Using an illumination discrimination paradigm to investigate the role of illumination priors in colour perception

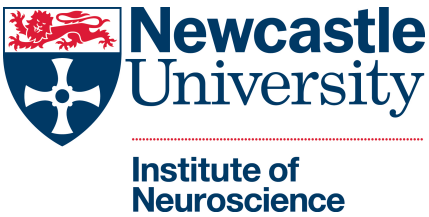
Stacey Aston

Student Number: 071626134

Supervisor: Prof. Anya Hurlbert

A thesis submitted for the degree of

Doctor of Philosophy



Institute of Neuroscience

Newcastle University

Newcastle upon Tyne

United Kingdom

September 2017

For my mum and dad.

Acknowledgements

Firstly, I'd like to thank my supervisor, Anya Hurlbert, for accepting me into her research group to complete my PhD studies. Anya has been supportive throughout my studies, encouraging independent research, supporting my attendance at conferences and summer schools, and connecting me with many vision science researchers. We may not have always seen eye to eye, but all in all, it has been a fruitful three years and I have learnt a hell of a lot.

In addition, I'd like to thank all those who have passed through Anya's laboratory during my time there and made the work day much more fun. I'd especially like to thank Naomi Gross and Daisy Fitzpatrick who were employed as research assistants during my final year of study and collected a great deal of the data presented in this thesis. Without them, I certainly would not have achieved as much as I did. Additional thanks go to Thomas Le Couteur Bisson, Matt Cranwell, Stuart Crichton and Brad Pearce who all shared in the PhD process with me at various stages. I must also not forget Angela Owen or Gabriele Jordan who have both given me vital encouragement when I was most in need. I'd also like to thank the undergraduates who completed their dissertations in the lab at various times. Two come to mind in particular who both collected fantastic data sets that have contributed to this thesis. They are Jay Turner and Olivia Kingston.

There are others in academia that I'd like to thank outside of Anya's research group. Firstly, David Brainard who kindly hosted me in his laboratory at UPenn for a week in early 2017, invited me to Stanford to meet with the ISETBIO team in the summer of 2016, and provided some important words of wisdom throughout my studies. Secondly, Ana Radonjić and Maria Olkkonen, who have both been terrific role models for a PhD student learning their trade and have also become great friends. In addition, David and Ana are collaborators on Chapters 3/4 and Maria on Chapter 6.

I would of course also like to express my eternal gratitude to my forever loving and supportive parents, Janet and Brian Aston. It has been a long road to get here but I am finally submitting a PhD thesis on my second attempt. For the first attempt, I was self-funding and I was a huge financial burden to my parents during that time. Without that experience though, I would not be where I am today. I will be forever indebted to them and hope that wherever I go in my career beyond this point, I will do them proud.

That brings me to my final thank you. The experience of writing a thesis has been just as much of a slog as everyone will tell you. However, I feel that process has been made a little easier by knowing that I have a Post-Doc position to go to. For that reason, I'd like to thank Marko Nardini and Ulrik Beierholm for taking a punt on a rather unknown quantity. I look forward to joining them in Durham and hope that I become a valuable member of their group.

Abstract

Previous studies suggest human colour constancy is optimised for natural daylight illuminations - a "blue bias" for colour constancy - but it is unclear how such a bias is encoded in the visual system. We use an illumination discrimination task to test two hypothesised mechanisms. Both hypotheses suggest that the human visual system has a prior expectation that illuminations are more likely to vary in a bluer region of chromaticity space. One hypothesis (the nature hypothesis) suggests this has developed in the human visual system through evolution, with selection of colour mechanisms that have reduced sensitivity to global bluer changes across a scene (a species prior). The second hypothesis suggests that the prior is learnt through experience with illuminations (the nurture hypothesis - an individual prior). In Chapter 3 we expand on previous results showing a "blue bias" for colour constancy when the illumination varies from a neutral reference, to show that the "blue bias" prevails in variants of the task where the illuminations are all chromatically biased. This result supports the nature hypothesis. However, depending on the chromatic bias, different biases can emerge in the threshold data that are more supportive of the nurture hypothesis. In Chapter 4 we explore individual differences in illumination discrimination ability, compare illumination discrimination ability with chromatic contrast detection ability, and develop ideal observer models for the task. The results in this Chapter are mostly in support of the nurture hypothesis. In Chapter 5 we show that illumination priors may play a role in the recent visual illusion of a dress photograph that appeared blue and black to some observers but white and gold to others. Finally, in Chapter 6, we search for evidence that observers can learn an illumination prior during a psychophysical task. We conclude that the "blue bias" is likely governed by both a learnt prior over the characteristics of daylight illuminations (the nurture hypothesis) and a generic reduction in sensitivity to bluer changes in an illumination (the nature hypothesis).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Overview

Our ability to see is largely taken for granted. We lack appreciation for the enormous amount of work our visual systems must do to process visual input and infer the distal stimulus whilst ensuring that our visual percept is accurate and consistent. Vision is used without conscious effort for object recognition, guiding our interactions with the world. It is especially important when we must react quickly to changes in the environment during activities like playing sports or driving cars. Yet, as effortless and instantaneous as these tasks seem, neuroscientists, psychologists, and computer scientists still struggle to create computational models of visual processing that mimic human behaviour. The problem boils down to the fact that human vision is an ill-posed problem.

Hohwy (2013) illustrates the under-determined nature of visual perception with the following example. Suppose an observer receives some visual input that stimulates the photoreceptive layer of cells at the back of the eye (the retina, see Section 1.3) in such a way as to cause a pattern of activity that represents the shape of half a bicycle immediately next to a rectangle (this is a simplified representation of how the visual system works but will do for this demonstration). This visual input could be caused by multiple physical worlds. For example, the distal stimulus could be a bicycle occluded by a wall, half a bicycle next to a wall, a swarm of bees flying in the shape of half a bicycle, or one of an infinite number of other causes. Hohwy (2013) makes the point that one of the causes listed above (occluded bicycle) can be considered more likely than the others. If the visual system incorporates this information as a probability distribution over the competing hypotheses, then it is using prior knowledge to overcome uncertainty in the sensory signal - the Bayesian Brain Hypothesis (Knill & Pouget, 2004; Mamassian, Landy & Maloney, 2003; Pouget, Beck, Ma & Latham, 2013). Estimating probability distributions over possible stimulus configurations, or in other words, forming priors for perception, begins to

constrain visual processing and frames the problem in terms of probability. What is the most likely distal stimulus given prior assumptions and the incoming data? Considering perception as inference will constitute a major theme of this work.

There is uncertainty in visual perception for other reasons too. Not only can there be many causes for the same visual input, but the same physical stimulus can lead to different visual stimulation depending on factors such as viewing angle, viewing distance and the incident illumination. In addition, the imperfections of the human visual system (caused by biological constraints) make for a noisy sensory signal with added uncertainty. Yet, the usefulness of a visual system is bounded by its ability to serve invariant perception in the form of stable object recognition or visual constancy (Dicarlo & Cox, 2007; Rust & Stocker, 2010). Constancy in vision may be considered in many different modalities such as size perception, shape perception, and, the subject of this thesis, colour perception. Colour constancy is the term used to describe the phenomenon whereby the human visual system maintains a relatively stable colour percept of the world despite changes in illumination; or at least this is the most common form of the definition as colour constancy can also be studied under illumination gradients, differences in viewing angle or changes in scene configuration. Indeed, spatial variations in the illumination colour within a scene can be on the same order of magnitude as variations throughout the day (Nascimento, Amano & Foster, 2016).

In the studies that follow we investigate the role of prior assumptions for illumination properties in colour constancy. We will expand on previous results from an illumination discrimination paradigm that is considered a method of establishing thresholds for colour constancy. Previous studies using this paradigm find evidence of a "blue bias" for colour constancy; illumination discrimination is worse for bluer illumination changes along the axis of daylight chromaticities (Pearce, Crichton, Mackiewicz, Finlayson & Hurlbert, 2014; Radonjić, Pearce, Aston, Krieger, Cottaris, Brainard & Hurlbert, 2016*b*). In Chapter 3 we show that the "blue bias" prevails in variants of the task, suggesting a reduced sensitivity to global bluer changes in a scene. However, we find that depending on the task parameters,

different biases can emerge in the threshold data. We will pay particular attention to the implications of inter-individual differences in performance on the task in Chapter 4, where we also ask if differences in chromatic contrast detection ability can predict differences in illumination discrimination ability. In addition, we develop ideal observer models for the illumination discrimination task to assess whether differences in the sensory machinery of the visual system can account for the inter-individual differences in performance or whether more general properties of human observers predict the "blue bias". Chapter 5 applies this work to a photograph of a blue and black (or white and gold) dress that went viral on the internet in 2015. Finally, in Chapter 6, we search for evidence that observers can learn an illumination prior during a psychophysical task. In what remains of this Chapter, we introduce the main concepts of vision science and review the relevant literature.

1.2 The Input

1.2.1 Light

Vision begins with a light source. Light is electromagnetic radiation and is made up of photons. Photons travel in waves with a wavelength determined by the energy of the photon. The more energy photons have the shorter the wavelength, while less energetic photons travel with longer wavelengths. Wavelength is not the only parameter that defines a wave; waves also have a speed and a frequency. For light waves, the speed is constant (in air) while the frequency of the wave is related to the wavelength; the longer the wavelength the lower the frequency and vice versa. Sources of light generally emit photons that are travelling at a variety of wavelengths, and the total number of photons emitted by the light source can vary greatly for each wavelength. The human visual system is only sensitive to a small proportion of the light that makes up the electromagnetic spectrum (visible light, wavelengths between approximately 400 and 700 nm; Figure 1.1).

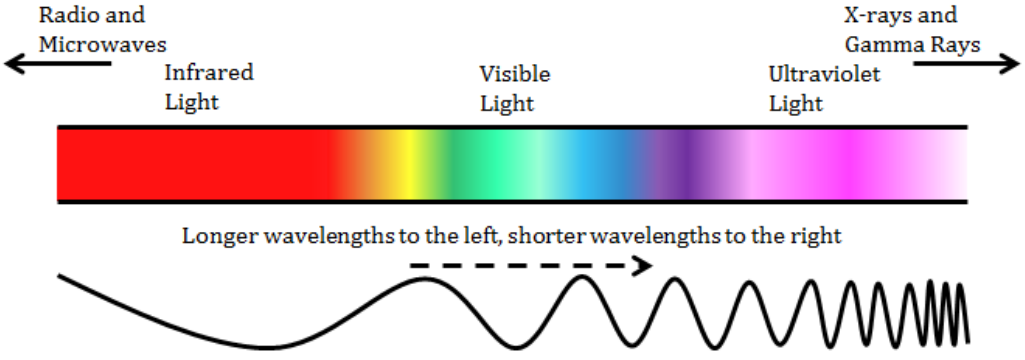


Figure 1.1: The electromagnetic spectrum.

1.2.2 Quantifying Light

Light can be quantified in several ways, taking into account wavelength or disregarding it. Spectral measures of light take wavelength into account. If we are not interested in measuring light at each wavelength and prefer a global measure, we simply sum the spectral measure of interest over all wavelengths. Throughout the text, we will quantify light using both spectral and global measures.

1.2.2.1 Radiometry

Radiometric quantities characterise the properties of light regardless of the observer. The only radiometric quantities we review here are spectral irradiance and spectral radiance, although there are many more. Spectral irradiance defines the amount of light (formally referred to as the radiant flux) falling on a point on a surface as a function of wavelength (Figure 1.2.A) and is expressed in $W/m^2/nm$, or amount of energy per unit time, per unit area, per wavelength (watts equal to joules per second). Spectral radiance defines the amount of light emitted, reflected, or transmitted in a given direction from a point on a surface (Figure 1.2.B) and is expressed in units of $W/m^2/sr/nm$, or amount of energy per unit time, per unit area, per solid angle, per wavelength.

The distinction between spectral irradiance and spectral radiance is subtle, but is enough to ensure that one cannot be obtained from the other unless strict conditions are met.

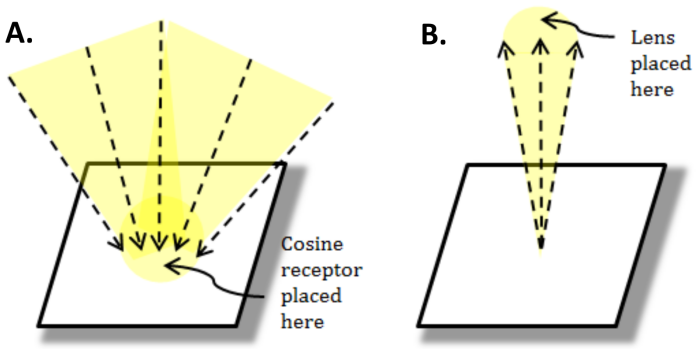


Figure 1.2: Radiometric measures of light. A. Irradiance defines all the light falling on a particular point (or space). B. Radiance defines all the light coming from a particular point (or space).

Suppose a measurement of spectral radiance $R(\lambda)$ has been taken from a point p on a surface S. We can calculate the spectral irradiance $I(\lambda)$ from $R(\lambda)$ using the equation

$$I(\lambda) = \frac{\pi R(\lambda)}{r} \tag{1.1}$$

if and only if S is a Lambertian surface (a surface which reflects each wavelength of light equally in all directions), where r is the proportion of light that the surface reflects.

Unfortunately, the world is littered with non-Lambertian surfaces and this equation is rarely useful. Instead, to obtain measurements of both spectral radiance and spectral irradiance, one must utilise two different types of measuring equipment. The main difference between radiance meters (used to measure spectral radiance) and irradiance meters (used to measure spectral irradiance) is the lens. Radiance meters use a standard lens which is focused at the point to be measured from (Figure 1.3.B). Irradiance meters however are fitted with a cosine receptor. The cosine receptor is placed at the point where one wishes to measure the incident light and performs the necessary function of gathering the light falling on that point from every direction, correcting the power of the light according to the cosine of the angle (Figure 1.3.A).

A measurement of spectral irradiance or spectral radiance can be used to define the spectral power distribution (SPD) of a light source (although the spectral radiance measurement must be taken from a Lambertian surface). Hence, a SPD defines the power of a light

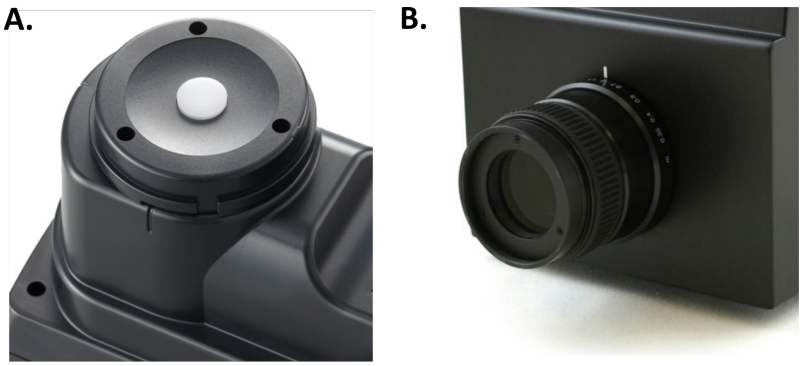


Figure 1.3: Light measurement devices. A. Cosine receptor used for irradiance measurements. B. Standard lens used for radiance measurements.

source at each wavelength across the visual spectrum. Spectral radiance measures can also be used to characterise the light reflected from a particular surface. At different points in this thesis, both radiance and irradiance measures of light will be used.

1.2.3 Reflectance

If objects can be said to have a 'true" colour, one would define this to be a correlate of the object's surface spectral reflectance function. An object's surface spectral reflectance function $S(\lambda)$ specifies the quantity of light that the surface will reflect for each wavelength across the visual spectrum. When the light emitted from a source $I(\lambda)$ hits an object, the light that is reflected from the object $R(\lambda)$ is the point-wise product of the two functions such that $R(\lambda) = I(\lambda)S(\lambda)$ (Figure 1.4). The light that an object reflects can be quantified by a spectral radiance measurement taken from the object's surface.

1.3 The Human Eye

Light enters the eye through the pupil, which has the capability to expand and contract to modulate the amount of light that enters the eye. After passing through the pupil, light is focused on the retina at the back of the eye by the lens. The lens also has the capacity to expand and contract in order to focus light from scenes at a variety of distances.

The retina has several layers; an illustration of the retinal layers can be seen in Figure 1.5.

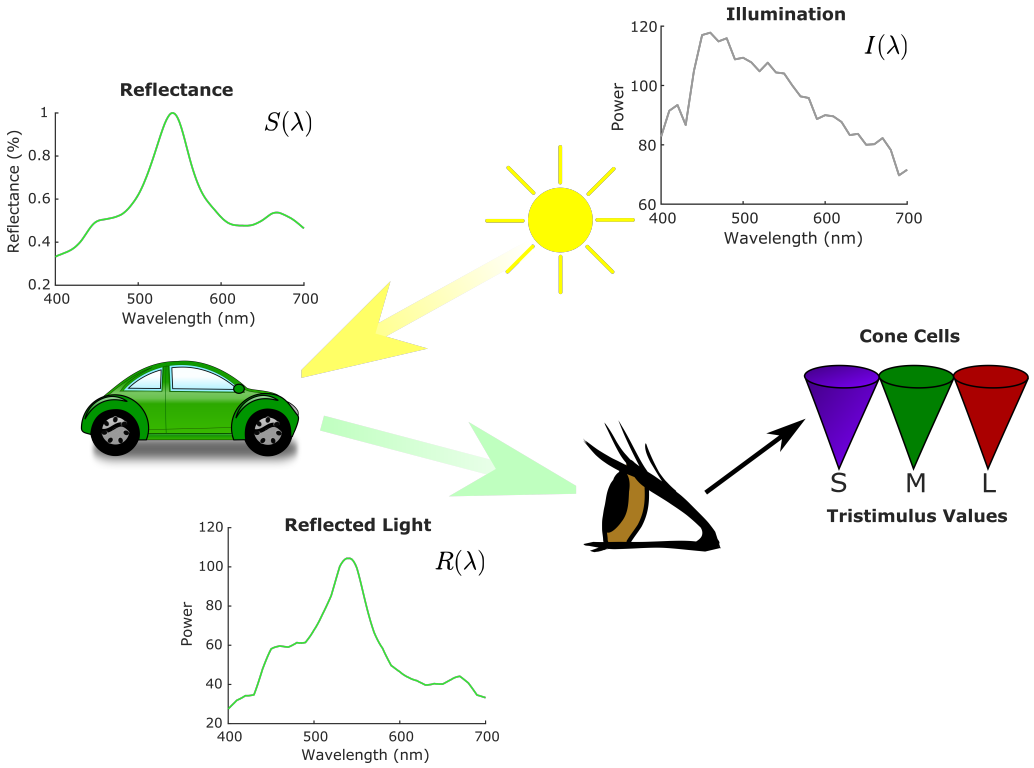


Figure 1.4: Determining the light signal reflected to the eye. The illumination $I(\lambda)$ is reflected from an object according to its surface spectral reflectance function $S(\lambda)$. The reflected light is the point-wise product of these two functions $R(\lambda) = I(\lambda)S(\lambda)$. Once the reflected light is transmitted through the optics of the eye, it reaches the retina at the back of the eye where the information is effectively encoded as just three numbers by the three cone types in the retina (in a trichromatic eye).

Notice that the photoreceptors, the rods and cones, are located at the back of the retina and that light must pass through all other retinal layers before reaching the photoreceptor layer. Once the light reaches the photoreceptor layer, the rods and cones begin responding to the light signal by hyperpolarising, informing the bipolar cells, who in turn inform the ganglion cells, that a light signal has been detected. The axons of the ganglion cells form the optic nerve which delivers the information about the light signal received to the lateral geniculate nucelus (LGN) for further processing.

1.3.1 Scotopic, Mesopic and Photopic Vision

The rod and cone photoreceptors in the human eye have distinctly different functions. The rods mediate scotopic vision (vision at low light levels), while cones mediate photopic

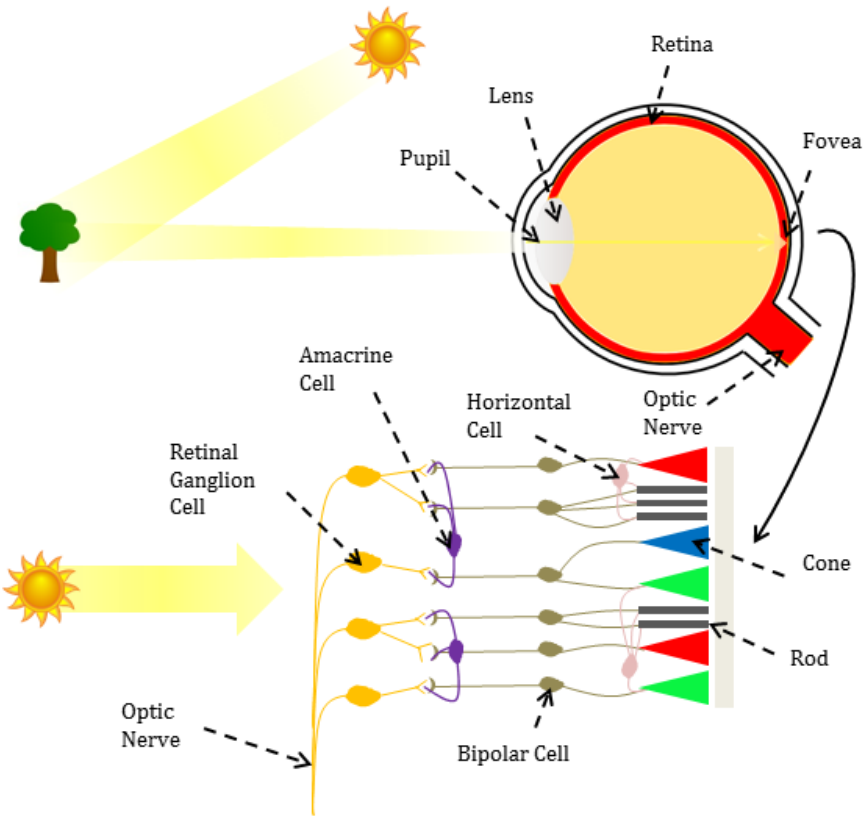


Figure 1.5: The human eye and a section through the retina.

vision (vision at higher light levels). In scotopic conditions, the cones are unresponsive, whereas in photopic conditions the rods are responsive, but are fully saturated. The term mesopic vision can also be used to represent vision in intermediate lighting levels where both rods and cones are contributing to visual processes (Fairchild, 2005). For all studies in this thesis, we will be working at phototopic lighting levels and so do not concern ourselves with the rods.

1.3.2 Trichromacy

Normal human vision is trichromatic, the retina possessing three types of cone receptors: the short (S), medium (M) and long (L) wavelength cones. The cones are named with respect to their spectral sensitivity functions. According to the Stockman & Sharpe (2000) cone fundamentals (Figure 1.6), L-cones have their peak sensitivity at 545 nm, M-cones

at 525 nm, and S-cones at 420 nm. They are also referred to colloquially as the "red", "green", and "blue" cones and we will adopt this terminology at times in the text.

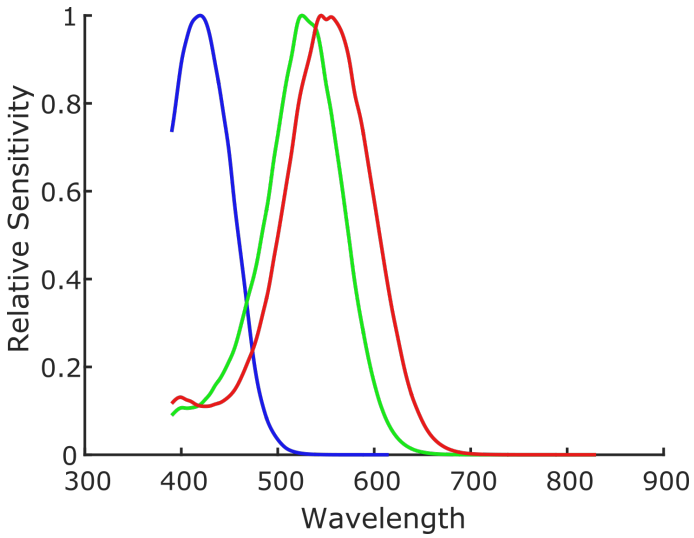


Figure 1.6: Cone sensitivity functions (Stockman & Sharpe, 2000; CVRL.org)

Humans with colour deficiencies can have fewer cone types (monochromats or dichromats), or can have a cone type whose sensitivity is shifted (anomalous trichromats) (Westland, 2002). Recent research has also led to the discovery of human tetrachromats who have four cone types in their retina (Jordan *et al.*, 2010). In these studies, we are concerned only with colour normal observers (those possessing the standard three L, M, and S cone types). While all observers are tested for colour vision deficiencies prior to participating in the various experiments, we do not explicitly test for tetrachromacy in our observers so cannot rule out the possibility of tetrachromats among our participant pool. Considering the infancy of the tetrachromacy research field, however, and the uncertainty that remains over whether observers who express more than three cone types in their retina can exploit the information they obtain, we do not concern ourselves with this issue here.

