotis, D., Karamitopoulos, L., Kalliris, G. & Almpandidou, S. (2022), 'A Novel Smartphone-Based Color Test for Detection of Color Vision Defects in Age Related Macular Degeneration', *Journal of Ophthalmology* **2022**, 1–9.
Available from: <https://doi.org/10.1155/2022/9744065>
- Kaufman, P., Alm, A. & Adler, F. (2003), *Adler's Physiology of the Eye: Clinical Application*, Mosby.
Available from: <https://books.google.co.uk/books?id=2YlsAAAAAMAAJ>
- Kinnear, P. R. & Sahraie, A. (2002), 'New Farnsworth-Munsell 100 hue test norms of normal observers for each year of age 5-22 and for age decades 30-70', *British Journal of Ophthalmology* **86**(12), 1408–1411.
Available from: <http://bjo.bmj.com/cgi/doi/10.1136/bjo.86.12.1408>
- Klug, K., Herr, S., Ngo, I. T., Sterling, P. & Schein, S. (2003), 'Macaque Retina Contains an S-Cone OFF Midget Pathway', *Journal of Neuroscience* **23**(30), 9881–9887.
Available from: <https://www.jneurosci.org/content/23/30/9881>
- Knoblauch, K., Vital-Durand, F. & Barbur, J. L. (2001), 'Variation of chromatic sensitivity across the life span', *Vision Research* **41**(1), 23–36.
Available from: <https://www.sciencedirect.com/science/article/pii/S0042698900002054>
- Koefoed, V. F., Miles, T., Cason, J. B. & Troche, R. (2020), 'Colour vision classification – comparing CAD and CIE 143:2001 International recommendations for colour vision requirements in transport', *Acta Ophthalmologica* **98**(7), 726–735.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/aos.14450>
- Kohl, S., Baumann, B., Broghammer, M., Jägle, H., Sieving, P., Kellner, U., Spiegel, R., Anastasi, M., Zrenner, E., Sharpe, L. T. & Wissinger, B. (2000), 'Mutations in the CNGB3 gene encoding the β -subunit of the cone photoreceptor cGMP-gated channel are responsible for achromatopsia (ACHM3)

- linked to chromosome 8q21', *Human Molecular Genetics* **9**(14), 2107–2116.
Available from: <https://doi.org/10.1093/hmg/9.14.2107>
- Kohl, S., Varsanyi, B., Antunes, G. A., Baumann, B., Hoyng, C. B., Jägle, H., Rosenberg, T., Kellner, U., Lorenz, B., Salati, R., Jurklies, B., Farkas, A., Andreasson, S., Weleber, R. G., Jacobson, S. G., Rudolph, G., Castellán, C., Dollfus, H., Legius, E., Anastasi, M., Bitoun, P., Lev, D., Sieving, P. A., Munier, F. L., Zrenner, E., Sharpe, L. T., Cremers, F. P. M. & Wissinger, B. (2005), 'CNGB3 mutations account for 50% of all cases with autosomal recessive achromatopsia', *European Journal of Human Genetics* **13**(3), 302–308.
Available from: <http://www.nature.com/articles/5201269>
- Kolb, H. (2003), 'How the Retina Works', *American Scientist* **91**(1), 28.
Available from: <http://webvision.med.utah.edu/wp-content/uploads/2011/01/2003-01Kolb.pdf>
- Kolb, H. (2011), Simple anatomy of the retina, in 'Webvision', Moran Eye Center. Retrieved November 26, 2022.
Available from: <https://webvision.med.utah.edu/book/part-i-foundations/simple-anatomy-of-the-retina/>
- Kolb, H. (2013), Photoreceptors, in 'Webvision', Moran Eye Center. Retrieved November 26, 2022.
Available from: <https://webvision.med.utah.edu/book/part-ii-anatomy-and-physiology-of-the-retina/photoreceptors/>
- Kolb, H. & Marshak, D. (2003), 'The midget pathways of the primate retina', *Documenta Ophthalmologica* **106**(1), 67–81.
Available from: <https://link.springer.com/content/pdf/10.1023/A:1022469002511.pdf>
- 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', *J. Opt. Soc. Am. A* **17**(3), 517–526.
Available from: <https://opg.optica.org/josaa/abstract.cfm?URI=josaa-17-3-517>
- Lakowski, R. (1972), 'The Pickford-Nicolson anomaloscope as a test for acquired dyschromatopsias', *Modern problems in ophthalmology* **11**, 25–33.
- Lamb, T. D. (2022), 'Photoreceptor physiology and evolution: cellular and molecular basis of rod and cone phototransduction', *The Journal of Physiology* **600**(21), 4585–4601.
Available from: <https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/JP282058>
- Lamb, T. D. & Pugh, Edward N., J. (2006), 'Phototransduction, Dark Adaptation, and Rhodopsin Regeneration The Proctor Lecture', *Investigative Ophthalmology & Visual Science* **47**(12), 5138–5152.
Available from: <https://doi.org/10.1167/iovs.06-0849>
- Lawrance-Owen, A. J., Bosten, J. M., Hogg, R. E., Bargary, G., Goodbourn, P. T. & Mollon, J. D. (2014), 'Counterphase modulation flicker photometry:

- phenotypic and genotypic associations', *J. Opt. Soc. Am. A* **31**(4), A226–A231.