In photopic lighting conditions, trichromats represent the reflected light spectrum by effectively encoding the information as just three numbers, the responses of the three cone types, that are referred to as the tristimulus values, $[L, M, S]$ (Figure 1.4). The cones may be considered as photon catchers, their spectral sensitivity curves determining the

probability that the cone will catch a photon travelling with a certain wavelength. Each cone's response may then be quantified by counting the number of photon catches or isomerisations within a given time interval. These response properties lead to the principle of univariance that describes the inability of a single cone type to distinguish between different wavelengths of light (Rushton, 1972). For example, consider a hypothetical case where an L-cone is exposed to perfectly monochromatic light stimuli of wavelengths 545 nm and 582 nm (stimuli l_1 and l_2, respectively). If the relative power of stimulus l_2 is 1.25 times that of l_1, the number of isomerisations per time interval will be identical for both stimuli and the cone will be unable to distinguish between them (Figure 1.7). The principle of univariance gives rise to the phenomenon of metamerism. Two light stimuli are considered metameric if they induce identical tristimulus values.

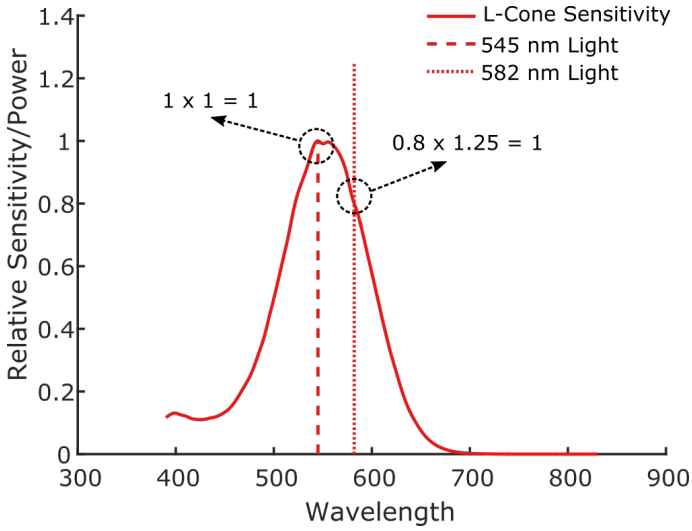


Figure 1.7: An illustration of the principle of univariance. An L-cone will respond the same to a theoretical monochromatic light stimulus at a wavelength of 545 nm, l_1, as it will to a theoretical monochromatic light stimulus at a wavelength of 582 nm, l_2, if the relative power of l_2 is 1.25 times the power of l_1. In this case, an L-cone is unable to distinguish the wavelengths of light and would respond to both light stimuli identically.

1.4 Colour Vision

If the cones are blind to wavelength information, how then do human observers see in colour? The trick is to interpret not the raw cone responses, but instead to trade-off the

cone responses, reducing the correlation between the colour channels caused by the large overlap in the spectral sensitivity functions of the different cone types.

1.4.1 Second Order Mechanisms

Zaidi (1997) showed that responses of the L and M cones are highly correlated for a selection of natural and man-made objects ($r = .99$). However, when Zaidi (1997) computed the $L + M$ and $L - M$ contrasts between the two sets of cone responses, little correlation remained ($r = .21$). This observation had been made earlier by Buchsbaum & Gottschalk (1983). Similarly, Ruderman, Cronin & Chiao (1998) performed a principal components analysis on a set of tristimulus values obtained from hyperspectral images of natural scenes and found that the first three principal axes of variation in the data were a luminance axis ($L + M + S$) and two chromatic axes ($L - M$ and $S - (L + M)$). This suggests it may be optimal to trade-off the cone responses in such a way in visual processing, and indeed, electrophysiological recordings and behavioural data provide evidence for the existence of such channels (three mechanisms: $L+M+S$, $L-M$, and $S-(L+M)$) early in the visual system (Derrington, Krauskopf & Lennie, 1984; Krauskopf, Williams & Heeley, 1982). We will refer to these as the second-order colour mechanisms.

The luminance mechanism (or channel) is considered such because it tracks changes in overall stimulus intensity. Conversely, the chromatic mechanisms define an isoluminant plane. Changes that modulate the $L - M$ mechanism trade off the responses of the L and M-cones creating a "red-green" chromatic channel. Similarly, changes that modulate the $S - (L + M)$ mechanism create a "blue-yellow" chromatic channel, trading off S-cone responses with the sum of the L and M-cone responses. Krauskopf *et al.* (1982) concluded from a set of psychophysical data that there are three second-order channels but that these are governed by six independent mechanisms, one for each half-axis of change (redder changes: $L - M$, greener changes: $M - L$, bluer changes: $S - (L + M)$, yellower changes: $(L+M) - S$, luminance increments: $L+M+S$, and luminance decrements $-L-M-S$).

In a first experiment, Krauskopf *et al.* (1982) defined a three-dimensional colour space

with axes $L+M$ (the luminance axis), $L-M$ (the red-green axis), and $S-(L+M)$ (the blue-yellow axis). They refer to these as the cardinal axes of this colour space, terminology that we will adopt throughout the text. They established thresholds for their participants for stimulus changes from the white point of the space (the average adaptation state of the participant) along each direction of each cardinal axis both before and after a habituation stimulus. The habituation stimulus was sinusoidally modulated along one of the cardinal axes such that the average chromaticity of the habituation stimulus was equal to the chromaticity of the white point of the space. The rationale is that the average adaptation state of the observer remains the same while the sensitivity of the hypothesised cardinal mechanism is selectively reduced. The test stimulus for which thresholds were established was a Gaussian enveloped pulse and they found thresholds for this stimulus using a yes/no staircase procedure. When Krauskopf *et al.* (1982) compared thresholds before and after habituation to each cardinal axis in the isoluminant plane of their colour space ($L+M$ is fixed), they found an increase in thresholds along the axis of habituation but not along the orthogonal isoluminant cardinal axis. Thresholds along intermediate axes were moderately increased. However, when the habituation stimulus modulated along intermediate axes, chosen to equally stimulate the two cardinal axes, thresholds were increased in all directions around the white point with only slight evidence of selective habituation. They take this as evidence that the cardinal mechanisms of their colour space have a special status.

In a second experiment, Krauskopf *et al.* (1982) used a sawtooth habituation stimulus rather than a sinusoidally modulating one. In this case, two types of habituation stimuli could be used. For example, for the red-green ($L-M$) cardinal axis, there was a habituation stimulus that modulated slowly from red to green in the slow phase of the sawtooth wave then returned quickly to red, and another where the stimulus modulated slowly from green to red in the slow phase then returned quickly to green. The test stimuli in this case was also changed to be a step change that did not return to the white point until the participant had responded, ensuring only the increment/decrement half of the axis was detecting the change. Using these stimuli, they found that thresholds were variably

increased depending on the sign of the sawtooth (the direction of the slow change). For a slow red to green phase, thresholds for detection of the green step change were increased more than for the red step change. For a slow green to red phase, thresholds for detection of the red step change were increased more than for the green step change. There was a similar pattern for thresholds along the blue-yellow cardinal axes. These results suggest separate mechanisms for the detection of increment/decrement changes along each chromatic cardinal axes. In a different study, they found similar effects for the luminance axis (Krauskopf, 1980).

1.4.2 Higher-Order Mechanisms

Krauskopf, Williams, Mandler & Brown (1986) later re-analysed the data of Krauskopf *et al.* (1982) and found evidence of higher order mechanisms (that appear later in the visual pathway or require further processing). They decomposed the threshold elevations as a function of test angle using Fourier analysis and found that the phase of the second harmonic tracked the angle of the test. This characteristic of the data is not consistent with the assumption of only two second order chromatic mechanisms, but rather with the existence of further higher order chromatic mechanisms that are tuned to intermediate directions (non-cardinal) in the isoluminant plane.

Webster & Mollon (1991) also suggested the existence of higher order mechanisms based on colour matching data. They used a similar procedure to Krauskopf *et al.* (1982), showing a sinusoidally modulating adapting stimulus followed by a test stimulus. However, in their procedure, participants then adjusted a second test field presented in the other half of the visual field to match the appearance of the first test. They adapted participants separately to the two cardinal chromatic axes and to six intermediate axes (at 22.5°, 45°, and 67.5° angles to cardinal axes). In all conditions, they find a selective effect on colour appearance matches. Matches were increased in saturation for tests along the adapting axes relative to tests along an orthogonal axes. In addition, intermediate tests were shifted in hue towards an orthogonal axis. This was true for the six intermediate adapting axes in

addition to the cardinal axes, although two out of four observers displayed more selectivity for the cardinal chromatic mechanisms than others. The authors concluded that, if the chromatic mechanisms act independently, then at least eight mechanisms are needed to encode such changes in colour appearance. However, they also acknowledge that the two cardinal mechanisms may not adapt independently. Zaidi & Shapiro (1993) also suggested that the two cardinal mechanisms interact, proposing a model where the two mechanisms jointly adapt that accounted for their data. In recent times, there has been suggestions of a greater number of higher-order colour mechanisms (Hansen & Gegenfurtner, 2013), but Shepard, Swanson, McCarthy, Eskew & Eskew Jr. ,Rhea T. (2016) offered a six mechanism model (three bipolar channels) explanation that could account for the data, akin to the original suggestion of Krauskopf *et al.* (1982). In sum, uncertainty remains over the number of chromatic mechanisms that are responsible for colour processing and there is no clear consensus among researchers in this field.

1.4.3 Where is colour processed?

In the last section, we introduced the idea of cone-opponent or second-order colour mechanisms and went on further to discuss the possibility of higher-order colour mechanisms. However, we did not discuss where in the visual pathway such mechanisms may arise. We briefly mentioned the study of Derrington *et al.* (1984) who found physiological evidence of second-order mechanisms. More specifically, they found neurons in the lateral geniculate nucleus (LGN) of macaque that display chromatic opponency. Other groups had previously recorded from cells that displayed chromatic opponency (Wiesel & Hubel, 1966; De Valois, Abramov & Jacobs, 1966) but Derrington *et al.* (1984) were the first to quantify them. However, cone-opponent processing begins much earlier than the LGN, starting in retinal processing with the biplolar and ganglion cells. The three cardinal second-order mechanisms are then processed somewhat distinctly along different channels as the signals are passed from retina to cortex (Gegenfurtner & Kiper, 2003).

The postulated higher-order mechanisms are often considered as recombinations of the

cardinal channels (Krauskopf *et al.*, 1986; Webster & Mollon, 1991), but as Eskew (2009) points out, the idea of higher-order mechanisms is two pronged. The first part of the argument is that there are further chromatic channels beyond the cardinal channels, the second is that the mechanisms are to be found at some later point in the visual pathway. Eskew (2009) goes on to argue that given our lack of knowledge over which cortical areas are even the most important for colour processing, and that there are numerous feedback connections from cortex to the LGN, it is speculative at best to posit that higher-order colour mechanisms must exist at a cortical level - if they even exist at all (as mentioned above, Eskew's group favours the six mechanisms explanation).

1.5 Colour Constancy

It has been proposed that trichromatic colour vision evolved to enable the detection of ripe fuit against the foliage (Mollon, 1989). Although this explanation for the natural pressures that led to trichromatic vision in humans is debated (Gegenfurtner & Kiper, 2003), there are studies that show the usefulness of colour in visual processing. The addition of colour information has been shown to enable faster object recognition and improve visual memory (Gegenfurtner & Rieger, 2000; Spence, Wong, Rusan & Rastegar, 2006). However, if a percept of colour is to provide useful information about the properties of objects in the world, then our percept of colour must be somewhat invariant, and indeed, the human visual system is said to be approximately colour constant, maintaining a relatively stable colour percept of surfaces across illumination changes (Brainard & Radonjić, 2014; Foster, 2011; Hurlbert, 1998; Maloney, 1999; Smithson, 2005).

This is no easy feat. As we have already detailed, the trichromatic human visual system maps an infinite dimensional stimulus space to a space with only three dimensions. This is a mapping from the space of all theoretically possible reflected light spectra to the isomerisations, or a set of tristimulus values. We have seen that the spectrum of light reflected from a surface is determined by both the surface's spectral reflectance function and the spectrum of the incident illumination. The spectral reflectance function of a

surface is a fixed property of the object; thus a strategy for colour constancy would be to perceive a correlate of spectral reflectance that is independent of the incident illumination. However, the reduction in dimensionality confounds surface reflectance and illumination information in an irreversible way. Moreover, if the light illuminating an object changes, changing the spectrum of light reflected to the eye, and so too the corresponding tristimulus values, it is impossible to tell from the change in the tristimulus values alone that this was due to a change in the illumination rather than the surface's spectral reflectance. This rules out a one-to-one mapping between isomerisations and perceived surface colour. The mechanisms by which perceived surface colour is determined by the visual system, and hence by which approximate colour constancy is achieved, remain unknown.

Several seminal studies have shown that the human visual system rarely displays perfect colour constancy (e.g. Arend & Reeves, 1986; Arend, Reeves & Goldstein, 1991; Brainard, 1998; Brainard, Brunt & Speigle, 1997), although the level of reported colour constancy varies greatly between studies (see Table 1 in Foster (2011)). Importantly, colour constancy tends to be better for real scenes (real surfaces under changing illuminations) compared to simulated scenes (Maloney, 1999). Further, the level of colour constancy achieved by observers can be modulated by altering task instructions. Arend & Reeves (1986) showed that when asked to adjust the colour of a reference patch seen under one illumination to match that of a test patch seen under another illumination in terms of hue and saturation (an appearance match), observers display less colour constancy than when asked to make the patch to look as if it were cut from the same piece of paper. In addition, Radonjić & Brainard (2016) showed that instructional effects are greater for simpler experimental stimuli. However, Arend & Reeves (1986) allowed observers to freely move their gaze between the two scenes (viewed binocularly), meaning adaptation to the illumination in either the reference or test scene was never complete. Similarly, while Radonjić & Brainard (2016) restricted the view to a single scene (either monocularly viewed for the simple stimulus and stereoscopically viewed for the more complex stimulus), multiple illuminations were present in the scene and observers could shift their gaze to focus on

either part of the image. Bramwell & Hurlbert (1996) suggested the use of haploscopically or successively viewed stimuli to allow full adaptation to the illumination in each scene. In this case, the difference between the two types of instructions is reduced.

Regardless, as real world scenes are likely to contain multiple illuminations (due to factors like illumination gradients and shadows; Nascimento *et al.*, 2016), a more complicated definition of colour constancy is required: colour constancy at the colour appearance or sensory level and colour constancy at the object identification or cognitive level. The difference between these two types of colour constancy, as we define them, can be thought of as a distinction between conscious vs. unconscious levels. If colour constancy is happening at the sensory level, then the observer does not have access to the alternative colour percept of a surface caused by the change in the illumination. However, if constancy is achieved at a cognitive level, the observer may be expected to have access to the alternative colour percept. Note that we do not claim that these two definitions of colour constancy are competing hypotheses, rather we suggest that colour constancy happens at both levels. In this thesis, we employ a behavioural task that obtains an upper limit of sensory level colour constancy as will be discussed later.

1.5.1 Strategies for colour constancy

1.5.1.1 Illumination estimation

It is generally assumed that as part of the colour constancy process, the human visual system attempts to recover information about the surface reflectance function of the object by estimating the incident illumination and discounting this from the scene (generally referred to as the illumination-estimation hypothesis (Maloney, 1999; Maloney & Yang, 2003)). Illumination estimation may be considered as either a sensory or cognitive level solution to colour constancy. Above, we defined a cognitive solution as a part of the process that the observer is conscious of. Other researchers (for example Troost & Weert, 1991) use the term to refer to computations that may simply be higher level, in which case,

they would consider the computation of separate representations of the illumination and surface reflectance to be a cognitive mechanism (if both representations were accessible by the visual system), even if this results in colour constancy at the sensory level (the observer is unaware of the change in reflected light). If we adopt our use of the terminology in the previous paragraph, elements of illumination estimation may be both sensory and cognitive. Part of the process may involve unconscious estimation of the illumination such that the observer's phenomenal experience is of a scene where an illumination estimate has already been discounted. Alternatively, the observer may form a conscious estimate of the illumination in the image, using this knowledge to assign colour names to objects that are not a direct correlate of their sensory or phenomenal experience (c.f. the paper match of Arend & Reeves, 1986).

Most computational algorithms for colour constancy take the illumination-estimation approach, attempting to recover a representation of the illumination (Hurlbert, 1998; Maloney, 1999; Maloney & Yang, 2003). However, perfect recovery of the spectral reflectance function and the incident illumination spectrum from the set of tristimulus values is an ill-posed problem unless the visual system uses knowledge of the statistics of the environment to form constraints (Maloney, 1999).

1.5.1.2 Natural constraints

One way to do this is to re-express the set of natural surface reflectance functions and/or the set of natural illumination spectra as linear combinations of a small set of basis functions. Maloney & Wandell (1986) showed that if the representation of possible surface reflectance and illumination spectra as linear models is to allow the colour constancy problem to be solved analytically, further constraints must be met. Firstly, the linear model that characterises illuminations must have dimension less than or equal to the number of receptor types in the imaging device; secondly, the linear model that characterises surfaces must have dimension strictly less than the number of receptor types. As the number of photoreceptor types in the retina responsible for colour vision is three, we

require illuminations to be constrained by a set of at most three basis functions and surfaces to be constrained by a set of at most two. Analysis of natural illumination spectra and sets of surface reflectance functions suggest that this is not the case; linear models of higher dimensionality are often needed to represent the set of measurements (Jaaskelainen, Parkkinen & Toyooka, 1990; Judd, MacAdam & Wyszecki, 1964; Maloney, 1986; Marimont & Wandell, 1992). However, if the sensitivity of human photoreceptors is taken into account, a linear model of lower dimension can be found for surface reflectance.

1.5.1.3 Prior knowledge

A second approach to constraining the problem would be the formation of illumination or surface reflectance priors. Priors place probabilities on the occurrence of different stimulus characteristics or parameters that may be encountered in the world and are believed to be formed through experience. They are a cornerstone of Bayesian theories of visual perception (Allred, 2012; Mamassian, Landy & Maloney, 2003; Brainard, 2009).

The effect of a prior on sensory processing can be modelled as follows. Suppose the observer receives some noisy sensory input I. A variety of different world states W could have given rise to the sensory input and we assume the observer knows the generative model that describes the probability of different world states producing various sensory input. This gives rise to the likelihood function, or the probability of the sensory input I given the different world states W, $P(I|W)$. The observer wants to recover a best guess as to the state of the world and can do that if they calculate the posterior function $P(W|I)$ (the probability of different world states given the sensory input). Suppose also that, over time, the observer has accumulated knowledge of the probability of encountering different world states, $P(W)$ (the prior), then Bayes' Theorem tells us that $P(W|I) = cP(I|W)P(W)$, where c is the normalisation constant $1/P(I)$. Note that Bayes' Theorem describes the optimal way to combine two sources of information and hence is also used in the cue combination literature to describe the optimal way to combine evidence either within or between senses where the prior in replaced by the likelihood function for the second sensory

input (e.g. Ernst & Banks, 2002).

In the visual perception literature, there is evidence for the use of priors in judgements of lighting direction (Adams, Graf & Ernst, 2004), colour (Hansen, Olkkonen, Walter & Gegenfurtner, 2006), and speed (Stocker & Simoncelli, 2006), as well as evidence of observers learning priors while performing a novel task (e.g. Bejjanki *et al.*, 2016; Tassinari *et al.*, 2006; Laquitaine & Gardner, 2017). In colour perception, the human visual system is thought to learn priors over daylight chromaticities as they display high levels of regularity - measured daylight chromaticities tend to fall on the Planckian locus (Hernández-Andrés, Romero, Nieves & Lee, 2001; Judd, MacAdam & Wyszecki, 1964; Spitschan, Aguirre, Brainard & Sweeney, 2016) - and surface reflectance properties, often thought to be tied to contextual cues and object perception (Hansen, Olkkonen, Walter & Gegenfurtner, 2006; Olkkonen & Allred, 2014; Witzel, Valkova, Hansen & Gegenfurtner, 2011). As will become clear, the task that we use here is designed to explore the hypothesis that observers possess a daylight illumination prior and, more generally, to assess whether colour constancy mechanisms are optimised to maintain constant surface colour perception in more natural lighting.

Note that these two approaches to constraining the colour constancy problem are not distinct. For example, one could use the constraints imposed by requiring illuminations that are physically realisable to place a uniform prior over the set of possible coefficient vectors that would be used to recombine the basis functions (Forsyth, 1990). Brainard & Freeman (1997) took this one step further, specifying priors over the coefficients that scale the basis functions for both illuminations and surfaces. Specifically, they used a set of measured daylight chromaticities to specify a "daylight prior" for illuminations. By incorporating these priors they developed an illumination estimation algorithm that can recover good estimates of the illumination spectrum present on a scene. They later showed that a modified version of the model that no longer estimates the illumination spectrum but rather the illumination chromaticity (using an equivalent illumination approach to colour constancy (Brainard & Maloney, 2011)) can be fit well to behavioural data. However,

model performance is improved by incorporating a prior for illuminations that is broader than the "daylight prior" and includes more atypical illumination chromaticities (Brainard, Longère, Delahunt, Freeman, Kraft & Xiao, 2006).

1.5.1.4 Lightness algorithms

It may not be necessary, however, to form an explicit estimate of the illumination in order to achieve the constant perception of colours. Indeed, several early models of colour constancy employ algorithms that imply colour constancy is achieved through adaptation to the illumination in distinct chromatic channels. Such methods fall in the general class of linear transforms (Hurlbert, 1998) and include "lightness" algorithms, and algorithms such as von Kries scaling and Retinex (see Hurlbert, 1998, Maloney, 1999 or Foster, 2011 for a review of such algorithms). We will only review the most simple of these algorithms briefly here to emphasise why they fail and why more complex colour constancy algorithms are needed to explain human behaviour.

von Kries (1878) suggested that the visual system compensates for a change in illumination (or discounts the illumination) by separately scaling each cone type, a form of adaptation. In practice, von Kries scaling is implemented as a linear transformation of the tristimulus values. As the theory assumes that each cone type adapts independently, the transformation matrix is a diagonal matrix such that the transformed tristimulus values corresponding to a particular point in the scene, $[L', M', S']$, are defined as:

$$\begin{bmatrix} L' \\ M' \\ S' \end{bmatrix} = \begin{bmatrix} l & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & s \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix} \tag{1.2}$$

The diagonal elements of the transformation matrix, l, m, and s, are the factors that each cone type is scaled by and may be considered a correlate of the illumination. Multiple methods have been proposed by which to calculate these scaling values. For example, one is the grey-world hypothesis that assumes the average reflectance across a scene is neutral

(flat) on average (Evans, 1951). In this case, l, m, and s would be found by taking the inverse of the average tristimulus values across the whole scene. A second example is the brightest-is-white hypothesis that assumes that the brightest surface in the scene is white and reflects exactly the illumination (Land & McCann, 1971; Brill & West, 1981; Barnard, Cardei & Funt, 2002). In this case, l, m, and s would be taken from the brightest point in the scene. Estimation of the illumination (or of l, m, and s), is a limiting factor for how well these algorithms can do; firstly at performing a successful colour constancy correction and secondly at predicting human behaviour. Indeed, these assumptions often fail to hold. The average reflectance across scenes is rarely neutral and the brightest point in the image is likely to be chromatically biased (Foster, 2011).