Available from: <https://opg.optica.org/josaa/abstract.cfm?URI=josaa-31-4-A226>
- Le Sueur, H., Mollon, J. D., Granzier, J. & Jordan, G. (2014), 'Counterphase modulation photometry: comparison of two instruments', *J. Opt. Soc. Am. A* **31**(4), A34–A37.
Available from: <https://opg.optica.org/josaa/abstract.cfm?URI=josaa-31-4-A34>
- Lee, B. B., Dacey, D. M., Smith, V. C. & Pokorny, J. (1999), 'Horizontal cells reveal cone type-specific adaptation in primate retina', *Proceedings of the National Academy of Sciences* **96**(25), 14611–14616.
Available from: <https://www.pnas.org/doi/abs/10.1073/pnas.96.25.14611>
- Lee, S. K., Sonoda, T. & Schmidt, T. M. (2019), 'M1 Intrinsically Photosensitive Retinal Ganglion Cells Integrate Rod and Melanopsin Inputs to Signal in Low Light', *Cell Reports* **29**(11), 3349–3355.e2.
Available from: <https://doi.org/10.1016/j.celrep.2019.11.024>
- Lewis, P. M. & Rosenfeld, J. V. (2016), 'Electrical stimulation of the brain and the development of cortical visual prostheses: An historical perspective', *Brain Research* **1630**, 208–224.
Available from: <https://www.sciencedirect.com/science/article/pii/S0006899315006897>
- Li, P., Garg, A. K., Zhang, L. A., Rashid, M. S. & Callaway, E. M. (2022), 'Cone opponent functional domains in primary visual cortex combine signals for color appearance mechanisms', *Nature Communications* **13**(1), 6344.
Available from: <https://www.nature.com/articles/s41467-022-34020-2>
- Lucas, R. J., Peirson, S. N., Berson, D. M., Brown, T. M., Cooper, H. M., Czeisler, C. A., Figueiro, M. G., Gamlin, P. D., Lockley, S. W., O'Hagan, J. B., Price, L. L., Provencio, I., Skene, D. J. & Brainard, G. C. (2014), 'Measuring and using light in the melanopsin age', *Trends in Neurosciences* **37**(1), 1–9.
Available from: <https://doi.org/10.1016/j.tins.2013.10.004>
- Lucas, R., Ponsonby, A.-L., McMichael, A., van der Mei, I., Chapman, C., Coulthard, A., Dear, K., Dwyer, T., Kilpatrick, T., Pender, M., Taylor, B., Valery, P. & Williams, D. (2007), 'Observational analytic studies in multiple sclerosis: controlling bias through study design and conduct. The Australian Multicentre Study of Environment and Immune Function', *Multiple Sclerosis Journal* **13**(7), 827–839. PMID: 17881396.
Available from: <https://doi.org/10.1177/1352458507077174>
- Mancuso, K., Neitz, M. & Neitz, J. (2006), 'An adaptation of the Cambridge Colour Test for use with animals', *Visual Neuroscience* **23**(3-4), 695–701.
Available from: <https://www.cambridge.org/core/journals/visual-neuroscience/article/an-adaptation-of-the-cambridge-colour-test-for-use-with-animals/705C02A3474C98B20CD0E52DC45C3AA1>

- Marechal, M., Delbarre, M., Tesson, J., Lacambre, C., Lefebvre, H. & Froussart-Maille, F. (2018), 'Color Vision Tests in Pilots' Medical Assessments', *Aerospace Medicine and Human Performance* **89**(8), 737–743.
Available from: <http://www.ingentaconnect.com/content/10.3357/AMHP.5009.2018>
- Margrain, T. H., Birch, J. & Owen, C. G. (1996), 'Colour Vision Requirements of Firefighters', *Occupational Medicine* **46**(2), 114–124.
Available from: <https://doi.org/10.1093/occmed/46.2.114>
- Maritime and Coastguard Agency (2020), MSN 1886 Amendment 1 MLC and ILO 188 medical standards: Medical examination system and appointment of approved doctors and medical and eyesight standards., Technical Report July 2020, Maritime and Coastguard Agency.
Available from: <https://www.gov.uk/government/publications/msn-1886-amendment-1-mlc-and-ilo-188-medical-standards>
- Markwell, E. L., Feigl, B. & Zele, A. J. (2010), 'Intrinsically photosensitive melanopsin retinal ganglion cell contributions to the pupillary light reflex and circadian rhythm', *Clinical and Experimental Optometry* **93**(3), 137–149.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1444-0938.2010.00479.x>
- Martin, D., Fowlkes, C. & Malik, J. (2004), 'Learning to detect natural image boundaries using local brightness, color, and texture cues', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**, 530–549.
Available from: <https://people.eecs.berkeley.edu/~efros/courses/AP06/Papers/martin-pami-04.pdf>
- Masri, R. A., Weltzien, F., Purushothuman, S., Lee, S. C. S., Martin, P. R. & Grünert, U. (2021), 'Composition of the Inner Nuclear Layer in Human Retina', *Investigative Ophthalmology & Visual Science* **62**(9), 22–22.
Available from: <https://doi.org/10.1167/iovs.62.9.22>
- MATLAB (2020), *9.8.0.1451342 (R2020a) Update 5*, The MathWorks Inc., Natick, Massachusetts.
- McKee, S. P., Klein, S. A. & Teller, D. Y. (1985), 'Statistical properties of forced-choice psychometric functions: Implications of probit analysis', *Perception & Psychophysics* **37**(4), 286–298.