1.5.1.5 Spatial ratios of cone excitations

Foster & Nascimento (1994) showed that it is possible to compute a signal at the level of retinal processing that is invariant to global illumination changes. After simulating daylight illuminations on a set of Munsell surfaces, they found minimal variation in the excitation coordinates of the M and L-cones recorded from a signal surface across illumination changes, but larger variations in S-cone signals (their Figure 2). However, when they considered the spatial ratio between two surfaces under illumination $I_1(\lambda)$ to the spatial ratio of the same two surfaces under illumination $I_2(\lambda)$, they found almost perfect invariance (their Figure 3). The same group has since shown that the invariance of spatial cone ratios holds in natural scenes and hence could be a useful computation for colour constancy (Nascimento, Ferreira & Foster, 2002; Foster, Amano & Nascimento, 2015).

1.5.2 Measuring colour constancy

Colour constancy is generally assessed using either asymmetric matching (matching of colours across scenes with either simultaneous or successive views, e.g. Arend & Reeves, 1986; Brainard *et al.*, 1997; Troost & Weert, 1991), achromatic adjustment (adjust a patch of colour to appear neither blue, yellow, red, or green, e.g. Brainard, 1998), colour naming

(e.g. Troost & Weert, 1991), or by asking observers to make judgements about illumination vs. surface reflectance changes (operational colour constancy tasks, e.g. Craven & Foster, 1992). These types of measures each have methodological issues (Foster, 2003; Maloney, 1999). As we have already mentioned, task instructions can influence surface colour adjustments, a problem for both asymmetric matching and achromatic adjustment tasks. Colour naming studies lack the resolution required to detect deviations from constancy, as often only a fixed set of colour names are allowed. Finally, while operational colour constancy tasks may avoid these issues, a correct response on the task does not directly imply colour constancy. One cannot conclude that correct discrimination of illumination vs. surface reflectance changes implies the colour appearance of objects remained stable.

Pearce *et al.* (2014) introduced a new method for measuring colour constancy by means of illumination discrimination thresholds: the illumination discrimination task (IDT). This task provides a natural measure of colour constancy and is the approach we take in the current study. On each trial of this task, observers first see a Mondrian-papered scene illuminated by a reference light. After small delays they are presented with two comparisons; one is the target light and the other a test light. The target light is always identical to the reference while the test varies from the reference along one of four specified axes in the $CIE\ xy$ chromaticity plane; the test is either bluer, yellower, redder or greener in colour appearance than the reference. A staircase procedure that modulates the distance of the test from the reference along each axis of change (distance measured in a perceptually uniform colour space - CIELUV) is used to find thresholds for correct identification of the target among the two comparison illuminations. The recovered threshold may then be considered as an upper limit of sensory colour constancy. For illumination changes to an illumination with chromaticity closer to the reference than the threshold, the observer remains colour constant, for chromaticities further away they do not. The logic is as follows. If an observer cannot correctly and reliably detect the target light, then they cannot discriminate between the two comparison scenes and the two comparison scenes must appear the same to the observer. Thus, the observer remained colour constant as

colour appearance remained stable despite an illumination change.

1.5.3 Evidence for a "blue bias" in colour constancy

An important feature of the IDT is that the bluer and yellower axes of change are parametrised to fall along the Planckian locus such that their chromaticities match those typical of daylights. The redder and greener axes of change fall along a line perpendicular to the Planckian locus in the $CIE\ u'v'$ chromaticity plane, taking chromaticities that are more atypical of natural illuminations. This allows one to probe whether human colour constancy mechanisms are optimised for daylights. Pearce *et al.* (2014) found a "blue bias" in the IDT thresholds, observers displaying better colour constancy for illumination changes along the bluer axis of change (mean thresholds were higher along this axis suggesting observers are more tolerant to bluer illumination changes from their neutral D67 reference illumination).

Measurements of daylight chromaticities from Hernández-Andrés *et al.* (2001) are skewed towards the bluer side of the D67 reference light (but see also Nascimento *et al.* (2016) who show that local illumination colours follow a Gaussian distribution centred on $\approx$ D57 which is in the yellow region of our experimental illuminations). Pearce *et al.* (2014) digitised this data from the plots in the original paper to compare it to the illuminations they used in the experiment (their Figure 1.D). They conclude that "[b]lueish illuminations are the most common among daylight illuminations". We obtained the raw data from `http://colorimaginglab.ugr.es/pages/Data#__doku_granada_daylight_spectral_database` and plot it in Figure 1.8.B. In addition, we plot histograms of the $CIE\ xy$ coordinates of the measurements, showing that while the chromaticities of the measured daylights are skewed towards the bluer illuminations used in the IDT, daylights are more likely to have a chromaticity corresponding to illuminations classified as yellower in the experiment. This is also a feature of a more recent set of daylight measurements (Figure 3 in Spitschan *et al.*, 2016). However, if we consider the measurements relative to the chromaticity of equal energy white ($(x, y) = (0.33, 0.33)$, or illuminant E), then at

least along the x dimension, one would conclude that illuminations are likely to be bluer relative to neutral.

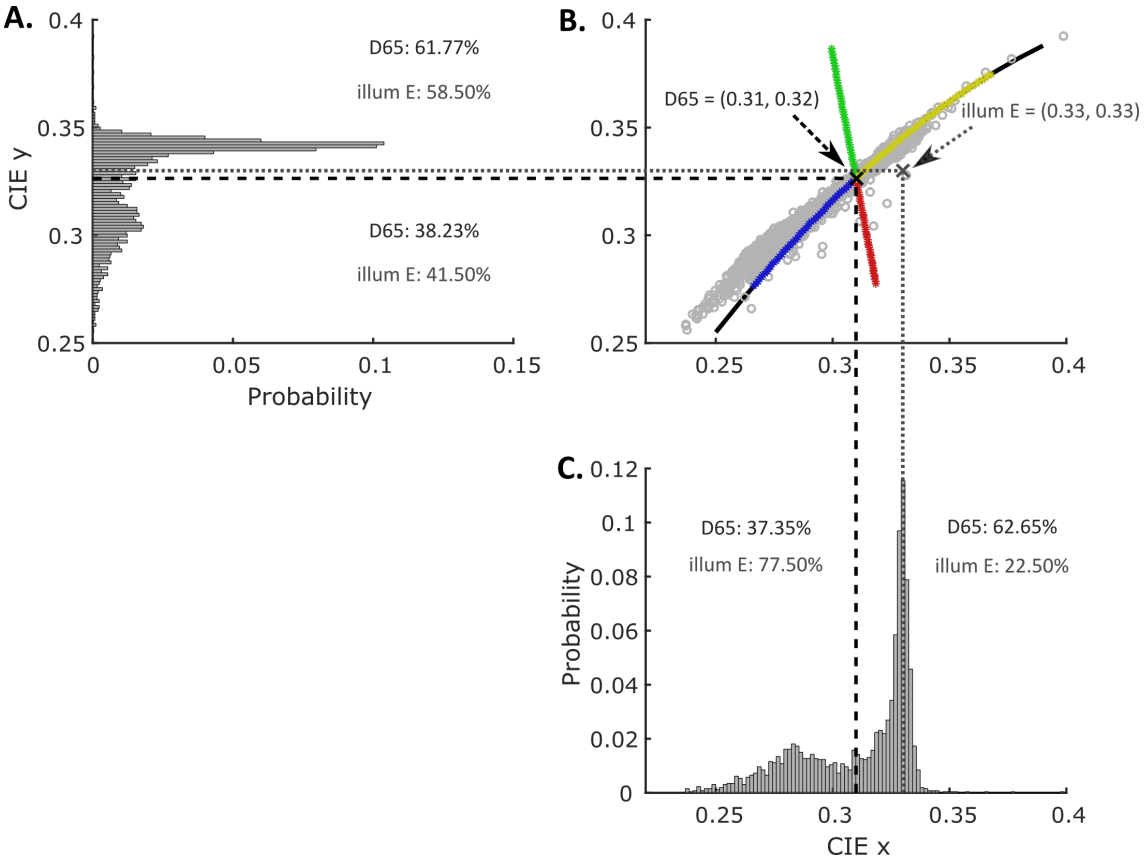


Figure 1.8: Measurements of daylight chromaticities (Hernández-Andrés *et al.*, 2001). A. Histogram of the $CIE\ y$ values of the measurements. B. The measurements plotted in the $CIE\ xy$ isoluminant plane (grey open circles) with the experimental illuminations from the IDT (bluer, yellower, redder, and greener illuminations) superimposed. The black solid line is the Planckian locus. C. Histogram of the $CIE\ x$ values of the measurements. The black cross and black dashed lines correspond to the reference illumination chromaticity (D65/D67, which are equivalent to 2 d.p.). The grey cross and grey dotted lines correspond to illuminant E, or equal energy white. The percentages refer to the proportion of illuminations with $CIE\ x/y$ chromaticity falling to either side of the reference illumination chromaticity, or the chromaticity of illuminant E.

While this contradicts the claim of Pearce *et al.* (2014) (if we assume they call illuminations with both x and y chromaticity values smaller than their neutral reference blue and yellow otherwise), it does not contradict their hypothesis that human colour constancy mechanisms are optimised for the statistics of the natural world. Of course, a statement that daylights are more likely to be blueish relative to the neutral reference light used in the experiment would be incorrect (it should be noted that the reference also serves as

the average adaptation point during the experiment and is used as the white point of the perceptually uniform space in which thresholds are calculated). These data (Hernández-Andrés *et al.*, 2001; Spitschan *et al.*, 2016) are more supportive of the statement that daylights are more likely to be yellowish (relative to the neutral reference in the experiment). However, there is more variability among daylights whose chromaticities fall in the bluer region of the experimental illuminations compared to those in the yellower region. Moreover, the measurements of daylights encompass the bluer illuminations used in the experiment better than the yellower illuminations (Figure 1.8). Put another way, bluer daylights may be more saturated than yellower daylights (again, relative to the neutral reference illumination). It may still be the case that colour constancy mechanisms are optimised for natural illuminations in the sense that they are optimised to deal well with the variation in natural illumination chromaticities. In this framework, one would expect better colour constancy (reduced discrimination) along the bluer axis of change in the experiment as the more saturated experimental illuminations (further away from the reference light) along this axis are more likely in the natural world than the more saturated experimental illuminations along the yellower axis of change.

The results of Pearce *et al.* (2014) have been replicated by Radonjić *et al.* (2016*b*) and Alvaro, Linhares, Moreira, Lillo & Nascimento (2017). Radonjić *et al.* (2016*b*) repeated the study of Pearce *et al.* (2014) with a slightly different real scene and also with a well-matched computer simulated scene. In both cases, their findings agreed with those of Pearce *et al.* (2014). Thresholds along the bluer direction of chromatic change in the illumination were higher than those along the three other axes of change both for the real and simulated scene. Participants had worse ability to discriminate the illumination changes along the bluer axis of chromatic change relative to the others, indicating colour constancy was better or that they remained colour constant for larger chromatic changes along this axis.

Alvaro *et al.* (2017) used a modified version of the IDT, rendering hyperspectral images of natural scenes under daylight illuminations. They used four different scenes, two rural

and two urban scenes. In this study, the authors only tested colour constancy along the Planckian locus (the bluer and yellower chromatic directions of change) and did not test the redder and greener directions of change. Like the previous studies, they found colour constancy to be better for the bluer illuminations, a finding that also held in a group of colour vision deficient observers (red-green dichromats).

The idea that colour constancy mechanisms are optimised for natural illumination changes was explored earlier by Delahunt & Brainard (2004). Initially, Delahunt & Brainard (2004) had observers adjust a test patch in simulated scenes of matte surfaces to appear achromatic (achromatic matching method). In their first experiment, the scene was rendered under one of five experimental illuminations. A neutral (D65) illumination (relative to which constancy was calculated), and a blue, yellow, red, and green illumination. The latter four illuminations were specified such that their chromaticity (measured from a Lambertian surface) was equidistant from the neutral illumination in the isoluminant CIELUV plane (60 $\Delta E_{u^*v^*}$ away, see 2.1.5). Importantly, as in the IDT, the blue and yellow illuminations fell along the Planckian locus, an estimate of the daylight locus (representing typical illumination changes), while the red and green illuminations are parametrised to fall off the axis of daylight changes (atypical illumination changes). While there was no significant main effect of direction of illuminations change (blue, yellow, red, or green) on the level of colour constancy achieved by observers, constancy was best for the blue illumination, suggesting an underlying "blue bias". In a second experiment, the authors hypothesised that if prior knowledge of likely illumination changes in the natural world does promote better colour constancy for natural illumination changes, any such effect may not be visible in their data as constancy was high in general. To investigate this further, they designed a set of stimuli that would result in worse constancy for all conditions by introducing invalid cues to the illumination change. For these stimuli, the pattern of constancy levels across the four illumination change conditions is consistent with the first experiment (blue > green > yellow > red), but they now find a significant main effect. In a final experiment, the authors confirmed that the effect was independent of the magni-

tude of the difference between the neutral and experimental illuminations. They repeated both their first and second experiments with a blue, yellow, red, and green illumination that were closer to the neutral illumination in the isoluminant CIELUV plane (30 $\Delta E_{u^*v^*}$ away) and found the same results.

Radonjić & Brainard (2016) found mixed evidence for a "blue bias" in colour constancy. In this study, observers completed both a colour selection ("find the best match") and asymmetric matching task under a blue and yellow illumination change condition. Again, constancy was calculated relative to a neutral (D65) illumination. The blue and yellow illuminations were defined to fall along the Planckian locus (colour temperatures of 12000 K and 4500 K, respectively). Colour constancy was significantly better for the blue compared to the yellow illumination change when scenes were more complex (simulated three-dimensional scenes), but not for more simple scenes (diffusely illuminated two-dimensional scenes of matte surfaces). Other studies also found no evidence for optimisation of colour constancy mechanisms for the illuminations that are most likely in nature (Brainard, 1998; Foster, Amano & Nascimento, 2003; Rüttiger, Mayser, Sérey & Sharpe, 2001).

In sum, there is mixed evidence for a "blue bias" in colour constancy, or for an optimisation of colour constancy mechanisms to properties of the natural world. In addition, it is unclear what the underlying mechanisms are that would mediate such a behavioural effect. The bias may come about through a learnt illumination prior that captures the natural variability in daylight illuminations (daylights are more variable along the bluer end of the spectrum). Alternatively, the bias may arise through reduced sensitivity to "blueish" global changes in scene chromaticity, the human visual system having evolved to be less sensitive to these types of changes, promoting colour constancy under bluer illumination changes. While it is also true that the ability to learn an illumination prior will have been selected for during evolution, we consider these two hypotheses as different due to the following distinction. The former hypothesis, that will be referred to from here on as the nurture hypothesis, suggests that colour constancy mechanisms will develop differently

during different individuals' lifetimes due to different illumination exposures - observers will develop their own specific prior. This may be expected to manifest as individual differences in behaviour (see Chapter 4). The latter hypothesis, that we will refer to as the nature hypothesis, suggests that in some sense, the human visual system has an engrained illumination prior, learnt during evolution, in the form of a reduced sensitivity to bluer changes in illumination - a species prior. In this case, constancy mechanisms are expected to be consistent across observers. The experiments of this thesis attempt to distinguish between these two hypotheses.

1.6 Publications arising from this thesis

1.6.0.1 Peer reviewed papers

Aston, S., & Hurlbert, A. C. (2017) What #theDress reveals about the role of illumination priors in color perception and color constancy. *Journal of Vision*. (From Chapter 5).

Aston, S., Gross, N., Fitzpatrick, D., Olkkonen, M. & Hurlbert, A. C. (in prep) Memory effects, central tendency, serial dependency or just task bias? An investigation using illumination hue discrimination. (From Chapter 6).

Aston, S. & Hurlbert, A. C. (in prep) Illumination Discrimination for Chromatically Biased Scenes. (From Chapter 3).

1.6.0.2 Conference abstracts

Aston, S., Groombridge, J., Pearce, B. & Hurlbert, A. (2015) Thresholds for colour constancy measured via illumination discrimination depend on adaptation point. *ECVP 2015 - PERCEPTION*. (From Chapter 3).

Aston, S., Olkkonen, M. & Hurlbert, A. C. (2017) Memory bias for illumination colour. *VSS 2017 - Journal of Vision*. (From Chapter 6).

Aston, S., Olkkonen, M. & Hurlbert, A. C. (2017) Memory effects, central tendency, serial dependency or just task bias? An investigation using illumination hue discrimination. *ECVP 2017*. (From Chapter 6).

Chapter 2

Measurements and Calibration

2.1 Quantifying Visual Responses to Light

2.1.1 Photometry

Photometric measurements characterise light in terms of what the human eye perceives. In photometry, radiometric measurements are weighted by the luminosity function, $V(\lambda)$. Weighting radiometric measurements by the luminosity function accounts for the fact that the eye is not equally sensitive to light at all wavelengths. In section 1.2.2.1, we introduced spectral radiance $R(\lambda)$ and spectral irradiance $I(\lambda)$. Equivalent photometric measures can be defined by weighting the spectral radiance and the spectral irradiance by the luminosity function and multiplying by 683 to transform to photometric units; this gives us the spectral luminance, $S_l(\lambda)$, and spectral illuminance, $S_i(\lambda)$, respectively. We can express these quantities as:

$$S_l(\lambda) = 683R(\lambda)V(\lambda) \tag{2.1}$$

and

$$S_i(\lambda) = 683I(\lambda)V(\lambda) \tag{2.2}$$

As we did with the radiometric quantities, we can drop the spectral information and consider just luminance and illuminance. The units for luminance are candelas per square meter (cd/m^2) and the unit for illuminance is lux which is equal to lumens per square meter ($lux = lm/m^2$).

2.1.2 Colour-matching Functions

The colour matching functions (CMFs) were introduced through psychophysical experiments of trichromatic matching and they are used to represent lights and surfaces by

$[X, Y, Z]$ tristimulus values (Hunt, 1998). The functions $\bar{x}$, $\bar{y}$ and $\bar{z}$ are illustrated in Figure 2.1. Again, the $[X, Y, Z]$ tristimulus values of a particular light or surface can be found by integrating over all wavelengths the point-wise multiplication of the three colour-matching functions by the SPD of the illumination or the spectral radiance of the light reflected from a surface. Formally, suppose a light has SPD $S(\lambda)$, then

$$X = 683 \int_{\lambda} \bar{x}(\lambda) S(\lambda) d\lambda, \tag{2.3}$$

$$Y = 683 \int_{\lambda} \bar{y}(\lambda) S(\lambda) d\lambda, \tag{2.4}$$

$$Z = 683 \int_{\lambda} \bar{z}(\lambda) S(\lambda) d\lambda. \tag{2.5}$$

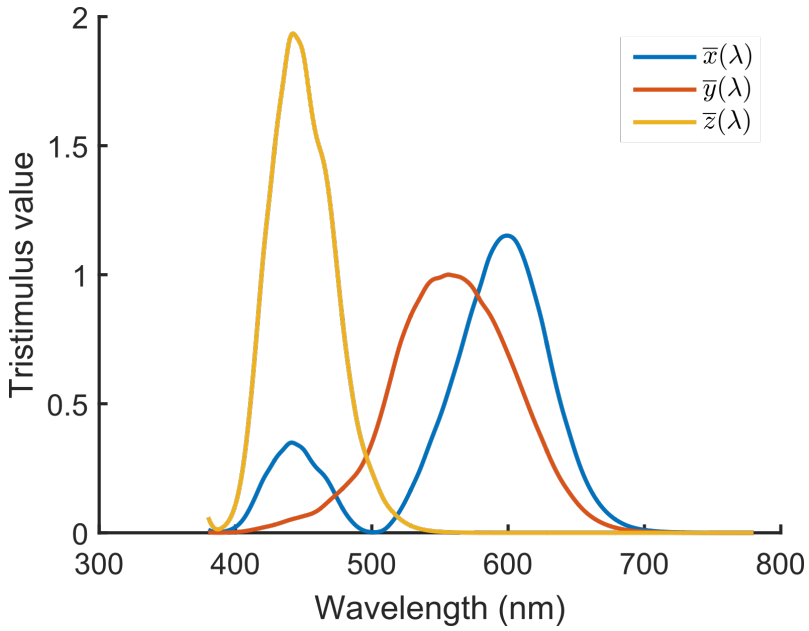


Figure 2.1: The CIE 2006 colour-matching functions. (Stockman & Sharpe (2000); CVRL.org)

2.1.3 Chromaticity

The CMF $\bar{y}$ is identical to the luminosity function, and hence, Y represents luminance. Luminance can be considered separately to chromaticity, a measure of 'colour'. A transformation of the $[X, Y, Z]$ tristimulus values takes us to the CIE xy chromaticity plane

at some level of luminance Y. We call these the $[Y, x, y]$ values. They are found using the following formula:

$$x = \frac{X}{X+Y+Z}, \tag{2.6}$$

$$y = \frac{Y}{X+Y+Z}. \tag{2.7}$$

2.1.4 A cone-opponent colour space

The $CIE\ Yxy$ colour space is useful for characterising illuminations and surfaces colorimetrically, but sometimes we may wish to represent them in terms of the excitation or response of different visual mechanisms. We have already seen that the first stage of visual processing (isomerisations of the cones in the retina) can be represented by a set of tristimulus values, $[L, M, S]$, calculated using the spectral sensitivities of the three cone types. Similarly, methods have been developed to calculate the responses of the three second-order mechanisms postulated by (Derrington *et al.*, 1984). The same authors defined one of the most widley used such colour spaces that has been coined the DKL colour space after the authors. Brainard (1996) provides a detailed development of the space as well as MATLAB code to compute values in the space that we utilise for colour conversions during this thesis. Briefly, suppose the current adaptation state (or background) has tristimulus values $[L_b, M_b, S_b]$. If the tristimulus values of the stimulus of interest are $[L_s, M_s, S_s]$ then the differential tristimulus values can be defined as $[\Delta L, \Delta M, \Delta S]$ where:

$$\Delta L = L_s - L_b, \tag{2.8}$$

$$\Delta M = M_s - M_b, \tag{2.9}$$

$$\Delta S = S_s - S_b. \tag{2.10}$$

The responses of the three second-order mechanisms, $[R_{Lum}, R_{L-M}, R_S]$ are defined as:

$$\begin{bmatrix} \frac{R_{Lum}}{k_{Lum}} \\ \frac{R_{L-M}}{k_{L-M}} \\ \frac{R_{S-(L+M)}}{k_{S-(L+M)}} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & \frac{-L_b}{M_b} & 0 \\ -1 & -1 & \frac{L_b+M_b}{S_b} \end{bmatrix} \begin{bmatrix} \Delta L \\ \Delta M \\ \Delta S \end{bmatrix} \tag{2.11}$$

The k_i are free constants and we follow the procedure for defining them suggested by Brainard (1996), choosing the k_i so that each mechanism has a response of 1 when stimulated in isolation by a stimulus with a pooled cone contrast of 1 ($(\Delta L/L_b)^2 + (\Delta M/M_b)^2 + (\Delta S/S_b)^2)^{1/2} = 1)$).

2.1.5 Perceptually uniform colour spaces

Another desirable property of a colour space is perceptual uniformity. MacAdam (1942) showed that tristimulus values that are equidistant in the $CIE\ Yxy$ colour space are not necessarily perceived to have equal colour difference; the space is not perceptually uniform. Further colour spaces have been developed that utilise behavioural data of colour discrimination thresholds, colour appearance data, and suprathreshold data to create colour spaces that are perceptually uniform. To do so, they transform the co-ordinates obtained from the CMFs ($[X, Y, Z]$).

One such example is the CIELAB colour space. CIELAB, like other colour spaces, is a three co-ordinate system. L^* represents the lightness dimension (higher values are brighter), a^* the red-green dimension (negative values are green, positive are red), and b^* the blue-yellow dimension (negative values are blue, positive are yellow). All points in CIELAB are defined relative to a white point. The transformation to CIELAB is defined

as follows:

$$L^* = 166f\left(\frac{Y}{Y_n}\right) - 16 \tag{2.12}$$

$$a^* = 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \tag{2.13}$$

$$b^* = 500\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \tag{2.14}$$

where

$$f(t) = \begin{cases} \sqrt[3]{t}, & t > \delta^3 \\ \frac{t}{3\delta^2} + \frac{4}{29}, & otherwise \end{cases}$$

with $\delta = 6/29$, and $[X_n, Y_n, Z_n]$ are the tristimulus values of the white point.

A second example is the CIELUV colour space. CIELUV has a similar three dimensions to CIELAB (a lightness (L^*), red-green (u^*), and blue-yellow (v^*) dimension) and is also defined relative to a white point. The transformation to CIELUV is defined as follows:

$$L^* = \begin{cases} \left(\frac{29}{3}\right)^3 (Y/Y_n), & (Y/Y_n) \leq \left(\frac{6}{29}\right)^3 \\ 166(Y/Y_n)^{1/3} - 16, & (Y/Y_n) > \left(\frac{6}{29}\right)^3 \end{cases} \tag{2.15}$$

$$u^* = 13L^*(u' - u'_n) \tag{2.16}$$

$$v^* = 13L^*(v' - v'_n) \tag{2.17}$$

where

$$u' = \frac{4X}{X + 15Y + 3Z} \tag{2.18}$$

$$v' = \frac{9Y}{X + 15Y + 3Z} \tag{2.19}$$

and $[X_n, Y_n, Z_n]$ are the tristimulus values of the white point.

Once in a perceptually uniform colour space, colour differences can be represented by a colour difference metric, defined as the Euclidean distance between the points. For

example, the distance between two points in CIELUV, $[L_1^*, u_1^*, v_1^*]$ and $[L_2^*, u_2^*, v_2^*]$ is:

$$\Delta E_{u^*v^*} = ((L_2^* - L_1^*)^2 + (u_2^* - u_1^*)^2 + (v_2^* - v_1^*)^2)^{1/2}. \tag{2.20}$$

A similar equation defines differences in CIELAB.

2.2 Testing for Colour Vision Deficiencies

In the studies presented in this thesis, participant's were screened for colour vision deficiencies using the 24-plate Ishihara Test and the Farnsworth-Munsell 100 Hue Test.

The Ishihara Test screens participants for red-green colour vision deficiencies. During the test, participants view a series of pseudo-isochromatic plates that contain a set of dots that have various sizes and colours. Hidden within the dots is a number that can only be seen by observers with normal colour vision as distinguishing the dots that make up the number from those that do not requires a certain level of colour discrimination. Participants read each number aloud as they go through the plates and errors are recorded. Specific combinations of errors (errors on a certain collection of plates) are indicative of different types of colour blindness and participants who display such a combination of errors were excluded from experiments.

In the Farnsworth-Munsell 100 Hue Test participants are required to sort coloured caps according to a hue gradient. The test consists of 85 coloured caps and participants sort these in subsets of 21/22 caps per tray. For each tray, the experimenter mixes up the caps, leaving the two end points fixed in the tray. The participant is then required to place the caps back in the tray to create a hue gradient between the two end points. After completion of the task, the ordering of the caps is used to assign participants a score at each of the 85 equally spaced points around the hue circle. If scores spike at two opposing points around the hue circle this can be indicative of a colour confusion line for the participants, suggesting a specific type of colour blindness. Any such participants were excluded from the experiments.

2.3 The Light Sources

In all experiments, lighting is provided by spectrally tunable LED light modules. These light modules are a form of solid state lighting (SSL). Prior to the introduction of SSL, the SPD of a light source was fixed, being pre-determined by the physical properties of the materials in the light source. Using a computer, the SPD of the light emitted by these LED light modules can be controlled in real-time via either a Bluetooth, USB or local wireless network (WiFi) connection. Each LED light module is multi-channel, meaning it contains a number of unique LEDs. The SPDs of each LED constitute a set of basis functions for all possible spectra that can be emitted by the light source. The spectrum emitted by the light modules is controlled by specifying the power of each individual LED channel (or basis function), which amounts to sending a list of weights to the light module consisting of a list of numbers bounded by zero and one. If there were n LED channels in the light module then a list of n weights must be sent to the light source, one for each channel, or one for each basis function. Hence, all SPDs that can be emitted by the light modules consist of all possible linear combinations of the basis functions. The methods for generating a set of weights to send to the light sources such that the SPD of the emitted light has particular specified characteristics will be discussed in Section 2.4. Depending on the experiment, different tunable LED prototypes were used that have a variable number of channels and communication method. The details of the two prototypes will be covered in the following subsections.

2.3.1 Mark I LEDMOTIVE Luminaires

The first prototype LED light module is the Mark I LEDMOTIVE luminaire (Figure 2.2.A) The Mark I Luminaire has 13 unique channels. The basis functions can be seen in Figure 2.3. Communication to the Mark I luminaire can be mediated by either a Bluetooth or a USB connection.

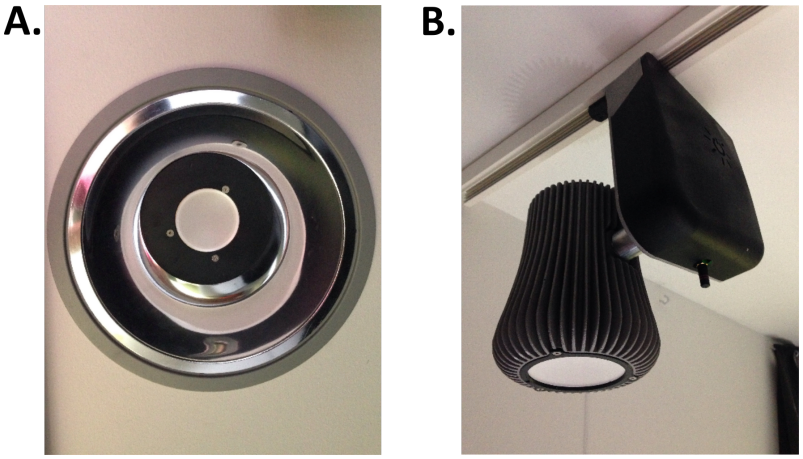


Figure 2.2: The luminaires. A. Mark I LEDMOTIVE luminaire. B. Hi-LED prototype I luminaire.

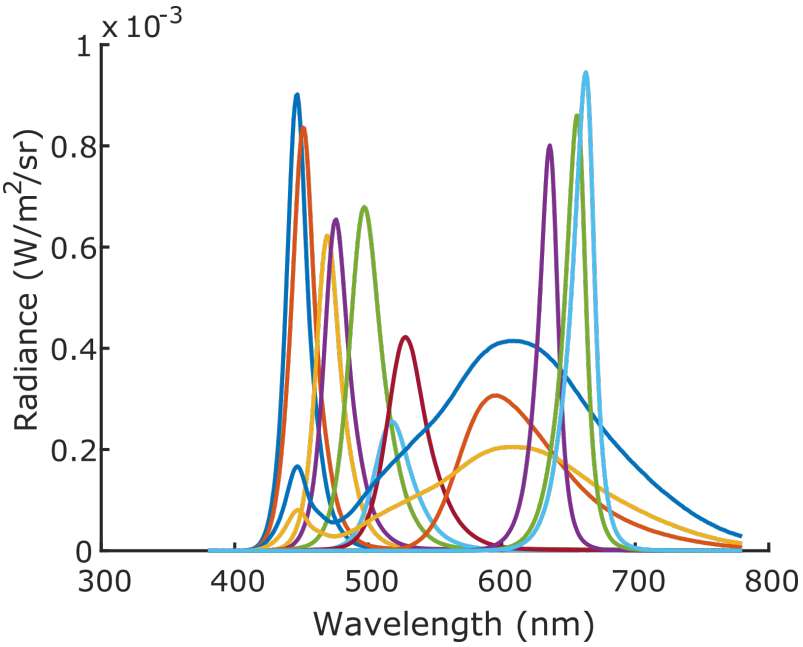


Figure 2.3: Basis functions for the Mark I LEDMOTIVE luminaire. The Mark I luminaire has 13 unique channels.

2.3.2 Hi-LED prototype I Luminaires

The second prototype LED light module is the Hi-LED prototype I luminaire (Figure 2.2.B). The HI-LED prototype I luminaire has 10 unique channels, two of which are duplicated. The basis functions can be seen in Figure 2.4. Communication to the HI-LED prototype I luminaire can be mediated by either a USB or WiFi connection.

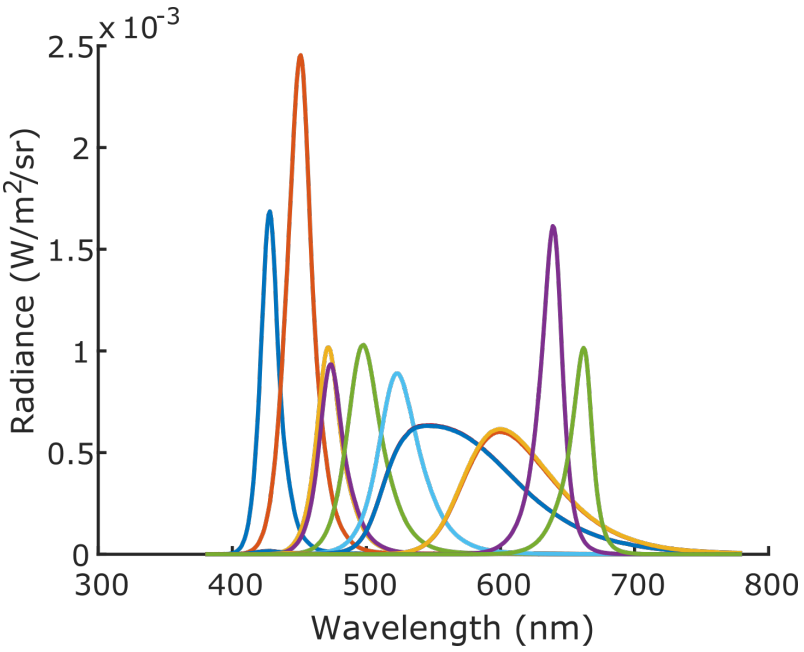


Figure 2.4: Basis functions for the Hi-LED prototype I luminaire. The HI-LED prototype I luminaire has 10 unique channels, with two duplicates.

2.4 Spectral Fitting Procedures

Fitting a SPD amounts to finding an array of weights that can be sent to the light modules that results in a light of given $[X, Y, Z]$ tristimulus values. When we do this, we always require the smoothest SPD possible as we want to use lighting conditions that are as natural as can be (lights in nature have relatively flat SPDs). The smoothness of a SPD $S(\lambda)$ can be defined as

$$\int_{\lambda} \left(\frac{dS}{d\lambda} \right)^2 d\lambda. \tag{2.21}$$

To see this, consider what $\frac{dS}{d\lambda}$ represents; it is the gradient of the $S(\lambda)$ at any given wavelength λ. Thus, $|\frac{dS}{d\lambda}|$ is a measure of the variation of $S(\lambda)$ at any given wavelength λ. To find the total variation in $S(\lambda)$ we can sum the variation at every wavelength; however, as there are infinitely many wavelengths at which we could measure the variation of $S(\lambda)$ this sum becomes an integral. Hence, the total variation (TV) of $S(\lambda)$ can be defined as

$$TV = \int_{\lambda} \left| \frac{dS}{d\lambda} \right|^p d\lambda, \tag{2.22}$$

where p is an optional constant that can be used to assign different weights to smaller or larger values of $|\frac{dS}{d\lambda}|$. We wish to highlight areas of large variability, so we choose a positive integer p such that we assign larger weights to larger values of $|\frac{dS}{d\lambda}|$ and smaller weights to smaller values of $|\frac{dS}{d\lambda}|$ (as opposed to choosing a negative integer p that would do the opposite). For our purposes, we set $p = 2$.

For a set of SPDs, $S_i(\lambda)$, of a set of metamers, the smoothest SPD will be given by

$$\min_i \int_\lambda \left(\frac{dS_i(\lambda)}{d\lambda}\right)^2 d\lambda. \tag{2.23}$$

Now, following Li & Luo (2001) we have

$$\int_\lambda \left(\frac{dS}{d\lambda}\right)^2 d\lambda = \frac{\Delta\lambda}{2}\left[\left(\frac{dS(\lambda_1)}{d\lambda}\right)^2 + \left(\frac{dS(\lambda_N)}{d\lambda}\right)^2 + 2\sum_{k=2}^{N-1}\left(\frac{dS(\lambda_k)}{d\lambda}\right)^2\right] \tag{2.24}$$

by the trapezium rule. Using differentiation from first principles for a discrete set of points we have

$$\frac{dS(\lambda_k)}{d\lambda} = \frac{S(\lambda_{k+1}) - S(\lambda_k)}{\Delta\lambda}, \quad \forall k \in [1, N-1] \tag{2.25}$$

and

$$\frac{dS(\lambda_k)}{d\lambda} = \frac{S(\lambda_k) - S(\lambda_{k-1})}{\Delta\lambda}, \quad \text{for } k = N. \tag{2.26}$$

Equations 2.25 and 2.26 can be substituted in to equation 2.24 to give

$$\begin{aligned}\int_\lambda \left(\frac{dS}{d\lambda}\right)^2 d\lambda = \frac{\Delta\lambda}{2}\Bigg[\left(\frac{S(\lambda_2) - S(\lambda_1)}{\Delta\lambda}\right)^2 + \left(\frac{S(\lambda_N) - S(\lambda_{N-1})}{\Delta\lambda}\right)^2 + \dots \\ \dots + 2\sum_{k=2}^{N-1}\left(\frac{S(\lambda_{k+1}) - S(\lambda_k)}{\Delta\lambda}\right)^2\Bigg].\end{aligned} \tag{2.27}$$

Pulling out the factor of $2\left(\frac{1}{\Delta\lambda}\right)^2$ gives

$$\int_\lambda \left(\frac{dS}{d\lambda}\right)^2 d\lambda = \frac{1}{\Delta\lambda}\left[\frac{1}{2}\left(S(\lambda_2)-S(\lambda_1)\right)^2 + \frac{1}{2}\left(S(\lambda_N)-S(\lambda_{N-1})\right)^2 + \dots \right. \\ \left. \dots + \sum_{k=2}^{N-1}\left(S(\lambda_{k+1})-S(\lambda_k)\right)^2\right]. \tag{2.28}$$

This becomes

$$\int_\lambda \left(\frac{dS}{d\lambda}\right)^2 d\lambda = \frac{1}{\Delta\lambda}\left[\left(\frac{S(\lambda_2)-S(\lambda_1)}{\sqrt{2}}\right)^2 + \left(\frac{S(\lambda_N)-S(\lambda_{N-1}}{\sqrt{2}}\right)^2 + \dots \right. \\ \left. \dots + \sum_{k=2}^{N-1}\left(S(\lambda_{k+1})-S(\lambda_k)\right)^2\right]. \tag{2.29}$$

If we set

$$\underline{r} = \begin{pmatrix} \frac{S(\lambda_2)-S(\lambda_1)}{\sqrt{2}} \\ S(\lambda_3)-S(\lambda_2) \\ \vdots \\ S(\lambda_{N-1})-S(\lambda_{N-2}) \\ \frac{S(\lambda_N)-S(\lambda_{N-1})}{\sqrt{2}} \end{pmatrix}, \tag{2.30}$$

then

$$\int_\lambda \left(\frac{dS}{d\lambda}\right)^2 d\lambda = \frac{1}{\Delta\lambda}||\underline{r}||^2, \tag{2.31}$$

where we consider $||\underline{x}||$ to be the usual Euclidean norm, i.e. $||\underline{x}|| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$.

If we take the matrix

$$\boldsymbol{D} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \cdots & \cdots & -1 & 1 & 0 \\ 0 & \cdots & \cdots & \cdots & -1 & 1 \\ 0 & \cdots & \cdots & \cdots & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}, \tag{2.32}$$

then we can write $\underline{r} = \boldsymbol{D}\underline{s}$ where

$$\underline{s} = \begin{pmatrix} S(\lambda_1) \\ S(\lambda_2) \\ \vdots \\ S(\lambda_{N-1}) \\ S(\lambda_N) \end{pmatrix}. \tag{2.33}$$

Hence, equation 2.23 becomes

$$\min_{\underline{s}} \frac{1}{\Delta\lambda} ||\boldsymbol{D}\underline{s}||^2, \tag{2.34}$$

and this is the equation we now wish to solve.

For our light modules, the basis functions for any illuminant are known (i.e. we know the SPDs of the n LED channels). These SPDs can be represented in an $(N \times n)$-matrix $\boldsymbol{A}$. So, any illuminant $\underline{s}$ can be expressed as $\underline{s} = \boldsymbol{A}\underline{w}$, for a vector of weights $\underline{w}$. We can now reformulate equation 2.34 further to say that the smoothest SPD will be found by solving

$$\min_{\underline{w}} ||\boldsymbol{D}\boldsymbol{A}\underline{w}||^2, \tag{2.35}$$

where we omit $\dfrac{1}{\Delta\lambda}$ as it is constant and will not affect the minimum.

By definition of matrix multiplication

$$\begin{aligned} ||\boldsymbol{D}\boldsymbol{A}\underline{w}||^2 &= (\boldsymbol{D}\boldsymbol{A}\underline{w})^T \boldsymbol{D}\boldsymbol{A}\underline{w}, & (2.36) \\ &= \underline{w}^T \boldsymbol{A}^T \boldsymbol{D}^T \boldsymbol{D}\boldsymbol{A}\underline{w}, & (2.37) \end{aligned}$$

so our problem is to find

$$\min_{\underline{w}} (\underline{w}^T \boldsymbol{A}^T \boldsymbol{D}^T \boldsymbol{D}\boldsymbol{A}\underline{w}). \tag{2.38}$$

If we require a smooth SPD with $[X, Y, Z] = [X_1, Y_1, Z_1]$ then we wish to solve equation

2.38 subject to

$$6.83 \times \boldsymbol{R}^T \boldsymbol{A} \underline{w} = \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix}, \tag{2.39}$$

where $\boldsymbol{R}$ is a matrix representing the CMFs; each column of $\boldsymbol{R}$ is a CMF..

This is now a quadratic programming problem and can be solved in MATLAB using the *quadprog* command.

Chapter 3

Illumination discrimination under chromatically biased illuminations

3.1 Introduction

Pearce *et al.* (2014) found a "blue bias" in illumination discrimination thresholds using the illumination discrimination task. Average thresholds were highest when the change in the illumination was along a bluer direction of chromatic change. This was interpreted as evidence for better colour constancy under blueish illuminations (observers were more tolerant to bluer illumination changes). However, as we mentioned, it is still unclear how this bias is encoded in the visual system. The experiment we present here allows us to further address this question.

The results of Pearce *et al.* (2014) were replicated by Radonjić *et al.* (2016*b*) in both real and simulated scenes. Radonjić *et al.* (2016*b*) also found that relative thresholds for the different directions of illumination change depend on the average chromaticity of surfaces in the scene (where by surface chromaticity we mean the chromaticity that the surface would have under a hypothetical equal energy white light). In particular, when the chromaticities of surfaces in the scene are biased towards reddish-blue, they find increased sensitivity to illumination changes when the illumination becomes more red, and decreased sensitivity when it becomes more green (relative to a neutral average surface chromaticity condition). A similar trend appears when the surfaces in the scene are biased such that their mean chromaticity is yellowish-green; sensitivity is decreased for bluer and redder illumination changes, but increased for greener illumination changes. In summary, thresholds tend to increase in directions chromatically opponent to the bias and decrease in the direction of bias.

Radonjić *et al.* (2016*b*) conclude that sensitivity to different directions of chromatic change in the illumination must be defined with respect to the surfaces in the scene. They offer an

explanation of these results by suggesting that as the bias in surface chromaticities leads to an increase in surfaces reflecting certain wavelengths of light (i.e. longer wavelengths are reflected more in the reddish-blue condition), there will be an increase in sensitivity to changes in illumination that contain more light at those wavelengths.

With this study, we address two questions. Firstly, if the explanation offered by Radonjić *et al.* (2016*b*) is correct, then we may expect that the same mean scene chromaticity bias caused by biasing the chromaticities of the illuminations in the task, while maintaining a surface ensemble with neutral average chromaticity, will not have the same effect on relative thresholds. In other words, we ask if sensitivity to different directions of chromatic change in the illumination must be defined with respect to the mean illumination chromaticity (temporally over the course of the experiment) in addition to the surface ensemble. Secondly, if relative thresholds do change with the reference illumination, do they change only when the reference illumination is moved away from the Planckian locus but not otherwise? Put differently, we first ask if the "blue bias" is specific to bluer illumination changes that mimic daylight changes. If such a distinction can be made it may be suggestive of the underlying visual mechanisms that mediate the "blue bias". Specifically, the behavioural data may be able to distinguish between our two competing hypotheses; the nature vs. nurture hypotheses (see Section 1.5.3). In the case of the nurture explanation, we may expect that observers will be more colour constant for illuminations that are closer in chromaticity to the daylight locus. Alternatively, the nature explanation would predict a reduced sensitivity to bluer illumination changes in general, not just ones that fall along the daylight locus.

To answer these questions, we examined performance of a group of observers who completed five versions of the IDT. One version was the standard IDT used by Pearce *et al.* (2014) and Radonjić *et al.* (2016*b*). In the other four versions of the task, the reference light became the most extreme test light (50 $\Delta E_{u^*v^*}$ away) from one of the four axes of change used in the original experiment. The corresponding test lights were also shifted in chromaticity space such that all illuminations used within a given condition were all either

blue, yellow, red or green in appearance. Our results show that the reference illumination, and hence the mean chromaticity of the stimulus during an experimental condition, modulates relative thresholds for the different directions of chromatic change. This is also true for reference illuminations that fall on the Planckian locus. However, we still see a "blue bias" in all reference illumination conditions. We show that the biases we see are not explained by a bias in low-level scene statistics.

3.2 Methods

3.2.1 Overview

Participants viewed a Mondrian-papered scene illuminated by LED lamps. After viewing the scene under a reference illumination, participants were required to indicate which of two successively presented comparison illuminations most closely matched the reference. One of the comparison illuminations would always match the reference (the target). The other comparison illumination (the test) varied from the reference such that it nominally became either bluer, yellower, redder or greener in colour. In a 5×4 repeated measures design, all participants ($n = 9$) completed the task for five different reference illumination conditions (neutral, blue, yellow, red and green). A staircase procedure was used to find thresholds for discrimination of an illumination change along each of the four axes of change for each reference condition.

3.2.2 Ethics

Ethical approval for this and all studies discussed later in the thesis was received from the Newcastle University Ethics Board. Written consent was received from all participants prior to participation in the studies.

3.2.3 Participants

Nine participants were recruited (4 male, 5 female, mean age of 23.4 ± 2.88 years). All participants had normal or corrected to normal visual acuity and no colour vision deficiencies, assessed using Ishihara Colour Plates.

3.2.4 The Scene

The scene was viewed through a porthole in a box of dimensions: height 45 cm, width 77.5 cm, depth 64.5 cm. The front of the box extended out towards the participant so that when looking through the porthole the viewing distance from the viewing plane to the back of the box was 81 cm (Figure 3.1.A). Both the scene and participants were immersed in the illumination (the only source of light in the room); their view, however, was restricted to inside the box (Figure 3.1.B). The top of the box was open to allow illumination of the scene. The bottom, rear and sides of the box were papered with a matte printed Mondrian, designed such that its average reflectance was flat (Figure 3.2.B). Each patch of the Mondrian was one of 24 unique chromaticities, chosen such that their average chromaticity was neutral (mean $CIE\ xy$ chromaticity $(x, y) \approx (0.33, 0.33)$ under a hypothetical equal energy white). The height and width of each patch varied in size from 2 mm to 42 mm (0.14 to 2.97 degrees of visual angle). The Mondrian was created by generating $100,000$ patches of varying pixel size and chromaticity and randomly placing them inside a 1001×1001 pixel frame that was initially set to $(R, G, B) = (0, 0, 0)$ everywhere. After the Mondrian generation process was complete, we checked for RGB values equal to $(0, 0, 0)$ to ensure that no black patches remained.