Available from: <http://link.springer.com/10.3758/BF03211350>
- Meinert, C. L. & Tonascia, S. (1986), Single-center versus multicenter trials, in 'Clinical Trials: Design, Conduct and Analysis', Oxford University Press.
Available from: <https://doi.org/10.1093/acprof:oso/9780195035681.003.0004>
- Merriam-Webster (n.d.), Gold standard, in 'Merriam-Webster.com dictionary', Merriam-Webster. Retrieved November 4, 2022.
Available from: <https://www.merriam-webster.com/dictionary/gold%20standard>
- Metha, A. B. & Vingrys, A. J. (1992), 'The C-100:a new dichotomiser of colour vision defectives', *Clinical and Experimental Optometry* **75**(3), 114–123.

- Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1444-0938.1992.tb00615.x>
- Ministry of Defence (2018), SP 950 Medical Policy Leaflet 6-7-7. Joint Service Manual of Medical Fitness, Technical report, Ministry of Defence.
Available from: http://data.parliament.uk/DepositedPapers/Files/DEP2019-0604/Joint_Service_Manual_of_Medical_Fitness.pdf
- Miquilini, L., Ratis, M. A. S., Lima, M. G., Bento-Torres, N. V. O., Lacerda, E. M. d. C. B., Cortes, M. I. T., Rodrigues, A. R., Silveira, L. C. d. L. & Souza, G. d. S. (2019), 'A proposed correction in the weighted method to score the Ishihara test', *BMC Research Notes* **12**(1), 295.
Available from: <https://bmcresearchnotes.biomedcentral.com/articles/10.1186/s13104-019-4320-2>
- Mitchell, D. & Rushton, W. (1971), 'Visual pigments in dichromats', *Vision Research* **11**(10), 1033–1043.
Available from: <https://www.sciencedirect.com/science/article/pii/0042698971901106>
- Mollon, J. & Cavonius, L. (2012), 'The Lagerlunda Collision and the Introduction of Color Vision Testing', *Survey of Ophthalmology* **57**(2), 178–194.
Available from: <http://vision.psychol.cam.ac.uk/jdmollon/papers/MollonCavoniusOnLagerlundaCollision.pdf>
- Mollon, J. & Reffin, J. (1989), 'A computer-controlled colour vision test that combines the principles of Chibret and of Stilling', *Proceedings of the Physiological Society* **414**(10-11), 20.
Available from: <http://vision.psychol.cam.ac.uk/jdmollon/papers/MollonReffin1989.pdf>
- Mollon, J. & Reffin, J. (2000), *Handbook of the Cambridge Colour Test*, Cambridge Research Systems.
- Montag, E. D. & Boynton, R. M. (1987), 'Rod influence in dichromatic surface color perception', *Vision Research* **27**(12), 2153–2162.
Available from: <https://www.sciencedirect.com/science/article/pii/0042698987901295>
- Mortari, L. (2015), 'Reflectivity in research practice: An overview of different perspectives', *International Journal of Qualitative Methods* **14**(5), 1–9.
Available from: <https://doi.org/10.1177/1609406915618045>
- Murphey, D. K., Yoshor, D. & Beauchamp, M. S. (2008), 'Perception Matches Selectivity in the Human Anterior Color Center', *Current Biology* **18**(3), 216–220.
Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0960982208000183>
- Murphy, R. A. (2015), Comparing Color Vision Testing Using the Farnsworth-Munsell 100-Hue, Ishihara Compatible, and Digital TCV Software, Master's thesis, Pacific University Oregon, College Of Optometry.
Available from: <https://core.ac.uk/download/pdf/48859253.pdf>
- Nagel, W. A. (1907), 'Zwei Apparate für die augenärztliche Funktionssprüfung', *Ophthalmologica* **17**(3), 201–222.
Available from: <https://www.karger.com/DOI/10.1159/000291204>

- Nathans, J., Davenport, C. M., Maumenee, I. H., Lewis, R. A., Hejtmancik, J. F., Litt, M., Lovrien, E., Weleber, R., Bachynski, B., Zwas, F., Klingaman, R. & Fishman, G. (1989), 'Molecular Genetics of Human Blue Cone Monochromacy', *Science* **245**(4920), 831–838.
Available from: <http://www.jstor.org/stable/1703715>
- Nathans, J., Merbs, S. L., Sung, C.-H., Weitz, C. J. & Wang, Y. (1992), 'Molecular genetics of human visual pigments', *Annual Review of Genetics* **26**(1), 403–424. PMID: 1482119.
Available from: <https://doi.org/10.1146/annurev.ge.26.120192.002155>
- Nathans, J., Piantanida, T. P., Eddy, R. L., Shows, T. B. & Hogness, D. S. (1986), 'Molecular Genetics of Inherited Variation in Human Color Vision', *Science* **232**(4747), 203–210.
Available from: <http://www.sciencemag.org/cgi/doi/10.1126/science.3485310>
- National Research Council (1981), *Procedures for Testing Color Vision: Report of Working Group 41*, The National Academies Press, Washington, DC.
Available from: <https://nap.nationalacademies.org/catalog/746/procedures-for-testing-color-vision-report-of-working-group-41>
- Neitz, J. & Neitz, M. (2011), 'The genetics of normal and defective color vision', *Vision Research* **51**(7), 633–651. Vision Research 50th Anniversary Issue: Part 1.
Available from: <https://www.sciencedirect.com/science/article/pii/S0042698910005699>
- Neitz, J. & Neitz, M. (2016), 'Evolution of the circuitry for conscious color vision in primates', *Eye* **31**(2), 286–300.
Available from: <http://www.nature.com/articles/eye2016257>
- Neitz, J., Neitz, M. & Kainz, P. M. (1996), 'Visual pigment gene structure and the severity of color vision defects', *Science* **274**(5288), 801–804.
Available from: <https://www.science.org/doi/abs/10.1126/science.274.5288.801>
- Neitz, M. & Neitz, J. (2000), 'Molecular Genetics of Color Vision and Color Vision Defects', *Archives of Ophthalmology* **118**(5), 691.
Available from: <http://archophth.jamanetwork.com/article.aspx?doi=10.1001/archophth.118.5.691>
- Nelson, J. H. (1938), 'Anomalous trichromatism and its relation to normal trichromatism', *Proceedings of the Physical Society* **50**(5), 661–702.
Available from: <https://doi.org/10.1088/0959-5309/50/5/305>
- Ng, J. S. & Liem, S. C. (2018), 'Can the Farnsworth D15 Color Vision Test Be Defeated through Practice?', *Optometry and Vision Science* **95**(5), 452–456.
Available from: <https://journals.lww.com/00006324-201805000-00006>
- Ng, J. S. & Morton, W. A. (2018), 'Case Report: Invalidation of the Farnsworth D15 Test in Dichromacy Secondary to Practice', *Optometry*

- and Vision Science* **95**(3), 272–274.
Available from: <https://journals.lww.com/00006324-201803000-00014>
- Noble, S., Scheinost, D. & Constable, R. T. (2019), ‘A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis’, *NeuroImage* **203**, 116157.
Available from: <https://www.sciencedirect.com/science/article/pii/S1053811919307487>
- Oliphant, D. & Hovis, J. K. (1998), ‘Comparison of the D-15 and City University (second) color vision tests’, *Vision Research* **38**(21), 3461–3465.
Available from: <https://www.sciencedirect.com/science/article/pii/S0042698998001175>
- O’Neill-Biba, M., Sivaprasad, S., Rodriguez-Carmona, M., Wolf, J. E. & Barbur, J. L. (2010), ‘Loss of chromatic sensitivity in amd and diabetes: a comparative study’, *Ophthalmic and Physiological Optics* **30**(5), 705–716.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.2010.00775.x>
- Pacheco-Cutillas, M., Edgar, D. F. & Sahraie, A. (1999), ‘Acquired colour vision defects in glaucoma—their detection and clinical significance’, *British Journal of Ophthalmology* **83**(12), 1396–1402.
Available from: <http://bj.o.bmj.com/cgi/doi/10.1136/bjo.83.12.1396>
- Padmos, P. & van Norren, D. (1971), ‘Cone spectral sensitivity and chromatic adaptation as revealed by human flicker-electroretinography’, *Vision Research* **11**(1), 27–42.
Available from: <https://www.sciencedirect.com/science/article/pii/0042698971902033>
- Paramei, G. V. (2012), ‘Color discrimination across four life decades assessed by the Cambridge Colour Test’, *J. Opt. Soc. Am. A* **29**(2), A290–A297.
Available from: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-29-2-A290>
- Paramei, G. V. & Oakley, B. (2014), ‘Variation of color discrimination across the life span’, *J. Opt. Soc. Am. A* **31**(4), A375–A384.
Available from: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-31-4-A375>
- Parry, N. R. A. (2015), Color vision deficiencies, in A. J. Elliot, A. Franklin & M. D. Fairchild, eds, ‘Handbook of Color Psychology’, Cambridge Handbooks in Psychology, Cambridge University Press, Cambridge, pp. 216–242.
Available from: <https://www.cambridge.org/core/books/handbook-of-color-psychology/color-vision-deficiencies/7547787B2DB6C9ED2AEED99FB2D90BE3>
- Patterson, S. S., Kuchenbecker, J. A., Anderson, J. R., Neitz, M. & Neitz, J. (2020), ‘A Color Vision Circuit for Non-Image-Forming Vision in the Primate Retina’, *Current Biology* **30**(7), 1269–1274.e2.
Available from: <https://www.sciencedirect.com/science/article/pii/S0960982220300841>

- Peters, A. (2002), 'The effects of normal aging on myelin and nerve fibers: A review.', *Journal of neurocytology* **31**(8-9), 581–593.
Available from: <https://doi.org/10.1023/A:1025731309829>
- Pokorny, J. & Smith, V. C. (1976), 'Effect of field size on red-green color mixture equations', *J. Opt. Soc. Am.* **66**(7), 705–708.
Available from: <http://opg.optica.org/abstract.cfm?URI=josa-66-7-705>
- Pokorny, J., Smith, V. C. & Lutze, M. (1987), 'Aging of the human lens', *Appl. Opt.* **26**(8), 1437–1440.
Available from: <http://opg.optica.org/ao/abstract.cfm?URI=ao-26-8-1437>
- Pokorny, J., Smith, V. C., Pinckers, A. J. L. G. & Cozijnsen, M. (1982), 'Classification of complete and incomplete autosomal recessive achromatopsia', *Graefe's Archive for Clinical and Experimental Ophthalmology* **219**(3), 121–130.
Available from: <https://doi.org/10.1007/BF02152296>
- Prins, N. (2012), 'The psychometric function: The lapse rate revisited', *Journal of Vision* **12**(6), 25–25.