The spectral reflectance distribution of each unique patch used to make up the Mondrian lining was found by first taking a radiance measurement from one sample of each unique patch with a CS2000 Konica Minolta spectroradiometer (Konica Minolta, Nieuwegein, Netherlands). The location of each patch is shown in Figure 3.2.A. These measurements were then subject to pointwise division by a measurement of the incident illumination

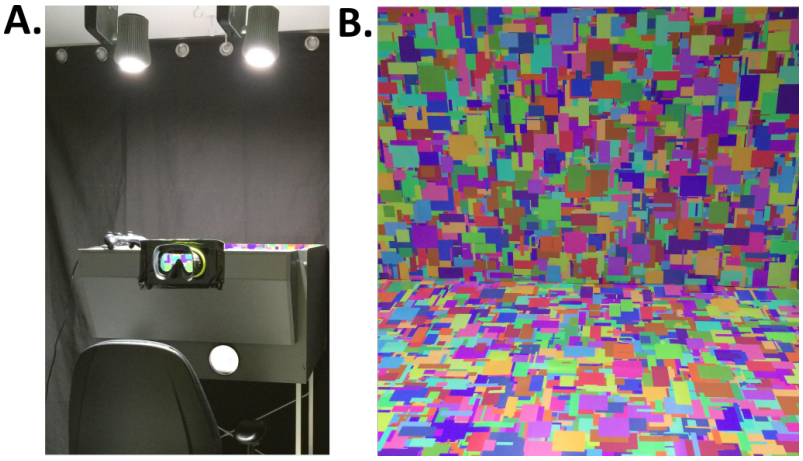


Figure 3.1: The experimental chamber. A. Participants positioned their head against the goggles, restricting their view to inside the stimulus box. The box was illuminated by two spectrally tunable LED luminaires. B. The participant's view inside the box (shown here illuminated by an arbitrary illumination that was not used during the experiment).

spectrum taken from a calibration tile to find the spectral reflectance distributions (Figure 3.2.B). We then calculated the chromaticity that each patch would have under a hypothetical equal energy white light using the CIE 2006 Yxy colour matching functions (Table 3.1).

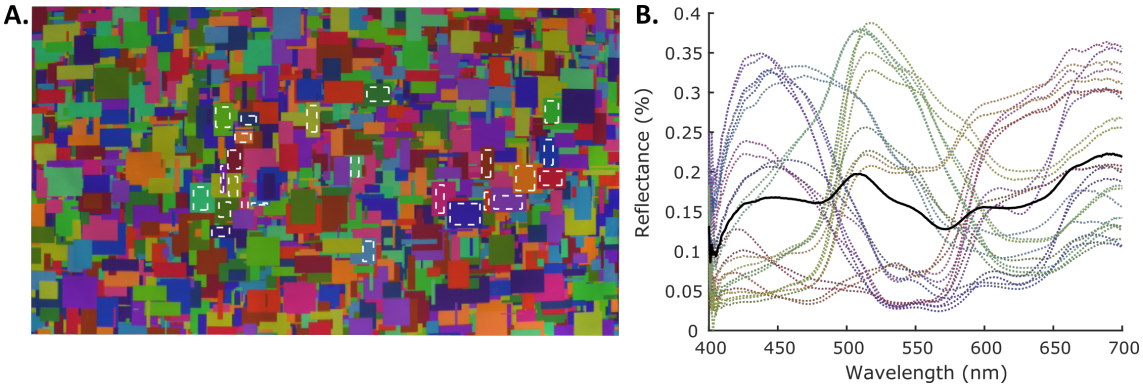


Figure 3.2: Parameters of the Mondrian. A. The locations of the 24 measured patches on the back wall of the stimulus box. B. The surface reflectance of each of the 24 unique patches used to make up the Mondrian lining of the box. Black solid line is the average reflectance of these 24 patches.

3.2.5 The Illuminations

The illuminations used in the experiment can be split into five subgroups corresponding to the five different conditions; a neutral, blue, yellow, red and green set (Figure 3.3). In

Table 3.1: *CIE xy* chromaticities of each unique patch in the Mondrian under a hypothetical equal energy white light. Also shown is the mean of these patches (M_{24}) and the mean taken across the whole hyperspectral image (M_h).

Patch	*CIE x*	*CIE y*	Patch	*CIE x*	*CIE y*
1	0.35	0.49	13	0.26	0.39
2	0.37	0.26	14	0.39	0.49
3	0.24	0.28	15	0.27	0.17
4	0.21	0.14	16	0.42	0.28
5	0.37	0.23	17	0.42	0.40
6	0.28	0.21	18	0.23	0.23
7	0.49	0.37	19	0.34	0.52
8	0.46	0.38	20	0.25	0.18
9	0.45	0.44	21	0.25	0.28
10	0.46	0.29	22	0.30	0.44
11	0.20	0.18	23	0.37	0.45
12	0.30	0.48	24	0.28	0.39
M_{24}	0.33	0.33	M_h	0.34	0.33

each set, the illuminations were generated such that they varied systematically away from a central point in the *CIE xy* chromaticity diagram. These central points were the reference illuminations for each experimental condition (neutral, blue, yellow, red and green). The *CIE xy* chromaticities of the five reference lights were: $(x, y) = (0.31, 0.33)$ (D65), $(x, y) = (0.25, 0.26)$, $(x, y) = (0.39, 0.39)$, $(x, y) = (0.32, 0.26)$ and $(x, y) = (0.30, 0.38)$, respectively (Figure 3.3.A). For each reference illumination, 20 test illuminations were generated that varied away from the reference in each of four distinct chromatic directions: nominally bluer, yellower, redder and greener. For neutral, blue and yellow reference illuminations, bluer/yellower test illuminations were parametrised to fall along the Planckian locus in order to mimic the chromaticities of daylight illuminations (defined in the *CIE xy* chromaticity plane as $y = 2.870x - 3x^2 - 0.275$; Wyszecki & Stiles (1967)); whereas for the red and green reference illuminations, the bluer/yellower test illuminations varied along a linear transformation of the Planckian locus. For neutral, blue and yellow reference illuminations, redder/greener test illuminations were parametrised to fall along the CCT line to the Planckian locus at the location of the reference illumination. Finally, in the red and green reference illumination conditions, the redder/greener test illuminations fell along a linear translation of the CCT line to the Planckian locus at $x = 0.31$ (D65),

maintaining orthogonality with the bluer/yellower changes in the $CIE\ u'v'$ chromaticity plane. Each set of 20 test illuminations were spaced one $\Delta E_{u^*v^*}$ apart (using the neutral reference illumination as the white point) (Figure 3.3.B). All illuminations were generated such that the luminance of a white polymer calibration tile was $Y = 50\ cd/m^2$ when illuminated by these lights.

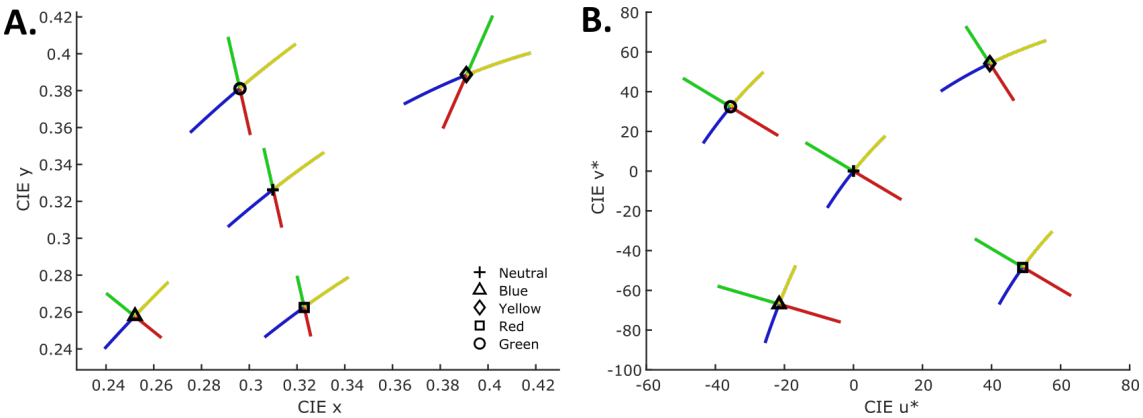


Figure 3.3: The IDT illumination chromaticities. A. Plotted in the $CIE\ xy$ chromaticity plane. B. Plotted in the $CIE\ u^*v^*$ chromaticity plane. The black open symbol marks the chromaticity of the reference illumination in each of the five conditions. The different axes of chromatic change are shown in their respective colours for each reference illumination condition.

We generated a further 30 illuminations along each axis that were 21 to 50 $\Delta E_{u^*v^*}$ away from the reference. While these illuminations were not used as part of the staircase procedure, one of them appeared every 10 trials from a randomly decided chromatic direction of change. We know from previous studies that these stimuli are supra-threshold for trichromat observers. The purpose of including such trials in the experiment was to ensure that the participant remained engaged with the task even when the staircases were approaching threshold and they were unlikely to detect the target. In addition, these trials can be used to ensure that our participants were attentive and performing the task to the best of their ability. If participants respond incorrectly to a large percentage of these trials (i.e. they display a high lapse rate), we can use this information to exclude participants from the analysis. For this reason, we will refer to these trials as catch trials. It was not necessary to use these trials in this particular experiment. We will meet them again in Chapter 4 however when they will prove useful for a different purpose.

The illuminations were produced using two 10-channel (9 unique) spectrally tunable LED lamps (see Section 2.3.2). The spectral power distribution of the light emitted from the lamps can be controlled in real-time by controlling the pulse width modulation of each individual LED channel, allowing the spectral power distribution to be specified by a set of 10 weights. In order to find a set of weights that produce a spectral power distribution with a certain luminance and $CIE\ xy$ chromaticity (as measured from the white calibration tile), we used custom MATLAB scripts to find the smoothest possible fit to the specified illumination characteristics. The smoothness constraint ensures that the spectra are comparable in smoothness to natural daylights. Full details of the spectral fitting procedure are reported elsewhere (Pearce *et al.*, 2014; Radonjić *et al.*, 2016*b*; Finlayson *et al.*, 2014) and in Chapter 2.

3.2.6 Spectral Calibration

Calibration was performed in two stages. Initially, we measured the spectral power distribution of each individual LED channel at maximum power. These will be referred to as the basis functions of the lamps. To measure the basis functions, a polymer white reflectance tile was placed flush against the back wall of the stimulus box with the Mondrian lining removed. A CS2000 Konica Minolta spectroradiometer (Konica Minolta, Nieuwegein, Netherlands) was used to take radiance measurements from the tile when illuminated separately by each LED channel. These basis functions were input to the spectral fitting code to find the weight vectors that would give illuminations of the specified chromaticities and luminance. The spectral power distributions of the fitted weights were then measured using the same procedure used to obtain the basis functions (Figure 3.4).

3.2.7 Hyperspectral Imaging and Illumination Modelling

Hyperspectral imaging with a Specim V10E camera was used to obtain image data for the analysis of scene statistics. A surface reflectance image of the back wall was obtained by first imaging the back wall under an arbitrary white light before removing the Mondrian

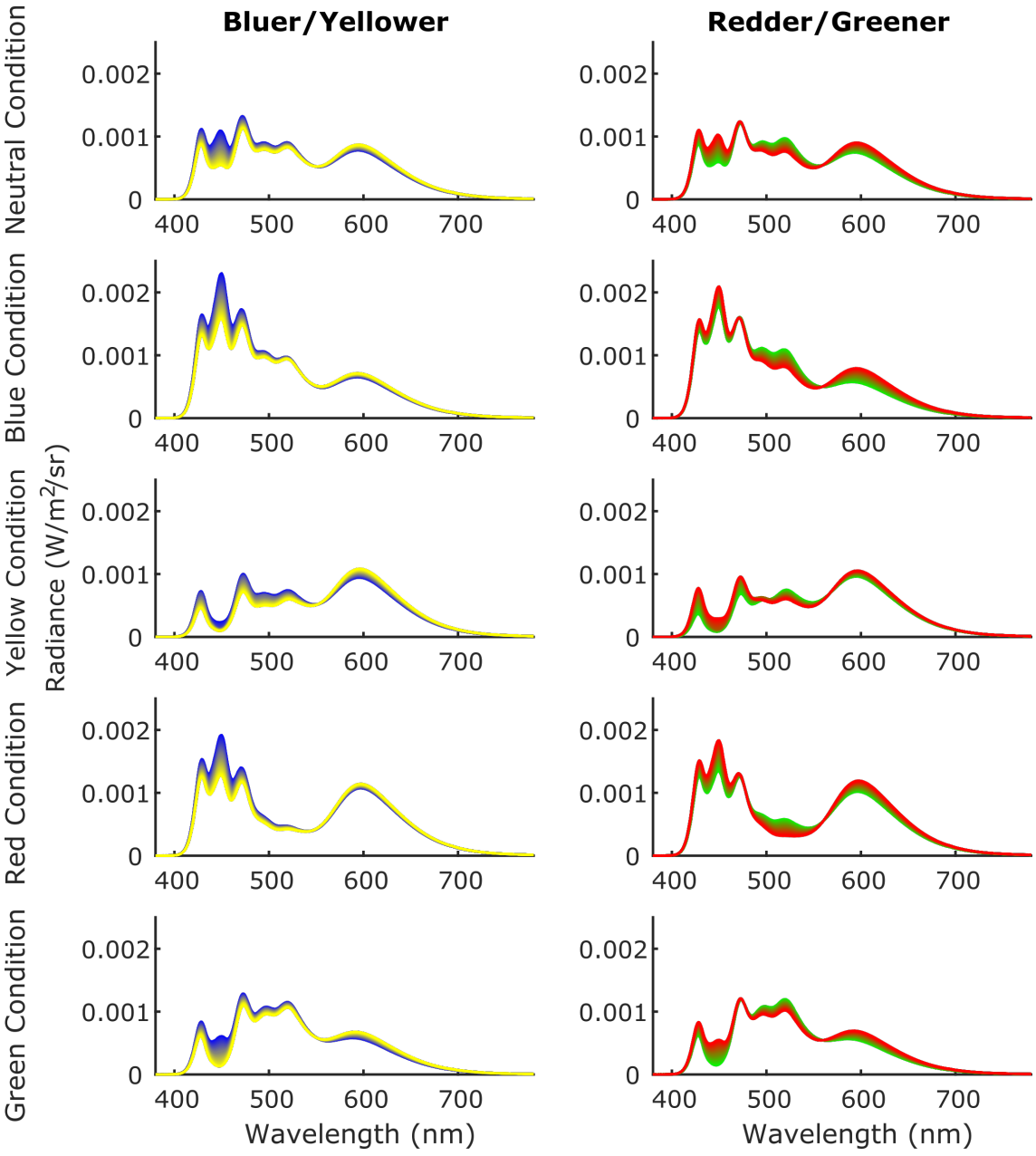


Figure 3.4: The spectral power distributions of the IDT illuminations. Each pair of plots across a row show bluer/yellower and redder/greener test illuminations for a given reference illumination condition.

lining from the box and then obtaining an image of the white reflectance tile covering the back wall. Because the tile is smaller than the wall, the latter image was obtained by combining three images, in which the white tile was placed either flush to the right wall of the box, in the centre of the box or flush to the left wall of the box. Thus, we obtained

a complete representation of the spatial gradients of irradiance on the back wall. Both hyperspectral images (the image of the Mondrian-papered back wall and the combined image of the white reflectance tile) were then cropped to remove any areas of the image that were above, below or to either side of the back wall of the box. The cropped hyperspectral image of the white reflectance tile was smoothed using a 2D Gaussian kernel. Finally, this smoothed image of the white reflectance tile was used to remove the illumination from the image of the Mondrian-papered back wall using point-wise division at each pixel (each pixel being a point in a data cube that represents a measured spectrum of light). To model the appearance of the Mondrian-papered back wall under each illumination, the measured SPDs of each illumination were combined with the measured surface reflectance at each pixel using point-wise multiplication. These images were used to find the mean scene chromaticity under each illumination used in the experiment in order to form the scene mean chromaticity look up table detailed in the data analysis section below.

3.2.8 Procedure

Participants visited the laboratory on five occasions, once for each reference illumination condition. On their first visit to the laboratory, all participants were tested for colour vision deficiencies using the Ishihara Colour Plates. In the experiment room, participants read the standardised instructions and were permitted to ask questions. All participants then received the same verbal instructions: *"On each trial, you will see the reference illumination followed by two comparison illuminations. You will use the gaming pad to indicate which of the two comparison illuminations most closely matched the reference"*. These instructions were repeated on subsequent visits but the participants did not read the standardised instructions a second time. On the first and all subsequent visits, before beginning the task, there was a two minute dark adaptation period. On each trial, the reference illumination was visible for 2000 ms. Each comparison illumination was displayed for 500 ms and between each illumination was 400 ms of darkness (Figure 3.5). The correct match randomly switched between the first and second comparison. Thresholds

for each direction of change were found using a 1-up, 3-down, transformed and weighted staircase procedure (Kaernbach, 1991). Three staircases were completed for each direction of change. Staircases terminated after six reversals (as this was an optimal amount of reversals to ensure the experiment was not too long and the staircases converged sufficiently according to simulations). There was no time limit on how long participants could wait to make their response and participants were told that they had the opportunity to take a break at any time during the experiment by remembering their response but not entering it until they were ready to continue. Participants were required to take a mandatory break after every 100 trials. This break could last for as long as the participant required; however, they did not leave the dark room during this period.

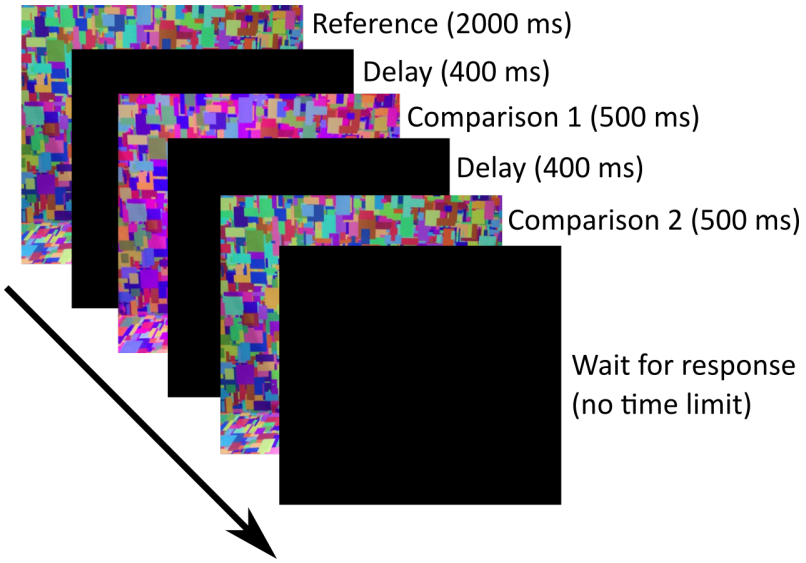


Figure 3.5: The illumination discrimination task (IDT).

3.2.9 Data Analysis

Thresholds were calculated by taking the mean of the last two reversals from each of the three staircases (a mean of six reversals in total). A look up table was created mapping each nominal staircase step (1 to 20) to a $\Delta E_{u^*v^*}$ between the measured chromaticities of each test light and the relevant reference (fixed white point illumination look up table). All reversal indices were first converted using the look up table and then an average taken

over the $\Delta E_{u^*v^*}$ values.

In addition, we also created a look up table mapping each nominal staircase step (1 to 20) to a $\Delta E_{u^*v^*}$ between the mean scene chromaticity under each test light and the relevant reference (scene mean look up table). The mean scene chromaticity was calculated from the surface spectral reflectance of the Mondrian-papered back wall (measured as above) and the measured illumination spectra, for each point in the scene. Converting nominal staircase steps to $\Delta E_{u^*v^*}$ values using this table accounts for any bias introduced by a bias in the scene statistics.

The neutral reference illumination (D65) was used as the white point for calculation of $\Delta E_{u^*v^*}$ in all reference illumination conditions (for both of the look up tables detailed above). However, it could be argued that if the participant adapts to the temporal average of the illumination during each condition of the experiment, then the reference illumination from each condition is a more appropriate choice for the respective white point that $\Delta E_{u^*v^*}$ values are defined with respect to. Hence, we made a third look up table in this way (variable white point illumination look up table). We show in the results section that our overall conclusions remain the same regardless of the look up table that we use.

All data are presented as mean and standard error unless stated otherwise. Graphed error bars also represent standard error. Where pairwise comparisons and simple main effects are reported, p-values have been corrected for multiple comparisons by applying a Bonferroni correction.

3.3 Results

3.3.1 Using the fixed white point illumination look up table

The data show that relative illumination discrimination thresholds for the different chromatic directions of illumination change depend on the reference illumination (Figure 3.6, Table 3.2). A qualitative look at these data reveals a pattern. While overall thresholds

are highest for the bluer direction of change regardless of reference illumination condition, thresholds for the direction of change that is chromatically opponent to the bias in the reference illumination (e.g. thresholds for the greener direction of change in the red reference condition) are increased relative to the other directions of chromatic change. In other words, whilst the "blue bias" prevails overall, a bias also emerges in the chromatic direction that is opponent to the reference chromaticity.

A 5×4 repeated measures ANOVA reveals a significant interaction effect of reference illumination and chromatic direction of illumination change on illumination discrimination thresholds ($F(12, 96) = 9.11, p < .001$). Moreover, there was a significant main effect of reference illumination, regardless of the direction of chromatic change ($F(4, 32) = 12.84, p < .001$), and a significant main effect of chromatic direction of illumination change, regardless of reference illumination condition ($F(3, 24) = 8.35, p = .001$). When thresholds are averaged over the different reference illumination conditions, thresholds for the bluer and yellower chromatic direction of change are significantly higher than redder thresholds ($p = .024$ and $p = .008$, respectively). There were no other significant differences between the different directions of chromatic change (Figure 3.6.A).

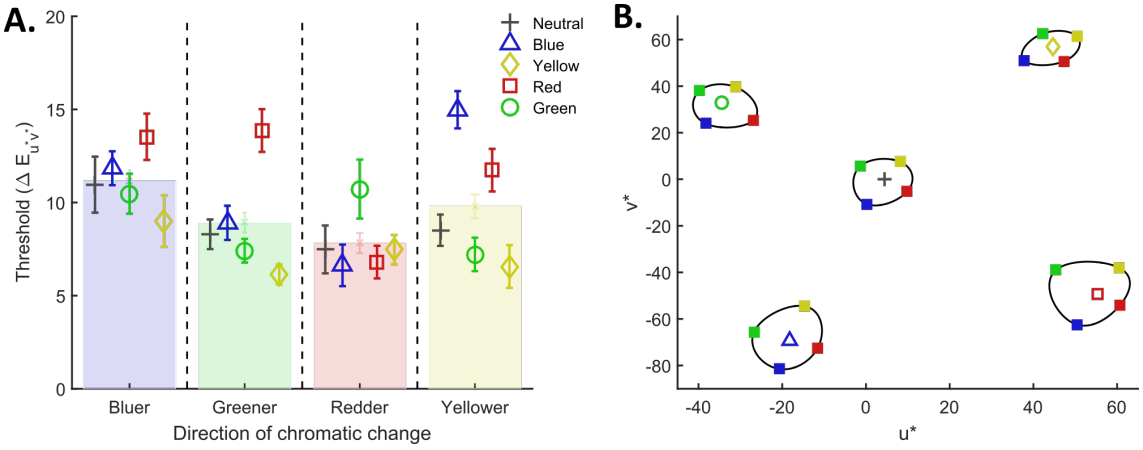


Figure 3.6: IDT thresholds calculated using the fixed white point look up table. A. Illumination discrimination thresholds for each chromatic direction of change and for each reference illumination condition. Transparent bars represent the main effect of chromatic direction of change. They are the thresholds for each chromatic direction averaged over reference illuminations. B. Mean thresholds plotted in the *CIE* u^*v^* chromaticity plane, with a discrimination contour formed by interpolating between the points, showing how the asymmetry in the discrimination ellipse changes as the reference moves around the colour space.

Table 3.2: Mean IDT thresholds using the fixed white point look up table. Values in brackets show the standard error.