Available from: <https://doi.org/10.1167/12.6.25>
- Purves, D. (2004), *Neuroscience*, Sinauer Associates.
Available from: <https://books.google.co.uk/books?id=YZGDbwAACAAJ>
- Rauscher, F. G., Chisholm, C. M., Edgar, D. F. & Barbur, J. L. (2013), 'Assessment of novel binocular colour, motion and contrast tests in glaucoma', *Cell and Tissue Research* **353**(2), 297–310.
Available from: <http://link.springer.com/10.1007/s00441-013-1675-x>
- Rayleigh, L. (1881), 'Experiments on Colour', *Nature* **25**(629), 64–66.
Available from: <http://www.nature.com/articles/025064a0>
- Regan, B. C., Julliot, C., Simmen, B., Viénot, F., Charles-Dominique, P. & Mollon, J. D. (2001), 'Fruits, foliage and the evolution of primate colour vision', *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **356**(1407), 229–283.
Available from: <https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2000.0773>
- Regan, B., Reffin, J. & Mollon, J. (1994), 'Luminance noise and the rapid determination of discrimination ellipses in colour deficiency', *Vision Research* **34**(10), 1279–1299.
Available from: <https://www.sciencedirect.com/science/article/pii/0042698994902038>
- Regan, D. (1974), 'Electrophysiological evidence for colour channels in human pattern vision', *Nature* **250**(5465), 437–439.
Available from: <https://www.nature.com/articles/250437a0>
- Ripamonti, C., Woo, W. L., Crowther, E. & Stockman, A. (2009), 'The S-cone contribution to luminance depends on the M- and L-cone adaptation levels: Silent surrounds?', *Journal of Vision* **9**(3), 10–10.
Available from: <https://doi.org/10.1167/9.3.10>

- Rodriguez-Carmona, M. & Barbur, J. (2017), Variability in normal and defective colour vision, in J. Best, ed., 'Colour Design', 2nd edn, Cambridge: Woodhead Publishing Limited, chapter 3, pp. 43–97.
Available from: <https://www.elsevier.com/books/colour-design/best/978-0-08-101270-3>
- Rodriguez-Carmona, M., Evans, B. E. W. & Barbur, J. L. (2020), 'Color vision assessment-2: Color assessment outcomes using single and multi-test protocols', *Color Research & Application* **46**(1), 21–32.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/col.22598>
- Rodriguez-Carmona, M., Harlow, A. J., Walker, G. & Barbur, J. L. (2005), The Variability of Normal Trichromatic Vision and the Establishment of the "Normal" Range, in 'Proceedings of 10th Congress of the International Colour Association', International Colour Association, pp. 979–982.
Available from: http://www.city-occupational.co.uk/papers/ICA_2005.pdf
- Rodriguez-Carmona, M. L. (2006), Variability of chromatic sensitivity: fundamental studies and clinical applications, PhD thesis, City University London. [Unpublished Doctoral thesis].
Available from: <https://openaccess.city.ac.uk/id/eprint/12058/>
- Rodriguez-Carmona, M., O'Neill-Biba, M. & Barbur, J. L. (2012), 'Assessing the severity of color vision loss with implications for aviation and other occupational environments.', *Aviation, space, and environmental medicine* **83**(1), 19–29.
Available from: <https://openaccess.city.ac.uk/id/eprint/1301/>
- Rogers, R. R. (2001), 'Reflection in Higher Education: A Concept Analysis', *Innovative Higher Education* **26**(1), 37–57.
Available from: <https://link.springer.com/article/10.1023/A:1010986404527>
- Rothen, N., Meier, B. & Ward, J. (2012), 'Enhanced memory ability: Insights from synaesthesia', *Neuroscience & Biobehavioral Reviews* **36**(8), 1952–1963.
Available from: <https://www.sciencedirect.com/science/article/pii/S0149763412000760>
- Rouw, R., Scholte, H. S. & Colizoli, O. (2011), 'Brain areas involved in synaesthesia: A review', *Journal of Neuropsychology* **5**(2), 214–242.
Available from: <https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1111/j.1748-6653.2011.02006.x>
- Rushton, W. A. H. (1972), 'Pigments and signals in colour vision', *The Journal of Physiology* **220**(3), 1–31.
Available from: <https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1972.sp009719>
- Ryan, M. (2011), 'Improving reflective writing in higher education: a social semiotic perspective', *Teaching in Higher Education* **16**(1), 99–111.
Available from: <http://www.tandfonline.com/doi/abs/10.1080/13562517.2010.507311>

- Sabesan, R., Hofer, H. & Roorda, A. (2015), 'Characterizing the human cone photoreceptor mosaic via dynamic photopigment densitometry', *PLOS ONE* **10**(12), 1–12.
Available from: <https://doi.org/10.1371/journal.pone.0144891>
- Schmidt, T. M., Do, M. T. H., Dacey, D., Lucas, R., Hattar, S. & Matynia, A. (2011), 'Melanopsin-positive intrinsically photosensitive retinal ganglion cells: From form to function', *Journal of Neuroscience* **31**(45), 16094–16101.
Available from: <https://www.jneurosci.org/content/31/45/16094>
- Schneck, M. E., Adams, A. J., Volbrecht, V. J. & Haegerstrom-Portnoy, G. (1990), Residual Cone Function in "Rod Monochromats" and the Mechanisms of Rod-Cone Interactions, in 'Noninvasive Assessment of the Visual System', Optica Publishing Group, p. ThA2.