Reference	Bluer	Greener	Redder	Yellower
Neutral	10.96 (1.42)	8.30 (0.75)	7.48 (1.21)	8.51 (0.79)
Blue	11.84 (0.86)	8.91 (0.87)	6.63 (1.05)	14.98 (0.94)
Green	10.47 (1.01)	7.41 (0.60)	10.73 (1.49)	7.21 (0.85)
Red	13.53 (1.17)	13.87 (1.08)	6.80 (0.83)	11.74 (1.08)
Yellow	9.00 (1.31)	6.14 (0.51)	7.47 (0.74)	6.57 (1.08)

As the interaction term was significant, we explored simple main effects. There was a simple main effect of chromatic direction of illumination change for the blue, green and red reference illumination conditions ($F(3, 24) = 15.45, p < .005$; $F(3, 24) = 5.37, p = .030$; and $F(3, 24) = 11.48, p < .005$, respectively), but not in the neutral and yellow reference illumination conditions ($F(3, 24) = 4.32, p = .070$ and $F(3, 24) = 2.88, p = .285$, respectively). In the blue reference illumination condition, thresholds differed between the following directions of change: bluer vs. redder ($p = .017$); greener vs. yellower ($p = .041$); and redder vs. yellower ($p = .004$). In summary, thresholds were highest for the yellower direction of change, but they were only significantly higher than the thresholds for the redder and greener directions, not bluer. For the red reference illumination condition, thresholds differed between the redder direction of chromatic change and all other directions of change ($p = .006$, $p = .012$, and $p = .008$, for blue, green, and yellow, respectively). Thresholds in the redder direction of change were significantly lower than thresholds in all other directions, with thresholds highest for the bluer and greener directions of change. No pairwise comparisons were significant for the green reference illumination condition, although the highest thresholds were for the bluer and redder directions of change.

There was a simple main effect of reference illumination on illumination discrimination thresholds in all chromatic directions ($F(4, 32) = 4.38, p = .024$ for bluer, $F(4, 32) = 14.15, p < .004$ for greener, and $F(4, 32) = 18.13, p < .004$ for yellower) except redder ($F(1.64, 13.11) = 4.32, p = .168$, with a Greenhouse-Geisser correction). Illumination discrimination thresholds for the bluer direction of change were lowest in the yellow reference illumination condition and highest in the red reference illumination condition, although no

pairwise comparisons among these thresholds were significant. For the greener direction of change, the red reference illumination produced a threshold significantly higher than all other reference illumination conditions except blue ($p = .039$, $p = .003$, and $p < .001$, for the neutral, green, and yellow reference illumination conditions, respectively). Finally, for the yellower direction of change, the highest threshold was in the blue reference illumination condition and this was significantly higher than in the neutral, green, and yellow reference illumination conditions ($p = .006$, $p = .002$ and $p = .005$, respectively). In addition, thresholds for the yellower direction of change in the red reference illumination condition were significantly higher than for the green and yellow reference illumination conditions ($p = .011$, and $p = .005$, respectively).

3.3.2 Using the scene mean look up table

If we repeat the ANOVA analysis above on the data obtained using the scene mean look up table a similar pattern of results emerge (Figure 3.7.A, Table 3.3). We can quantify the strength of the correspondence between the two threshold sets by calculating Pearson's correlation coefficient ($r = .979, p < .001$; Figure 3.7.B). We still find a significant interaction effect of reference illumination and chromatic direction of illumination change on illumination discrimination thresholds ($F(12, 96) = 7.58, p < .001$), a significant main effect of reference illumination ($F(4, 32) = 9.67, p < .001$), and a significant main effect of chromatic direction of illumination change ($F(3, 24) = 3.87, p = .022$). However, although thresholds in the bluer direction of change are still the highest when averaged over reference illumination conditions, no pairwise comparisons between the different directions are significant.

Using this look up table, there was only a simple main effect of chromatic direction of illumination change for the blue and red reference illumination conditions ($F(3, 24) = 7.27, p < .005$ and $F(3, 24) = 11.67, p < .005$, respectively), but not in the neutral, green, and yellow reference illumination conditions ($F(3, 24) = 2.32, p = .505$; $F(3, 24) = 4.40, p = .065$, and $F(3, 24) = 2.87, p = .285$, respectively). In the blue reference illumina-

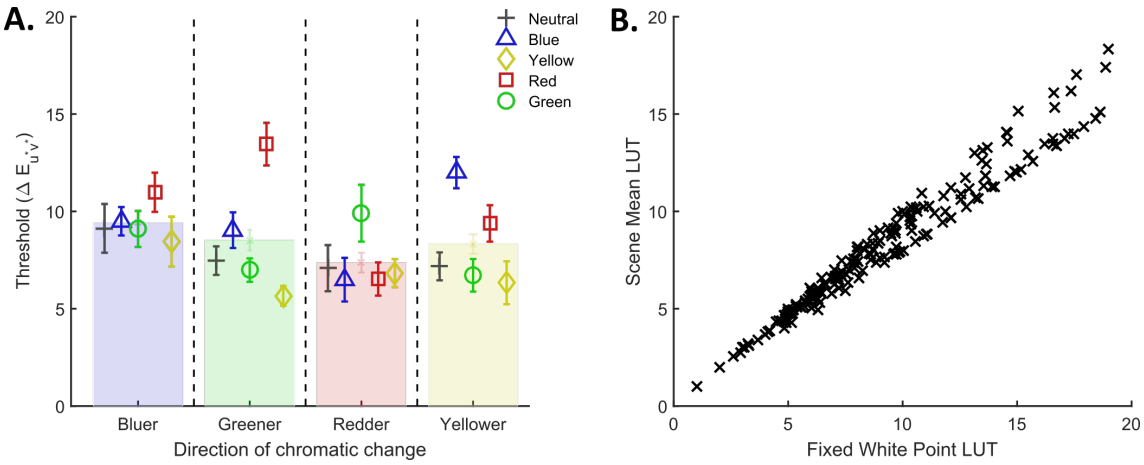


Figure 3.7: Thresholds calculated using the scene mean look up table. A. Thresholds across the different directions of change for the five reference illumination conditions. Transparent bars represent the main effect of chromatic direction of change. They are the thresholds for each chromatic direction averaged over reference illuminations. B. Scatter plot of the threshold data calculated using the scene mean look up table (LUT) plotted against threshold data calculated using the fixed white point illumination look up table (LUT).

Table 3.3: Mean IDT thresholds using the scene mean look up table. Values in brackets show the standard error.

Reference	Bluer	Greener	Redder	Yellower
Neutral	9.13 (1.25)	7.47 (0.73)	7.08 (1.18)	7.18 (0.72)
Blue	9.49 (0.73)	9.03 (0.92)	6.49 (1.12)	11.99 (0.80)
Green	9.10 (0.92)	6.98 (0.60)	9.90 (1.46)	6.72 (0.84)
Red	10.98 (1.01)	13.45 (1.09)	6.53 (0.85)	9.38 (0.94)
Yellow	8.45 (1.29)	5.66 (0.50)	6.83 (0.72)	6.33 (1.10)

tion condition, thresholds for the yellower direction of chromatic change were significantly higher than for the redder direction ($p = .028$). No other comparisons were significant. For the red reference illumination condition, thresholds in the redder direction of chromatic change were significantly lower than those for the bluer or greener direction of change ($p = .029$ and $p = .010$, respectively). No other comparisons were significant.

There was a simple main effect of reference illumination on illumination discrimination thresholds for the greener and yellower chromatic directions of change ($F(4, 32) = 16.10, p < .004$ and $F(4, 32) = 10.76, p < .004$, respectively), but not for bluer and redder ($F(4, 32) = 1.86, p = .568$ and $F(1.69, 164.76) = 3.62, p = .244$, with a Greenhouse-Geisser correction, respectively). For the greener direction of change, the red reference illumination

produced a threshold significantly higher than all other reference illumination conditions except blue ($p = .019$, $p = .002$, and $p < .001$, for the neutral, green, and yellow reference illumination conditions, respectively). For the yellower direction of change, the highest threshold was in the blue reference illumination condition and this was significantly higher than in the neutral, green, and yellow reference illumination conditions ($p = .011$, $p = .012$ and $p = .028$, respectively).

3.3.3 Using the variable white point illumination look up table

When thresholds are calculated using the reference illumination from each condition as the white point, the ordering of mean thresholds within each condition and across conditions remains the same (Figure 3.8.A, Table 3.4). Again, we quantified the strength of the correspondence between the two threshold sets by calculating Pearson's correlation coefficient ($r = .999, p < .001$; Figure 3.8.B).

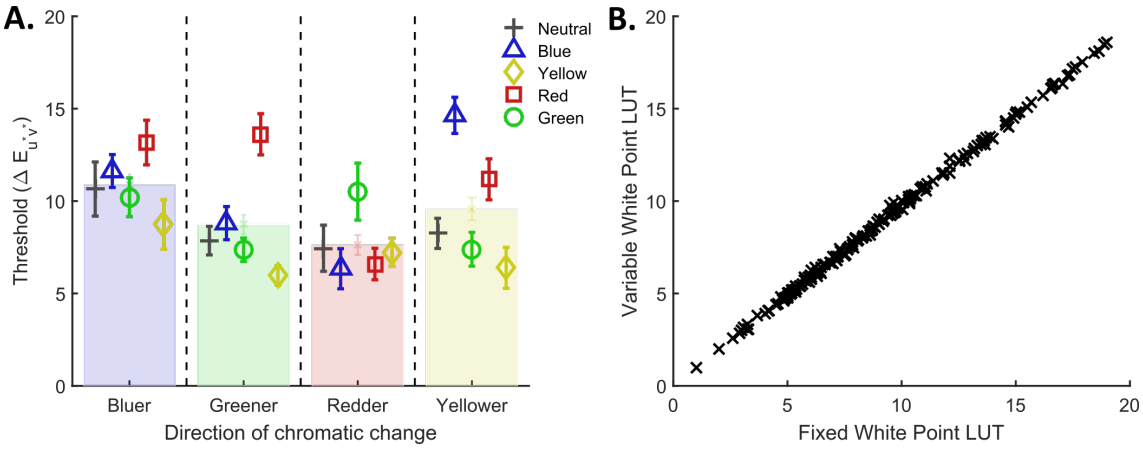


Figure 3.8: IDT thresholds calculated using the variable white point illumination look up table. A. Thresholds across the different directions of change for the five reference illumination conditions. Transparent bars represent the main effect of chromatic direction of change. They are the thresholds for each chromatic direction averaged over reference illuminations. B. Scatter plot of the threshold data calculated using the variable white point illumination look up table (LUT) plotted against threshold data calculated using the fixed white point illumination look up table (LUT).

Once again we can repeat the ANOVA analysis on these data. We find a significant interaction effect of reference illumination and chromatic direction of illumination change on illumination discrimination thresholds ($F(12, 96) = 9.11, p < .001$). Moreover, there was

Table 3.4: Mean IDT thresholds using the variable white point illumination look up table. Values in brackets show the standard error.

Reference	Bluer	Greener	Redder	Yellower
Neutral	10.65 (1.47)	7.86 (0.77)	7.45 (1.25)	8.25 (0.82)
Blue	11.63 (0.89)	8.81 (0.90)	6.34 (1.09)	14.64 (0.98)
Green	10.20 (1.05)	7.36 (0.63)	10.51 (1.55)	7.39 (0.91)
Red	13.17 (1.21)	13.61 (1.11)	6.59 (0.85)	11.18 (1.11)
Yellow	8.73 (1.34)	5.98 (0.53)	7.23 (0.76)	6.39 (1.11)

a significant main effect of reference illumination, regardless of the direction of chromatic change ($F(4, 32) = 12.70, p < .001$), and a significant main effect of chromatic direction of illumination change, regardless of reference illumination condition ($F(3, 24) = 8.34, p = .001$). When thresholds are averaged over the different reference illumination conditions, thresholds for the bluer and yellower chromatic direction of change are significantly higher than redder thresholds ($p = .024$ and $p = .007$, respectively). There were no other significant differences between the different directions of chromatic change (Figure 3.8.A).

As the interaction term was significant, we explored simple main effects. There was a simple main effect of chromatic direction of illumination change for the blue and red reference illumination conditions ($F(3, 24) = 15.80, p < .005$ and $F(3, 24) = 11.80, p < .005$, respectively), but not in the neutral, green, and yellow reference illumination conditions ($F(3, 24) = 4.20, p = .064$; $F(3, 24) = 4.57, p = .055$; and $F(3, 24) = 2.88, p = .285$, respectively). In the blue reference illumination condition, thresholds differed between the following directions of change: bluer vs. redder ($p = .013$); greener vs. yellower ($p = .045$); and redder vs. yellower ($p = .003$). For the red reference illumination condition, thresholds differed between the redder direction of chromatic change and all other directions of change ($p = .006$, $p = .010$, and $p = .010$, for blue, green, and yellow, respectively). Thresholds in the redder direction of change were significantly lower than thresholds in all other directions, with thresholds highest for the bluer and greener directions of change.

There was a simple main effect of reference illumination on illumination discrimination thresholds in all chromatic directions ($F(4, 32) = 4.49, p = .020$ for bluer, $F(4, 32) = 14.54, p < .004$ for greener, and $F(4, 32) = 17.17, p < .004$ for yellower) except redder

($F(1.63, 13.07) = 4.57, p = .148$, with a Greenhouse-Geisser correction). Illumination discrimination thresholds for the bluer direction of change were lowest in the yellow reference illumination condition and highest in the red reference illumination condition, although no pairwise comparisons among these thresholds were significant. For the greener direction of change, the red reference illumination produced a threshold significantly higher than all other reference illumination conditions except blue ($p = .028$, $p = .004$, and $p < .001$, for the neutral, green, and yellow reference illumination conditions, respectively).

Finally, for the yellower direction of change, the highest threshold was in the blue reference illumination condition and this was significantly higher than in the neutral, green, and yellow reference illumination conditions ($p = .005$, $p = .004$ and $p = .004$, respectively). In addition, thresholds for the yellower direction of change in the red reference illumination condition were significantly higher than for the green and yellow reference illumination conditions ($p = .029$, and $p = .007$, respectively).

3.4 Discussion

A group of observers completed five versions of the IDT that is considered a measure of colour constancy. One version (the neutral condition) was the standard IDT used by Pearce *et al.* (2014) and Radonjić *et al.* (2016*b*). In the other four versions (the blue, yellow, red and green conditions) the reference illumination, and all corresponding test illuminations, were shifted in chromaticity space such that the scene was chromatically biased in appearance. We had two main goals. Firstly, we aimed to establish if the results of Radonjić *et al.* (2016*b*), where relative thresholds to the different directions of chromatic change in the illumination depend on the mean scene chromaticity, is specific to a chromatic bias in the retinal image caused by a change in the scene surfaces, or if the effect is also seen when the chromatic bias is introduced by biasing the chromaticity of the illuminations. Secondly, we asked if the "blue bias" in thresholds for illumination discrimination (and so colour constancy), previously illustrated using the IDT, is a trait specific to versions of the task where bluer changes in the illumination occur along the

daylight axis.

We found that relative thresholds to the different directions of change are modulated by chromatically biasing the illumination set used in the task in a similar way to how biasing the mean scene surface reflectance affects thresholds. Like Radonjić *et al.* (2016*b*), we find that illumination discrimination thresholds are increased in the chromatic direction opponent to the direction of chromatic bias introduced to the scene. In our experiment, the chromatic bias is introduced by biasing the illumination set used during the experiment, while Radonjić *et al.* (2016*b*) introduced bias by modulating the mean surface reflectance of the scene but keeping the illumination set fixed. However, unlike Radonjić *et al.* (2016*b*), we do not consistently find that thresholds are decreased in the direction of chromatic bias. Regardless, these results show that illumination discrimination thresholds must be defined not only relative to the scene surface ensemble, but also with respect to the illumination set used in the task.

The results cannot be accounted for by simple adaptation (to the illumination) effects that would seem to predict opposing results to those that we find. Suppose, for example, that in the red reference illumination condition, the participant becomes adapted to the chromaticity of the reference illumination. Then we would predict a decrease in sensitivity to redder changes in the illumination and an increase in sensitivity to greener changes. This is the opposite of what our results show, suggesting simple adaptation effects cannot account for the asymmetry in thresholds. In future studies, all reference illumination conditions could be interleaved to equate adaptational state in all conditions of the task. If threshold asymmetries are still present, this would completely rule out this explanation.

Radonjić *et al.* (2016*b*) offered an explanation for their results that attributed the effect to a reduction of information about the illumination change. They suggest that in scenes where the mean surface reflectance is biased to reflect particular wavelengths of light, illuminations made up of wavelengths that the mean surface reflectance does not favour are harder to discriminate between as the scene reflects fewer wavelengths in the areas of the visual spectrum where the two illuminations differ. This explanation cannot be extended

to our results as the mean surface reflectance of the scene is kept constant across the different conditions of the task. In fact, in this experiment, the mean surface reflectance is relatively flat (Figure 3.2.B) and so illumination changes that result in perceptually equivalent colour changes should be equally perceptible on average across the whole scene. We offer an alternative explanation for the effect in terms of a daylight illumination prior.

Suppose our observers each have strong daylight illumination priors ($P(I)$, for illuminations I); their prior assumption is that the probability of different illumination chromaticities, for any particular scene, follows the frequency of daylight illumination chromaticities that they have experienced. Let's also suppose that our observers utilise a generative model that describes how illuminations of different chromaticities interact with surfaces in the world to give rise to different patterns of retinal excitation (although we could have a generative model that predicts stimulation to a later stage of visual processing and the same logic applies). Considering the vast amount of possible surface spectral reflectance functions that could be encountered in the world, the variety of scene configurations that are possible, and the number of metameric illumination spectra, this will give rise to a likelihood function (the probability of different retinal excitations E, given different illuminations chromaticities, I, or $P(E|I)$) that is highly variable. In this case, the prior will be given more weight in computations of the posterior ($P(I|E) = cP(E|I)P(I)$, where c is a normalisation constant that ensures the posterior integrates to 1), from which the illumination chromaticity is estimated. This will bias estimates towards daylight illumination chromaticities. Thus, observers will generate better (closer to the true value) illumination chromaticity estimates when the true illumination chromaticity is closer to the variation in daylights. The closer the estimate of the illumination chromaticity is to the true chromaticity, the better the colour constancy correction will be for the scene and the more likely it is that the colour appearance of surfaces in the scene appear close to the chromatic correlate of their fixed surface reflectance property. This makes a prediction that participants will display a bias for colour constancy (worse illumination discrimination) along the axis of change closest to daylight chromaticities, the effect that we see in the

behavioural data.

This explanation though is only half the story. Once an observer estimates the illumination chromaticity, they must still discount this from the scene to recover surface reflectance information. It is desirable then to form a spectral (rather than just chromatic) representation of the illumination. One such method for recovery of a spectral representation of an illumination from a chromatic one relies on a linear model of illuminations, where all illumination spectra can be re-expressed as a linear combination of a small number of basis functions (Brainard & Stockman, 2009). Linear models of illuminations are generally based on measurements of natural illumination spectra (Marimont & Wandell, 1992; Judd *et al.*, 1964). While some of the illuminations used in our experiment are constrained to have chromaticities on the Planckian locus, their spectral distributions differ from those of natural illuminations. If linear models formed through experience with natural illuminations are utilised by the visual systems of our observers to recover the spectral distribution of the experimental illuminations, then our illuminations must be well described by linear models of natural illuminations (or at least they must in cases where the observer remains colour constant). In Appendix A, we show that on average 64.95% ± 12.87% (mean ± SD) of the variation in the experimental illuminations can be captured by the CIE 3-component model of daylights (CIE, 2004); although the model explains a larger percentage of the variance in illuminations whose chromaticity is closer to the daylight axis.

Previously, we suggested our results may be explained by our observers invoking a prior over daylight chromaticities. As we now know that a large proportion of the variance in our illuminations can be captured by a linear model of natural daylight spectra, we could also take the approach of Brainard & Freeman (1997) who specified priors over the coefficients that recombined the basis functions of the linear model. Specifically, they used the CIE linear model of daylights to specify a "daylight prior" for illuminations. They incorporated this prior as part of an illumination estimation algorithm that recovered good estimates of the illumination spectrum present on a scene. They later showed that

a modified version of the model that no longer estimates the illumination spectrum but rather the illumination chromaticity (using an equivalent illumination approach to colour constancy (Brainard & Maloney, 2011)) can be fit well to behavioural data. However, model performance is improved by incorporating a prior for illuminations that is broader than the "daylight prior" and includes more atypical illumination chromaticities (Brainard *et al.*, 2006). It remains to be seen whether such a model would predict the biases we see in the behavioural data.

Our data show that the "blue bias" prevails in all reference illumination conditions and is not an effect specific to bluer illumination changes along the daylight axis. This suggests that if the "blue bias" that we see is evidence for an optimisation of colour constancy mechanisms for the statistics of the environment, this comes about through a reduction in sensitivity to global chromatic changes in a scene when the change is in a bluer direction (the nature hypothesis). This may seem contradictory to our previous explanation for increased colour constancy in the direction chromatically opponent to the change in the reference illumination in terms of a daylight prior. However, it is plausible that mechanisms of both kinds are at work. A reduction in sensitivity to global "blueish" chromatic changes in a scene could be a trait that has been engrained in the human visual system through evolution - a long term solution to colour constancy; whereas the formation of a daylight prior may be a more short term solution, observers learning the statistics of their environment through development.

Contrary to previous results, we do not find a main effect of direction of illumination change for the neutral reference illumination condition (Pearce *et al.*, 2014; Radonjić *et al.*, 2016*b*). However, note that in the present study, due to multiple experimental conditions (different reference illuminations), we actually perform simple main effects analyses. In other words, our p-value is Bonferroni corrected. If left uncorrected it is .014 and we would conclude a main effect of direction of illumination change, with thresholds highest for the blue direction of change.

Finally, we showed that the biases in our behavioural data are not explained by a bias in the

mean scene chromaticity, or the use of a fixed white point across experimental conditions, by analysing the results using three different look up tables. The reason for using the mean scene look up table is to determine whether representing the change between illuminations is better represented by the change between the average chromaticity of the scene under the two illuminations, effectively using the average scene chromaticity as our experimental estimate of the illumination chromaticity. It might be, for example, that the observer estimates the illumination by averaging the reflected light from the entire scene. If this average changed less evenly per step in the staircase than the generated illumination chromaticities, it might be that an uneven pattern of thresholds would emerge when thresholds are plotted as the generated illumination chromaticities. The fact that the pattern of thresholds still varies with chromatic direction of change means that the average scene chromaticities are not changing unevenly enough to explain the asymmetries in the data. Similarly, perceptually uniform colour spaces are known to become less perceptually uniform the further away one moves from the white point. In addition, the choice of white point will effect the transformation to the space (or really, the space itself). By repeating our analysis using the variable white point look up table we showed that the behavioural biases in thresholds are independent of such choices.

Chapter 4

Individual differences in illumination discrimination ability

4.1 Introduction

In Chapter 3 we showed that thresholds for discrimination of chromatic changes in illumination must be defined not only relative to the set of surfaces in the scene (Radonjić *et al.*, 2016*b*), but also with respect to any chromatic bias in the illumination set used during the experiment. Additionally, we showed that despite the effect that a chromatic bias in the illumination set has on relative thresholds, the "blue bias" (Pearce *et al.*, 2014) prevails overall. Thus far, we have only considered thresholds for illumination discrimination at a group level. We are yet to establish if the "blue bias" is a universal trait of our observers or if individual differences are present in the data, the bias stronger in some than others.

There are two types of individual differences: intra-individual differences that are observed within an individual across different task conditions or repeats, and inter-individual differences that are observed between different individuals within a given task condition and are preserved in different task conditions or repeats. In this Chapter, we are primarily interested in discovering whether inter-individual differences are present in observer performance on the IDT. If we find evidence of inter-individual differences, especially in the level of "blue bias" that an individual displays, this suggests variations in colour constancy mechanisms that may arise through differences in development and experience with the world (the nurture hypothesis). Put differently, this would be supportive of the hypothesis that observers possess a learned illumination prior that has been formed during their lifetime and hence is different in different observers leading to variations in performance. Alternatively, if we find no evidence of inter-individual differences, this suggests universal mechanisms for colour constancy that have become engrained through evolution (the nature hypothesis). In order to establish if differences in performance on the IDT across observers can be considered reliable evidence of inter-individual differences, we must first establish that differences in task performance across observers are greater than differ-

ences within observers. In other words, we must first ascertain whether the magnitude of intra-individual differences on the task are smaller than the magnitude of inter-individual differences.