Available from: <https://opg.optica.org/abstract.cfm?URI=NAVS-1990-ThA2>
- Schön, D. A. (1992), *The Reflective Practitioner*, 1st edn, Routledge.
Available from: <https://www.taylorfrancis.com/books/9781351883160>
- Schroeder, M. M., Harrison, K. R., Jaekel, E. R., Berger, H. N., Zhao, X., Flannery, M. P., St. Pierre, E. C., Pateqi, N., Jachimska, A., Chervenak, A. P. & Wong, K. Y. (2018), 'The Roles of Rods, Cones, and Melanopsin in Photoresponses of M4 Intrinsically Photosensitive Retinal Ganglion Cells (ipRGCs) and Optokinetic Visual Behavior', *Frontiers in Cellular Neuroscience* **12**(203), 1–14.
Available from: <https://www.frontiersin.org/article/10.3389/fncel.2018.00203/full>
- Seifirad, S. & Alquran, L. (2021), 'The bigger, the better? When multicenter clinical trials and meta-analyses do not work', *Current Medical Research and Opinion* **37**(2), 321–326. PMID: 33287578.
Available from: <https://doi.org/10.1080/03007995.2020.1860922>
- Seshadri, J., Christensen, J., Lakshminarayanan, V. & Bassi, C. J. (2005), 'Evaluation of the new web-based "Colour Assessment and Diagnosis" test', *Optometry and Vision Science* **82**(10), 882–885.
Available from: <https://oce.ovid.com/article/00006324-200510000-00008/PDF>
- Sharpe, L. T. & Nordby, K. (1990), Total Colour Blindness: An Introduction, in R. F. Hess, L. T. Sharpe & K. Nordby, eds, 'Night Vision: Basic, Clinical and Applied Aspects', Cambridge University Press, pp. 253–289.
- Shevell, S. (2003), *The Science of Color*, Elsevier Science.
Available from: <https://books.google.it/books?id=-fNJZ0xmTFIC>
- Shinomori, K., Scheffrin, B. E. & Werner, J. S. (2001), 'Age-related changes in wavelength discrimination', *J. Opt. Soc. Am. A* **18**(2), 310–318.
Available from: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-18-2-310>
- Simunovic, M. P. (2016), 'Acquired color vision deficiency', *Survey of Ophthalmology* **61**(2), 132–155.
Available from: <https://www.sciencedirect.com/science/article/pii/S0039625715001939>

- Sliney, D. H. (2002), 'How Light Reaches the Eye and Its Components', *International Journal of Toxicology* **21**(6), 501–509. PMID: 12537646.
Available from: <https://doi.org/10.1080/10915810290169927>
- Sliney, D. H. (2016), 'What is light? The visible spectrum and beyond', *Eye* **30**(2), 222–229.
Available from: <http://www.nature.com/articles/eye2015252>
- Smees, R., Hughes, J., Carmichael, D. A. & Simner, J. (2019), 'Learning in colour: children with grapheme-colour synaesthesia show cognitive benefits in vocabulary and self-evaluated reading', *Philosophical Transactions of the Royal Society B: Biological Sciences* **374**(1787), 20180348.
Available from: <https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2018.0348>
- Smith, V. C., Pokorny, J., Lee, B. B. & Dacey, D. M. (2001), 'Primate Horizontal Cell Dynamics: An Analysis of Sensitivity Regulation in the Outer Retina', *Journal of Neurophysiology* **85**(2), 545–558. PMID: 11160492.
Available from: <https://doi.org/10.1152/jn.2001.85.2.545>
- Smithson, H. (2005), 'Sensory, computational and cognitive components of human colour constancy', *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**(1458), 1329–1346.
Available from: <https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2005.1633>
- Solomon, S. G. (2021), Chapter 3 - Retinal ganglion cells and the magnocellular, parvocellular, and koniocellular subcortical visual pathways from the eye to the brain, in J. J. Barton & A. Leff, eds, 'Neurology of Vision and Visual Disorders', Vol. 178 of *Handbook of Clinical Neurology*, Elsevier, pp. 31–50.
Available from: <https://www.sciencedirect.com/science/article/pii/B9780128213773000180>
- Spitschan, M. & Woelders, T. (2018), 'The method of silent substitution for examining melanopsin contributions to pupil control', *Frontiers in Neurology* **9**.
Available from: <https://www.frontiersin.org/articles/10.3389/fneur.2018.00941>
- Squire, T. J., Rodriguez-Carmona, M., Evans, A. D. & Barbur, J. L. (2005), 'Color vision tests for aviation: Comparison of the anomaloscope and three lantern types', *Aviation Space and Environmental Medicine* **76**(5), 421–429.
Available from: https://www.peter2000.co.uk/aviation/misc/JAR_colour_study.pdf
- Steward, J. M. & Cole, B. L. (1989), 'What Do Color Vision Defectives Say About Everyday Tasks?', *Optometry and Vision Science* **66**(5), 288–295.
Available from: https://journals.lww.com/optvissci/Abstract/1989/05000/What_Do_Color_Vision_Defectives_Say_About_Everyday.6.aspx
- Stilling, J. (1878), *Die Prüfung des Farbensinnes beim Eisenbahn- und Marinepersonal*, Cassel : T. Fischer. [Translation: "The examination of the color sense of the railway and naval personnel"].

- Stockman, A. (2018), Report on Colour vision standards for STU officers, Technical report, UCL Institute of Ophthalmology. Version: 6.1.
- Stockman, A. (2019), ‘Cone fundamentals and CIE standards’, *Current Opinion in Behavioral Sciences* **30**, 87–93.
Available from: <https://doi.org/10.1016/j.cobeha.2019.06.005>
- Stockman, A. & Brainard, D. H. (2010), COLOR VISION MECHANISMS, in M. Bass, ed., ‘Handbook of Optics: Volume III - Vision and Vision Optics’, 3rd edition edn, McGraw-Hill Education, New York.
Available from: <https://www.accessengineeringlibrary.com/content/book/9780071498913/chapter/chapter11>
- Stockman, A. & Brainard, D. H. (2015), Fundamentals of color vision I: color processing in the eye, in A. J. Elliot, M. D. Fairchild & A. Franklin, eds, ‘Handbook of Color Psychology’, Cambridge Handbooks in Psychology, Cambridge University Press, p. 27–69.
Available from: <https://doi.org/10.1017/CB09781107337930.004>
- Stockman, A., MacLeod, D. I. A. & Johnson, N. E. (1993), ‘Spectral sensitivities of the human cones’, *J. Opt. Soc. Am. A* **10**(12), 2491–2521.
Available from: <https://opg.optica.org/josaa/abstract.cfm?URI=josaa-10-12-2491>
- Sumner, P. & Mollon, J. (2000), ‘Chromaticity as a signal of ripeness in fruits taken by primates’, *Journal of Experimental Biology* **203**(13), 1987–2000.
Available from: <https://doi.org/10.1242/jeb.203.13.1987>
- 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. PMID: 16424455.
Available from: <https://doi.org/10.1152/jn.00714.2005>
- Szmajda, B. A., Buzás, P., FitzGibbon, T. & Martin, P. R. (2006), ‘Geniculocortical relay of blue-off signals in the primate visual system’, *Proceedings of the National Academy of Sciences* **103**(51), 19512–19517.
Available from: <https://www.pnas.org/doi/abs/10.1073/pnas.0606970103>
- Szmajda, B. A., Grünert, U. & Martin, P. R. (2008), ‘Retinal ganglion cell inputs to the koniocellular pathway’, *Journal of Comparative Neurology* **510**(3), 251–268.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cne.21783>
- Tang, T., Álvaro, L., Alvarez, J., Maule, J., Skelton, A., Franklin, A. & Bosten, J. (2022), ‘ColourSpot, a novel gamified tablet-based test for accurate diagnosis of color vision deficiency in young children’, *Behavior Research Methods* **54**(3), 1148–1160.
Available from: <https://link.springer.com/10.3758/s13428-021-01622-5>
- Thomas, P. B. & Mollon, J. D. (2004), ‘Modelling the Rayleigh match’, *Visual Neuroscience* **21**(3), 477–482.
Available from: <https://www.cambridge.org/core/journals/visual-neuroscience/article/modelling-the-rayleigh-match/8793F93169529A16DD46BD905F55A251>

- Topley, H. (1959), 'Sight testing for the merchant navy.', *The British journal of physiological optics* **16**(1), 36–47.
- Trevethan, R. (2017), 'Sensitivity, Specificity, and Predictive Values: Foundations, Plabilities, and Pitfalls in Research and Practice', *Frontiers in Public Health* **5**.
Available from: <https://www.frontiersin.org/articles/10.3389/fpubh.2017.00307>
- Trieschmann, M., van Kuijk, F. J. G. M., Alexander, R., Hermans, P., Luthert, P., Bird, A. C. & Pauleikhoff, D. (2008), 'Macular pigment in the human retina: histological evaluation of localization and distribution', *Eye* **22**(1), 132–137.
Available from: <https://doi.org/10.1038/sj.eye.6702780>
- Turberville, D. (1684), 'Two letters from the great, and experienced oculist, Dr. Turberville of Salisbury to Mr. William Mulgrave SPS of Oxon containing several remarkable cases in Physick, relating chiefly to the Eyes.', *Phil Trans* **14**, 735–738.
- Ulrich, R. & Miller, J. (2004), 'Threshold estimation in two-alternative forced-choice (2AFC) tasks: The Spearman-Kärber method', *Perception and Psychophysics* **66**(3), 517–533.
Available from: <https://link.springer.com/article/10.3758/BF03194898>
- Usui, T., Kremers, J., Sharpe, L. T. & Zrenner, E. (1998), 'Flicker cone electroretinogram in dichromats and trichromats', *Vision Research* **38**(21), 3391–3396.
Available from: <https://www.sciencedirect.com/science/article/pii/S0042698997004665>
- Vemala, R., Sivaprasad, S. & Barbur, J. L. (2017), 'Detection of Early Loss of Color Vision in Age-Related Macular Degeneration – With Emphasis on Drusen and Reticular Pseudodrusen', *Investigative Ophthalmology & Visual Science* **58**(6), 247–254.
Available from: <https://doi.org/10.1167/iovs.17-21771>
- Verriest, G. (1963), 'Further studies on acquired deficiency of color discrimination*', *J. Opt. Soc. Am.* **53**(1), 185–195.
Available from: <http://opg.optica.org/abstract.cfm?URI=josa-53-1-185>
- Verriest, G. (1968), 'Comparative study of effectiveness of various tests for recognition of color vision abnormalities', *Archives des maladies professionnelles de medecine du travail et de securite sociale* **29**(6), 293–314.