We investigate these individual differences by first re-examining the data from Chapter 3 at an individual level. We find evidence of inter-individual differences in illumination discrimination ability, showing that the ordering of thresholds among our observers remains consistent across different directions of chromatic change in the illumination and across different reference illumination conditions. By collecting a new data set from a group of observers who completed the original (neutral) version of the IDT three times each, we show that although there are intra-individual differences in performance on the IDT that are suggestive of a learning effect, inter-individual differences are independent of them. Inter-individual differences remained consistent across multiple runs of the task. Finally, we collect a larger data set to investigate the extent of inter-individual differences in the strength of the "blue bias". The observers in the larger data set also completed a test of chromatic contrast detection ability. We compare these data to illumination discrimination data to show that chromatic contrast detection thresholds do not predict the "blue bias" in illumination discrimination thresholds, although they do explain a small amount of the inter-individual variation. In addition, we check the validity of our comparison between illumination discrimination and chromatic contrast discrimination thresholds with a control experiment and develop ideal observer models for the IDT.

Our motivation for these studies is as follows. Firstly, we believe that if observers learn an illumination prior during their lifetime (the nurture hypothesis) then we may expect to see larger inter-individual differences in the level of "blue bias" that observers display than if the prior was learnt during evolution (the nature hypothesis) and manifests as a reduced sensitivity to global, bluer chromatic changes in a scene. In order to assess this, we must first establish that the IDT can reliably pick up on inter-individual differences. We establish this in Experiment 2. In Experiment 3, we aim to further investigate whether the "blue bias" in the IDT can be attributed to a reduced sensitivity to bluer chromatic

changes by comparing observer performance on the IDT to performance on a low-level chromatic contrast detection task. If the same biases are not present in chromatic contrast detection thresholds, this is evidence that the IDT taps into higher-level processes beyond colour discrimination - such as invoking the use of a learnt illumination prior (the nurture hypothesis). Similarly, by developing a retinal processing model using ISETBIO in Experiment 5 we are able to rule out the influence of information loss at the retina on threshold asymmetries in the IDT.

4.2 Experiment 1: Chapter 3 revisited

In Experiment 1, we take another look at the data from Chapter 3. We use these data to take a first look at inter-individual differences in performance on the IDT. To do so, we will consider data from each of our observers separately and ask if their performance relative to one another is preserved across the chromatic directions of change and across different reference illumination conditions.

4.2.1 Results

We find evidence of inter-individual differences in illumination discrimination thresholds. Thresholds are consistently ordered among our observers across the different directions of chromatic change and across the different reference illumination conditions. Consider the black threshold trace (filled circles) across the different directions of chromatic change for the neutral reference condition in Figure 4.1.A. This participant (participant 8) always has one of the lowest, if not the lowest, illumination discrimination thresholds. Looking across the rest of the figures (Figure 4.1.B-E), it is clear that this participant maintains a low threshold value throughout the conditions. Similarly, participant 9 always displays high threshold values. To quantify the level of consistency in the ordering of the threshold values, we took an average threshold (for each participant) from each reference illumination condition (averaged over the different directions of chromatic change) and correlated these

across the different conditions (Table 4.1). The mean correlation coefficient is $r = .694 \pm .100$ (M $\pm$ SD; mean coefficient of variation is 49.04 ± 13.51), however, only one correlation coefficient meets a Bonferroni adjusted significance level of $\alpha = .05/10 = .005$.

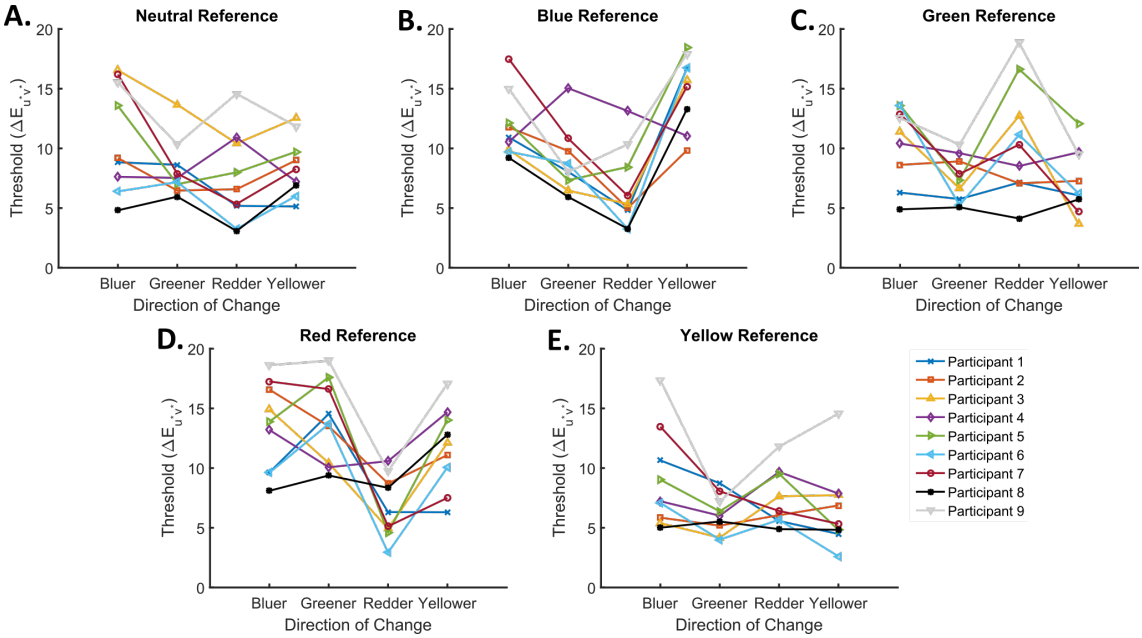


Figure 4.1: Inter-individual differences are consistent across different reference illumination conditions in the IDT. A-E. Individual threshold traces across the different directions of chromatic change in the illumination for each reference illumination condition. The traces for each participant are consistently coloured in each plot.

Table 4.1: Pearson's correlation coefficient between mean thresholds (across the different directions of chromatic change) in each reference illumination condition. Presented as $r(p)$.

	Blue	Green	Red	Yellow
Neutral	.491(.179)	.638(.065)	.660(.053)	.657(.054)
Blue		.754(.019)	.668(.049)	.797(.010)
Green			.765(.016)	.666(.050)
Red				.843(.004)

The main result that we find using the IDT is a reduced sensitivity to bluer chromatic changes in the illumination compared to the other directions of change. We have hypothesised that this effect is due to a learnt illumination prior for daylight illuminations whose variation can be captured by the Planckian locus. If such a prior is learnt through experience with the world, we may expect that different individuals possess different illumination priors due to differences in the illuminations to which they are most often exposed. In

this case, in addition to the overall differences in illumination discrimination ability that we highlight above, we may also expect to see inter-individual differences in the amount of "blue bias" that observers display. To assess this we must quantify each individual's "blue bias". We do this by subtracting thresholds for each of the greener, redder and yellower directions of change from thresholds for the bluer direction and then averaging over these differences and call this the "blue bias". Higher values indicate more bias and vice versa. For the neutral reference illumination condition, there is variability in the amount of "blue bias" displayed by each observer (minimum = -0.94, maximum = 9.02 $\Delta E_{u^*v^*}$). However, while the average "blue bias" does not differ across the different reference illumination conditions (repeated measures ANOVA, $F(4, 32) = 0.51, p = .729$; Figure 4.2.A), the ordering of the participants is only partially maintained (mean correlation coefficient of $r = .460 \pm .215$ (M SD); mean coefficient of variation is 25.26 ± 18.86; all non-significant at the $\alpha = .05/10 = .005$ Bonferroni adjusted significance level; 4.2.B-K).

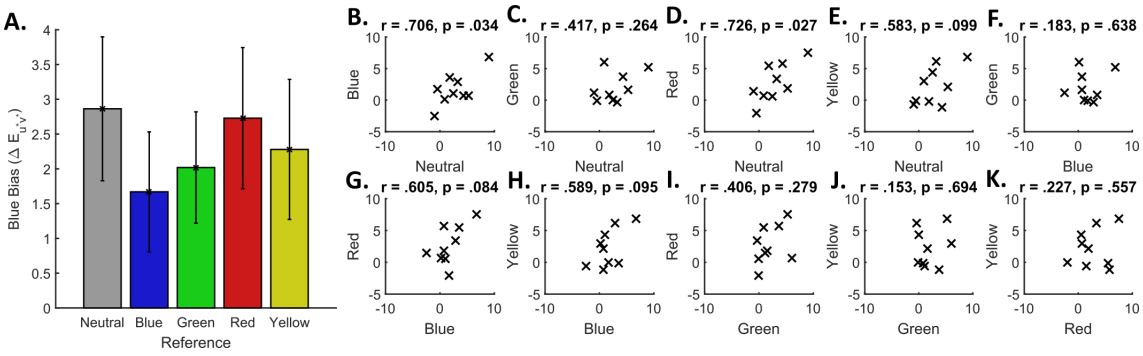


Figure 4.2: Quantifying the "blue bias" in different reference illumination conditions. A. The mean "blue bias" in the five reference illumination conditions. B-K. Scatter plots comparing the "blue bias" in each pair of reference illumination conditions with Pearson's correlation coefficient shown above.

4.2.2 Interim discussion

These data show that there are inter-individual differences in generic illumination discrimination ability; those who display increased/decreased sensitivity to illumination changes continue to do so across multiple directions of change and across different task conditions. What is less clear from these data is whether intra-individual differences are also present

in the level of "blue bias" displayed by observers. While we will return to these data in Section 4.6 when we begin to model individual differences in illumination discrimination ability, we need a different and larger data set initially in order to assess the extent of inter-individual differences in performance. In addition, to examine intra-individual differences we must assess how observers perform in repeated runs of the task. This is the route we take in the following Experiments.

4.3 Experiment 2: Reliability of the IDT

To establish the reliability of the IDT we recruited a group of participants who completed the original (neutral) version of task on three separate occasions. This allows us to assess whether there are intra-individual differences across repeated runs of the task. Moreover, whether intra-individual differences are present or not, we can ask if inter-individual differences are stable across repeated runs; do observers who display increased/decreased sensitivity to illumination changes, relative to the average, continue to do so in subsequent runs of the task?

4.3.1 Methods

4.3.1.1 Participants

Twelve participants were recruited (3 male, 9 female, mean age of 34.19 ± 16.40 years). All participants had normal or corrected to normal visual acuity and no colour vision deficiencies, assessed using Ishihara Colour Plates and the Farnsworth-Munsell 100 Hue Test.

4.3.1.2 The IDT

The task was the same as the neutral version of the IDT detailed in Chapter 3 with the addition of one extra feature: a set of trials were added at the start of the task to find a

range of values in which we were sure the participant could discriminate the test from the target. In short, the trials were in place to establish a range of values in which to start each staircase. In addition, the number of test illuminations along each axes of change was increased to 50 such that the maximum distance from target to test was 50 $\Delta E_{u^*v^*}$ along each axis of chromatic change. The experimental procedure, instructions and data analysis protocol were identical to those detailed in Chapter 3 (using the fixed white point LUT). Each participant attended the laboratory on three separate occasions (at least one day between each visit) and completed the neutral IDT on each visit.

4.3.2 Results

4.3.2.1 Intra-individual differences across repeated runs of the task

A 4×3 repeated measures ANOVA with direction of chromatic change (4 levels) and run (3 levels) as the independent variables and threshold value (in $\Delta E_{u^*v^*}$) as the dependent variable revealed a main effect of run on thresholds regardless of direction of change ($F(2, 22) = 11.37, p < .001$; Figure 4.3). Observers' performance on the IDT improved with practice, thresholds in the first run were significantly higher than thresholds in both the second and third runs ($p = .014$ and $p = .005$, respectively). There was no difference between overall thresholds in the second and third run ($p > .05$). There was also a main effect of direction of change on thresholds regardless of run ($F(1.80, 19.83) = 10.27$, $p = .001$, with a Greenhouse-Geisser correction). Thresholds were highest for the bluer direction of chromatic change. They were significantly higher than for the redder and greener directions of change ($p = .008$ and $p = .009$, respectively), but not for the yellower direction ($p = .096$). There were no significant differences between any other directions of change. Moreover, there was no significant interaction effect of run and direction of chromatic change on thresholds ($F(2.77, 30.42) = 0.92$, $p = .439$, with a Greenhouse-Geisser correction).

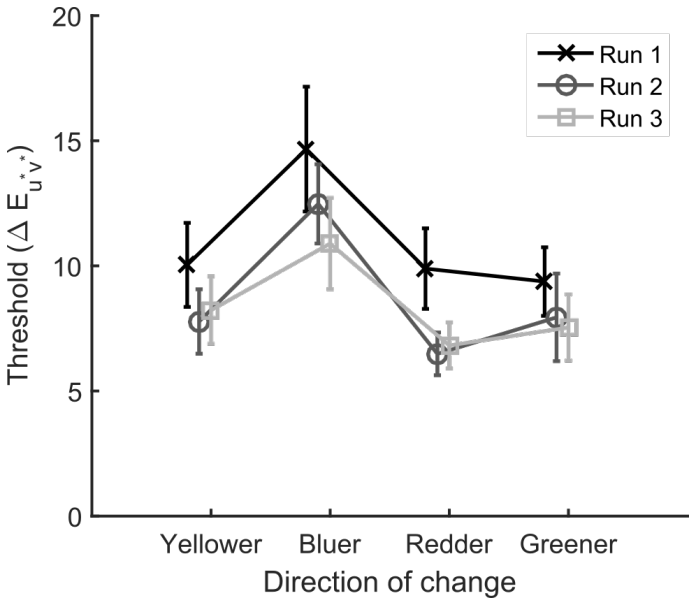


Figure 4.3: Intra-individual differences across runs on the original (neutral) IDT.

4.3.2.2 Inter-individual differences are consistent across directions of change and repeated runs of the task

Inter-individual differences in thresholds are consistent across the different directions of chromatic change and across multiple runs of the task. Considering data only from the first run of the task, we see that thresholds for the different directions of chromatic change are strongly related; although the correlation is only significant in two of the six cases (note that we set the significance level at $\alpha = .05/6 = .0083$ here to account for multiple comparisons; Figure 4.4). In addition, inter-individual differences in thresholds for a given direction of change are preserved across runs of the task with thresholds for a given direction of change strongly correlated across runs (six of the twelve correlations are significant at a Bonferroni adjusted significance level of $\alpha = .05/12 = .0042$; Figure 4.5).

4.3.2.3 Consistency of the blue bias

We quantified the amount of "blue bias" displayed by each observer in the three runs of the IDT in the same way as we did for the different reference illumination conditions in Experiment 1 (this Chapter). Here, we find large variability in the amount of "blue bias"

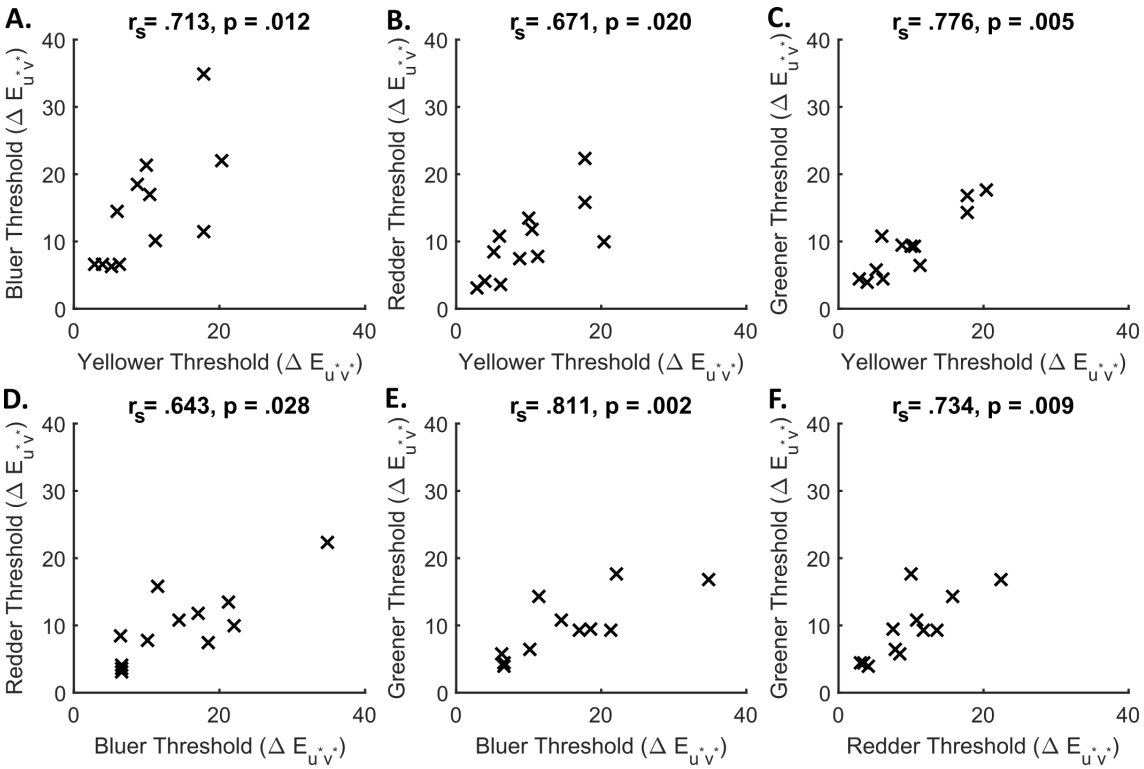


Figure 4.4: Inter-individual differences are consistent across the directions of chromatic change for the neutral IDT (data from the first run of the task only). A-F. Scatter plots comparing thresholds for the different chromatic directions with Spearman's correlation shown above.

across observers, which is strong evidence for the presence of inter-individual differences (minimum = −4.49, maximum = 15.86 $\Delta E_{u^*v^*}$). The average "blue bias" does not differ across the runs (repeated measures ANOVA, $F(2, 22) = 1.20$, $p = .159$; Figure 4.6.A). In addition, the correlations between the "blue bias" values calculated from the repeated runs of the task are significant in two out of three cases (at a Bonferroni adjusted significance level of $\alpha = .05/3 = .017$; Figure 4.6.B-D).

4.3.3 Interim discussion

Taken together, Experiments 1 and 2 provide strong evidence for the existence of inter-individual differences in performance on the IDT that are independent of intra-individual variation. While the decrease in average thresholds (across the chromatic directions of change) in repeated runs of the task suggest a learning effect on the IDT, the inter-

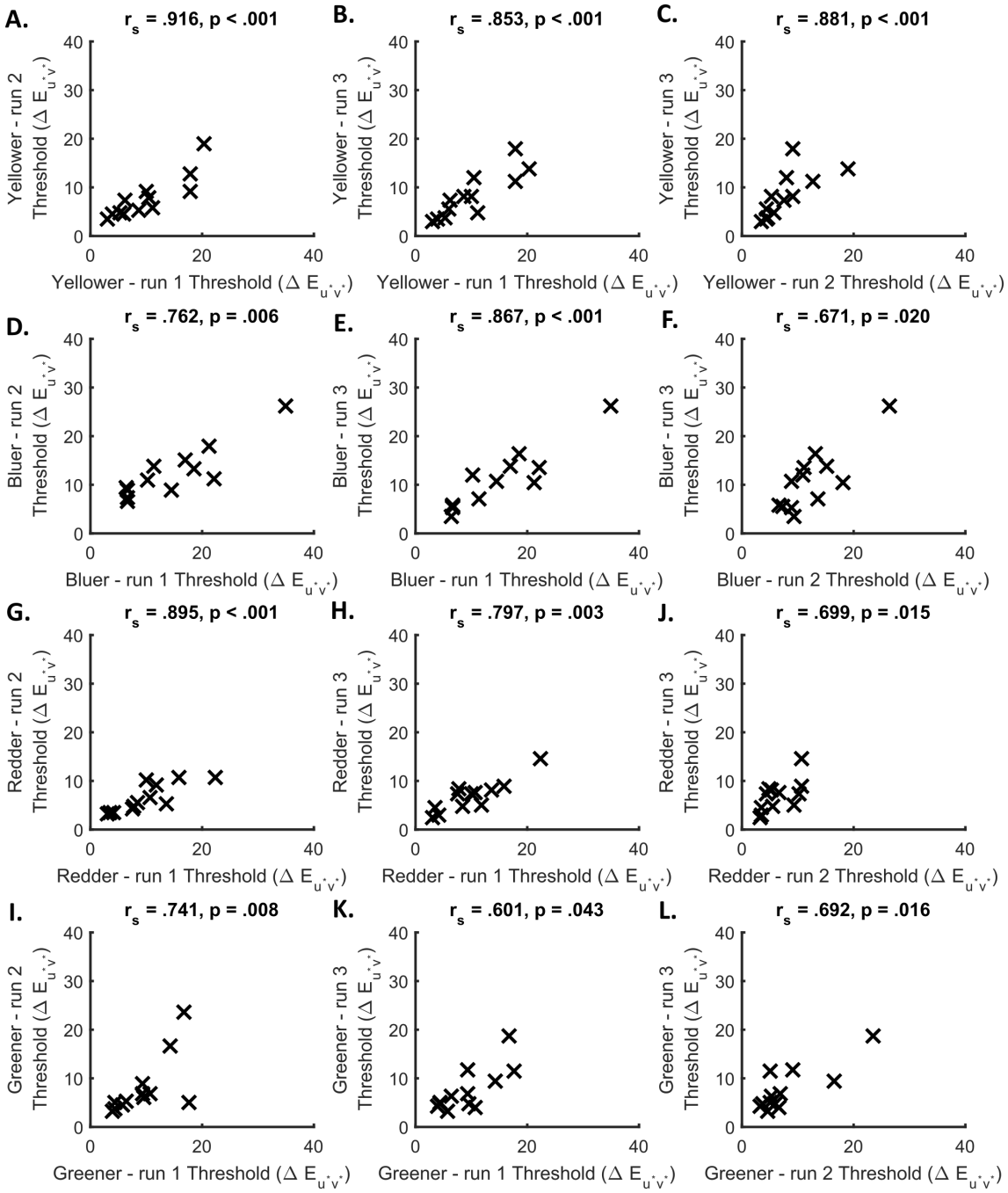


Figure 4.5: Inter-individual differences are consistent across multiple runs of the neutral IDT. A-L. Scatter plots comparing thresholds for the different chromatic directions of change across runs with Spearman's correlation shown above.

individual differences that we observe are well-maintained across the repeated runs of the task. A particularly interesting feature of these data is the inter-individual variability in the amount of "blue bias" displayed by observers. In Experiment 3, we asked if

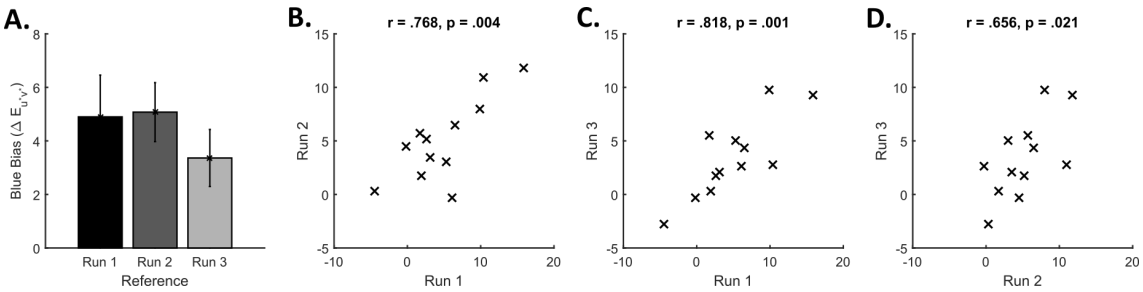


Figure 4.6: Quantifying the "blue bias" across repeated runs of the neutral IDT. A. The mean "blue bias" in the three repeats of the task. B-D. Scatter plots comparing the "blue bias" in each pair of repeats with Pearson's correlation coefficient shown above.

inter-individual differences in chromatic contrast discrimination ability explain the inter-individual differences we observe in the IDT.

4.4 Experiment 3: Inter-individual differences

In Experiment 3, we asked all participants from Experiment 2 to return to the laboratory to complete a chromatic contrast detection task (CCDT). In addition, we recruited further participants who completed the CCDT and one run of the IDT. The goal of this study was to ascertain whether performance on the CCDT (a measure of low-level colour contrast detection ability) can predict the inter-individual differences in performance on the IDT.

4.4.1 Methods

4.4.1.1 Participants

In addition to the 12 participants from Experiment 2, a further 43 participants were recruited, a total of 55 for the study (23 male, 32 female, mean age of 27.83 ± 12.10 years). All participants had normal or corrected to normal visual acuity and no colour vision deficiencies, assessed using Ishihara Colour Plates and the Farnsworth-Munsell 100 Hue Test.