- Verriest, G., Laethem, J. V. & Uvijls, A. (1982), 'A New Assessment of the Normal Ranges of the Farnsworth-Munsell 100-Hue Test Scores', *American Journal of Ophthalmology* **93**(5), 635–642.
- Verriest, G. & Uvijls, A. (1989), Results of the Oscar Test in Groups of Normal and Abnormal Subjects, in B. Drum & G. Verriest, eds, 'Colour Vision Deficiencies IX: Proceedings of the ninth symposium of the International Research Group on Colour Vision Deficiencies, held at St. John's College,

- Annapolis, Maryland, U.S.A., 1–3 July 1987', Springer Netherlands, Dordrecht, pp. 479–485.
Available from: https://doi.org/10.1007/978-94-009-2695-0_55
- Vingrys, A. J. & Cole, B. L. (1983), 'Validation of the Holmes - Wright lanterns for testing colour vision', *Ophthalmic and Physiological Optics* **3**(2), 137–152.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.1983.tb00593.x>
- Vogt, H., Hofmann, B. & Getz, L. (2016), 'Personalized medicine: evidence of normativity in its quantitative definition of health.', *Theoretical medicine and bioethics* **37**(5), 401 – 416.
Available from: <https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=mdc&AN=27638683&authtype=shib&site=ehost-live&scope=site&custid=s1089299>
- von Campenhausen, C. & Schramme, J. (1995), '100 Years of Benham's Top in Colour Science', *Perception* **24**(6), 695–717. PMID: 7478909.
Available from: <https://doi.org/10.1068/p240695>
- Waalder, G. H. M. (1927), 'Über die Erblichkeitsverhältnisse der verschiedenen Arten von angeborener Rotgrünblindheit', *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* **45**(1), 279–333.
Available from: <https://doi.org/10.1007/BF01847031>
- Walls, G. L. (1959), 'How good is the H-R-R test for color blindness.', *American Journal of Optometry and Archives of American Academy of Optometry* **36** 4, 169–193.
Available from: <https://oce.ovid.com/article/00000450-195904000-00001/HTML>
- Wandell, B. A., Brainard, D. H. & Cottaris, N. P. (2022), Chapter 11 - Visual encoding: Principles and software, in N. Santhi & M. Spitschan, eds, 'Circadian and Visual Neuroscience', Vol. 273 of *Progress in Brain Research*, Elsevier, pp. 199–229.
Available from: <https://www.sciencedirect.com/science/article/pii/S0079612322001339>
- Werblin, F. S. (1973), 'The Control of Sensitivity in the Retina', *Scientific American* **228**(1), 70–79.
Available from: <http://www.jstor.org/stable/24922957>
- Werner, J. S. & Steele, V. G. (1988), 'Sensitivity of human foveal color mechanisms throughout the life span', *J. Opt. Soc. Am. A* **5**(12), 2122–2130.
Available from: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-5-12-2122>
- Wichmann, F. A. & Hill, N. J. (2001), 'The psychometric function: I. Fitting, sampling, and goodness of fit', *Perception & Psychophysics* **63**(8), 1293–1313.
Available from: <http://link.springer.com/10.3758/BF03194544>
- Williams, D. R. (2011), 'Imaging single cells in the living retina', *Vision Research* **51**(13), 1379–1396. Vision Research 50th Anniversary Issue: Part 2.
Available from: <https://www.sciencedirect.com/science/article/pii/S0042698911001763>

- Wool, L. E., Packer, O. S., Zaidi, Q. & Dacey, D. M. (2019), 'Connectomic Identification and Three-Dimensional Color Tuning of S-OFF Midget Ganglion Cells in the Primate Retina', *Journal of Neuroscience* **39**(40), 7893–7909.
Available from: <https://www.jneurosci.org/content/39/40/7893>
- Wright, W. D. (1946), *Researches on Normal and Defective Colour Vision*, London: Henry Kimpton.
- Wuerger, S., Xiao, K., Fu, C. & Karatzas, D. (2010), 'Colour-opponent mechanisms are not affected by age-related chromatic sensitivity changes', *Ophthalmic and Physiological Optics* **30**(5), 653–659.
Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-1313.2010.00744.x>
- Wyszecki, G. & Stiles, W. S. (1982), *Color Science: Concepts and Methods Quantitative Data and Formulae*, 2nd edn, Wiley New York.
- Yue, L., Castillo, J., Gonzalez, A. C., Neitz, J. & Humayun, M. S. (2021), 'Restoring Color Perception to the Blind', *Ophthalmology* **128**(3), 453–462.
Available from: <https://doi.org/10.1016/j.ophtha.2020.08.019>
- Zachi, E. C., Costa, T. L., Barboni, M. T. S., Costa, M. F., Bonci, D. M. O. & Ventura, D. F. (2017), 'Color Vision Losses in Autism Spectrum Disorders', *Frontiers in Psychology* **8**, 1–11.
Available from: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01127>
- Zisman, F., Seger, K. R. & Adams, A. J. (1987), Specificity Evaluation of the Oscar Color Vision Test, in G. Verriest, ed., 'Colour Vision Deficiencies VIII', Springer Netherlands, Dordrecht, pp. 173–176.
Available from: https://doi.org/10.1007/978-94-009-4275-2_26
- Zrenner, E. (1984), Electrophysiology of Colour Vision, in G. Verriest, ed., 'Colour Vision Deficiencies VII', Springer Netherlands, Dordrecht, pp. 29–48.
Available from: https://doi.org/10.1007/978-94-009-6551-5_4