4.4.1.2 The CCDT

The chromatic contrast detection task (CCDT) was originally introduced by Cranwell, Pearce, Loveridge & Hurlbert (2015). The data we present here were collected using a later version of the task that differed from the task described by Cranwell *et al.* (2015) in several ways, although the premise of the task remained the same. Rather than detail the differences between the two tasks, we describe the parameters of the later version in full here.

The CCDT is designed to find chromatic (and achromatic) contrast detection thresholds along the three cardinal axes of DKL colour space (Derrington *et al.*, 1984, see also Section 2.1.4). These axes describe the red-green ($L-M$), blue-yellow ($S-(L+M)$) and luminance ($L+M$) channels at the second stage of visual processing. In the CCDT, arrows are presented against a neutral background ($D65$ at a luminance of 50 cd/m^2; $CIE\ Yxy = [50, 0.31, 0.32]$). If the participant detects the arrow they are required to respond using a button press to indicate whether the arrow was pointing left or right (left or right control keys on the keyboard). In the version of the task that we used here, luminance noise ($+5, +2.5, -2.5$ or -5 cd/m^2) is added to both the background and the arrow by dividing the whole screen into 2 mm squares (≈ 0.23 degrees of visual angle at the fixed viewing distance of 50 cm) and randomly deciding, for each pixel, whether to add no noise or to jitter the luminance by one of the four values listed previously. The noisy background was set to change at a rate of 6 Hz (every 10 frames of our 60 Hz screen) throughout the task. The arrows vary in chromaticity from the background along one of the axes of DKL space. The resolution along each axis is determined by the capabilities of the monitor which varies according to the background luminance that is chosen. We found the smallest possible step size along each cardinal DKL axis that could be achieved on our monitor using a background luminance of 50 cd/m^2. These were step sizes of 0.01 along the luminance axis, 0.002 along the red-green axis, and 0.008 along the blue-yellow axis. We generated 50 steps along each half-axis. The arrow was specified such that it fits inside a square of side length 3.75 mm (≈ 4.3 degrees of visual angle at the fixed viewing

distance of 50 *cm*). The horizontal position of the arrow's centre was fixed at the centre of the screen throughout the experiment. The vertical position of the arrow's centre varied by 0 to 4 *cm* ($\approx$ 4.6 degrees of visual angle at the fixed viewing distance of 50 *cm*) on each trial, a parameter that varied at random. The start of a new trial was indicated by the appearance of a small 5 *mm* square ($\approx$ 0.58 degrees of visual angle at the fixed viewing distance of 50 *cm*) appearing in the centre of the screen. The task was run on a 10-bit ASUS Proart LCD screen. A head rest was used to ensure participants maintained a distance of 50 *cm* from this screen. The monitor was controlled using a 64-bit Windows machine, equipped with an NVIDIA Quadro K600 10-bit graphics card, running MATLAB scripts that used Psychtoolbox routines (Brainard, 1997; Kleiner, Brainard & Pelli, 2007; Pelli, 1997). The stimuli were colorimetrically calibrated using a linearised calibration table based on measurements of the monitor primaries made with a Konica Minolta CS2000 spectroradiometer (Konica Minolta, Nieuwegein, Netherlands). Calibration checks were performed regularly throughout the study period and the calibration table was updated when needed.

4.4.1.3 Procedure

The 12 participants from Experiment 2 returned to the laboratory to complete the CCDT (Figure 4.7). The 43 participants that were recruited for this study attended the laboratory to complete the CCDT in addition to one run of the original (neutral) version of the IDT. The order in which the participants completed the two tasks was randomised to account for any generic learning effects.

The procedure for the IDT was the same as that described in Chapter 3. For the CCDT, participants first read the standardised instructions before being asked if they had any questions. All participants then received the following verbal instructions prior to starting the task: *"On each trial, an arrow will appear on the screen. If you see the arrow, indicate the direction that the arrow points using the control keys. Press the left control key if the arrow points left and the right control key if the arrow points right. On some trials, you*

will not see an arrow. In this case, you will be signalled the start of a new trial by the flash of a small white square in the centre of the screen". Prior to the practice trials and main trials, the instructions regarding how to issue a response were repeated to the participant by showing text on the screen that was read out by a text-to-speech engine. At the start of the task, participants were adapted to a uniform $D65$ background for 2 minutes, being asked to fixate on a small black cross in the centre of an otherwise uniform screen ($CIE\ Yxy = [50, 0.31, 0.32]$). On each trial, after an inter-stimulus interval between 1 and 3 seconds, an arrow was presented for 166.67 ms (10 frames of the 60 Hz screen). The participant had 5 seconds to respond by pressing either of the control keys to indicate the direction of the arrow. If they did not respond within this time period, a random response (correct/incorrect) was generated. Once a response was issued (or randomly generated) a small white square presented for 166.67 ms (10 frames of the 60 Hz screen) was presented to indicate the start of a new trial (Figure 4.7).

The task was run in three blocks, one for the luminance axis, one for red-green, and one for blue-yellow. Thresholds for each half-axis were determined using a 1-up, 3-down, transformed and weighted staircase procedure (Kaernbach, 1991). There were two staircases for each half-axis and so four staircases were interleaved in each block of the task. The staircases terminated after 50 trials or six reversals, whichever came first.

Each block of the task began with five unrecorded practice trials that were chosen to be supra-threshold for trichromat participants according to pilot data and previous studies (between 40 and 50 steps away). After this, a sequence of trials were completed that established a starting point for each pair of staircases; we call these the range finding trials. These began with five trials along each half-axis that were 20 steps away from the background. If the participant got four of these five correct, staircases for that half-axis started between 10 and 20 steps away (decided at random). If they got less than four correct, they completed five trials 30 steps away and so on such that staircases started in the intervals $10-20$, $20-30$, $30-40$ or $40-50$ steps away.

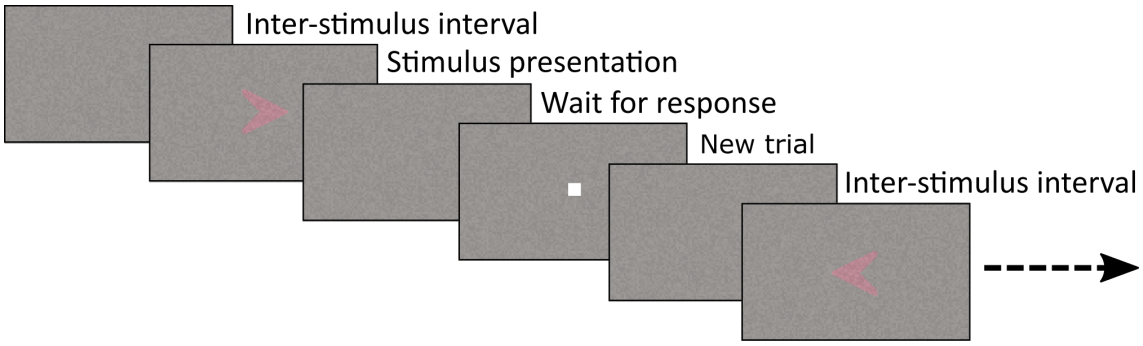


Figure 4.7: The chromatic contrast discrimination task CCDT (not to scale).

4.4.1.4 Data Analysis

Thresholds for the CCDT were calculated by taking the mean of the last two reversals from each of the two staircases (a mean of four reversals in total). A look-up table was created mapping each nominal staircase step (1 to 50) to the $\Delta E_{u^*v^*}$ value between the measured chromaticities of each arrow and the background. All reversal indices were first converted and then an average taken over the $\Delta E_{u^*v^*}$ values. The average background chromaticity (D65) was used as the white point for calculation of $\Delta E_{u^*v^*}$.

4.4.2 Results

4.4.2.1 The distribution of blue bias values in the IDT

Considering the data from all 55 observers, there is a main effect of chromatic direction of change on threshold values (Friedman's test, $\chi^2 = 56.35$, $p < .001$; Figure 4.8.A). Thresholds for the bluer direction of change were significantly higher than for the yellower, redder, or greener directions ($p < .006$ in all cases). In addition, thresholds for the yellower direction of change were significantly higher than for the redder or greener direction of change ($p = .018$ and $p < .006$, respectively). The redder and greener thresholds did not differ ($p > .05$).

This is evidence once again for a "blue bias" in illumination discrimination ability. However, if we consider the data at an individual level (Figure 4.8.A-B), we see that the blue

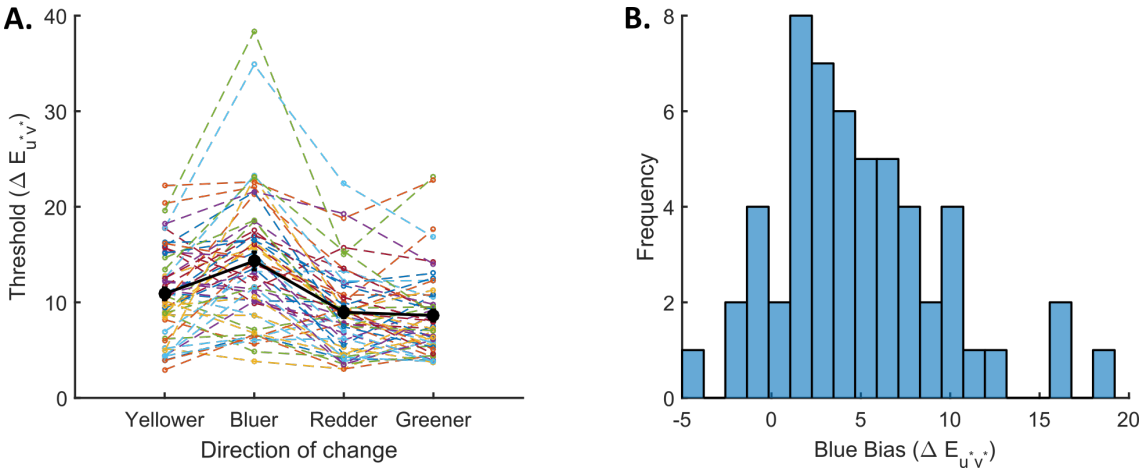


Figure 4.8: Inter-individual differences in relative performance on the IDT. A. Individual threshold traces for each observer. The average threshold trace is the solid black line. B. The distribution of "blue bias" values.

bias is not present in all individuals. On average, the "blue bias" values are significantly higher than zero ($t = 7.59$, $p < .001$). The mean "blue bias" is 4.83 ± 4.72 $\Delta E_{u^*v^*}$ (M $\pm$ SD). However, the largest value is 19.13 $\Delta E_{u^*v^*}$ while the minimum value is -4.49 $\Delta E_{u^*v^*}$, suggesting that the level of "blue bias" is highly variable across observers. In the following section, we ask if inter-individual differences in chromatic contrast discrimination ability can explain the variability we see in performance on the IDT.

4.4.2.2 Do colour contrast discrimination thresholds predict global illumination discrimination ability?

There is a main effect of chromatic direction of change on CCDT thresholds ($\chi^2 = 122.55$, $p < .001$; Figure 4.9.A). Blacker and whiter thresholds did not differ significantly from each other ($p > .05$), but both were significantly lower than thresholds for all other directions of change except redder ($p < .015$ in all cases). In addition, thresholds for the greener direction of change were not significantly different from redder or bluer thresholds ($p = 0.06$ and $p = 0.27$, respectively), but were significantly smaller than yellower thresholds ($p < .015$). Finally, thresholds for the redder direction of change were significantly smaller than for the bluer and yellower directions of change ($p < .015$ in both cases) which also

differed significantly from each other ($p < .015$). In short, thresholds for chromatic axes of change in the CCDT do not follow the same pattern as thresholds for the IDT (Figure 4.9.B).

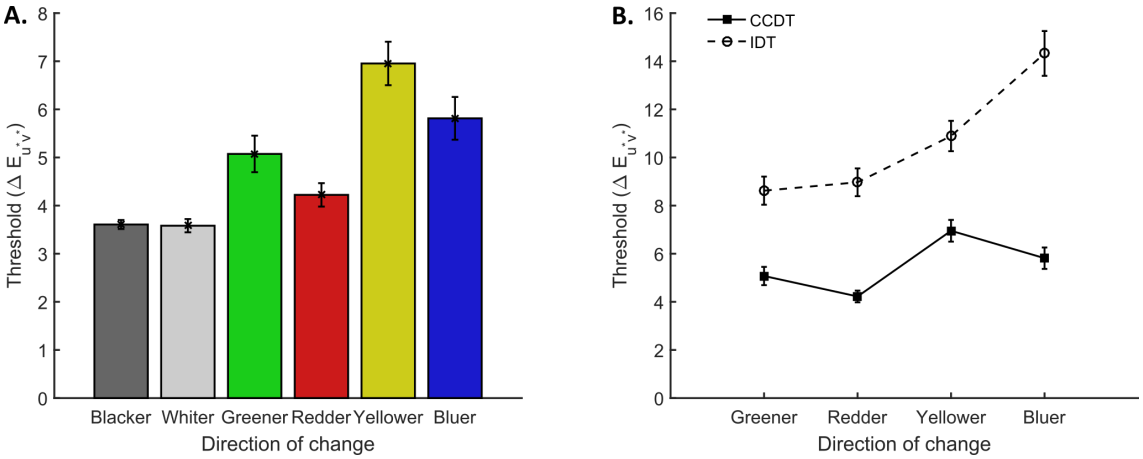


Figure 4.9: Thresholds for the chromatic contrast discrimination task. A. CCDT thresholds across all half-axes of change in the CCDT. B. CCDT thresholds compared to IDT thresholds. Error bars show ± 1 SEM.

We have already established that the "blue bias" seen in the IDT thresholds is not present in the CCDT data. We further illustrate this by computing Spearman's rank correlations between each chromatic axis of change in the IDT and each chromatic axis of change in the CCDT (Figure 4.10). If we adjust the level at which we accept significance using a Bonferroni correction we have $\alpha = 0.05/16 = 0.003$. In this case, only three of the 16 correlations are significant. With 16 comparisons, a Bonferroni correction is very conservative, but regardless, the average amount of variation that thresholds for a particular axis of the CCDT explain in IDT thresholds is only $12.41 \pm 5.31\%$ (M ± SD). This is not too surprising considering the differences in the axes of chromatic change used in the two tasks (Figure 4.11). In fact, if we make the assumption that the cardinal axes as they are defined in DKL colour space are an accurate and full description of second stage visual processing, then discrimination along each axis of chromatic change in the IDT will be governed by a pair of these mechanisms. For example, changes along the bluer axis in the IDT could be detected by both the blue and green half-axis ($S - (L + M)$ and $M - L$). If we make a further assumption that the second stage mechanisms act independently to

detect changes in the illumination in the IDT then we can use the CCDT thresholds to predict IDT thresholds.

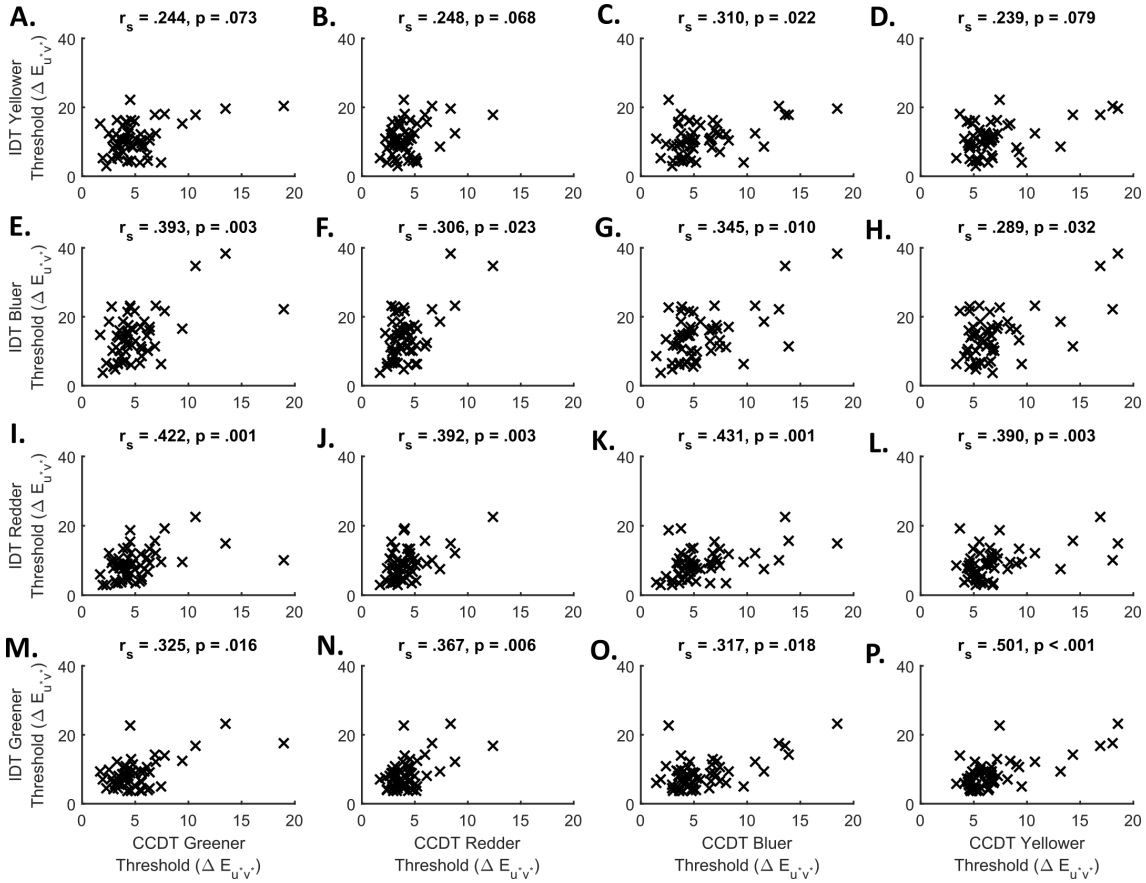


Figure 4.10: IDT vs. CCDT thresholds.

We first fit a psychometric function (PF) to the data for each half-axis of chromatic change in the CCDT for each observer. We chose to fit a PF from the family of Weibull functions. To fit the Weibull functions to the data, we used Palamedes Toolbox for MATLAB (Kingdom & Prins, 2010). We fixed the guess rate (γ) at 50% and allowed the threshold (α), slope (β), and lapse rate (λ) to vary; although the lapse rate was bounded such that $0.01 \leq \lambda \leq 0.05$. The resulting PF represents the probability that the observer would respond correctly for each stimulus value. We would like the probability that they detected the stimulus. We converted the correct response probabilities to detection probabilities using the formula below to form PFs for detection of the stimulus (see Figure 4.12 for a

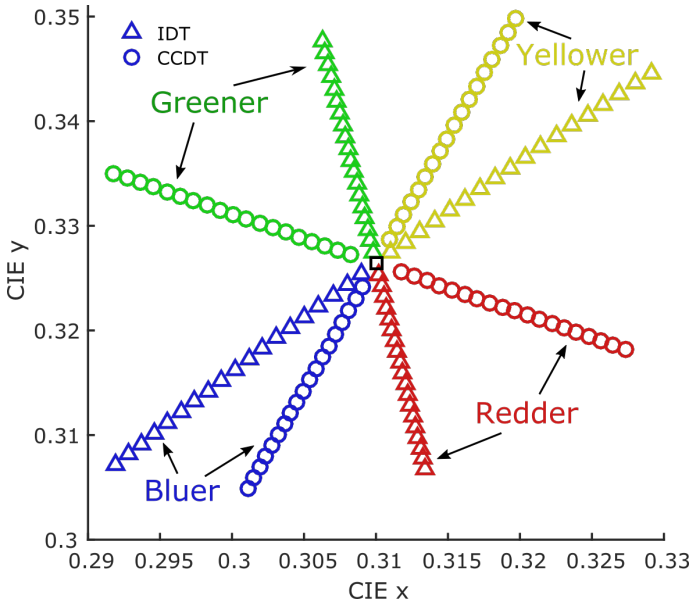


Figure 4.11: The chromatic axes of change in the IDT and CCDT.

derivation of this formula).

$$P(detect) = \frac{P(correct) - 0.5}{0.5 - P(lapse)}. \tag{4.1}$$

For each test illumination used in the IDT, we calculated its chromaticity in DKL space by considering each test relative to the reference illumination ($D65$, the white point or background in the calculation). We then used the PFs for detection of the stimulus along each half-axis in the CCDT to calculate the probability that at least one of the cone-opponent mechanisms (Luminance, red-green or blue-yellow) would detect the change in chromaticity from the reference illumination to each test illumination in the IDT. This is defined as:

$$P(detect_{IDT}) = 1 - [(1 - P(detect_{LUM})) + (1 - P(detect_{L-M})) + (1 - P(detect_{S-(L+M)}))]. \tag{4.2}$$

This formula can be used to find the probability of the visual system detecting a change at each stimulus level. We can then convert back to correct response probabilities for the

IDT as follows:

$$P(correct) = P(detect)(1 - P(lapse)) + (1 - P(detect))P(guess). \tag{4.3}$$

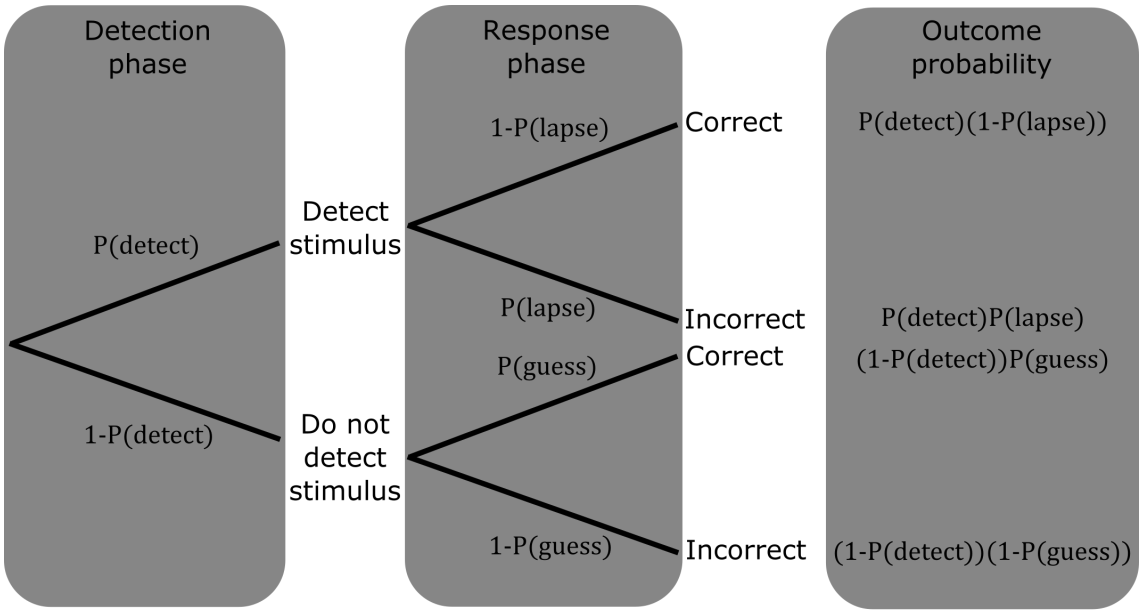
By fitting a PF to these values, we can find the stimulus level where the participant is predicted to respond correctly 75% of the time based on their CCDT thresholds. For these fits, we again fit a PF from the family of Weibull functions. Here, however, we also fix the lapse rate at the value obtained from the catch trials in the IDT for each individual. In some cases, this leads to a maximum percentage correct p% for certain observers such that $p < 75$, making it impossible to infer the 75% correct probability from fitted PF. For these observers, we took the stimulus level that corresponded to a percentage correct of $(p-1)$%, where it is sensible to assume the staircases would converge for such a participant.

Predictions of IDT thresholds made in this way still fail to capture the pattern of thresholds that we observe in the IDT (Figure 4.13.A). While predicted thresholds are close to measured thresholds for the redder and greener directions of chromatic change in the IDT, they fail to capture the elevation of yellower and bluer thresholds completely. However, if we calculate the correlation between each predicted and measured threshold set, we find that the predictions do represent inter-individual variability well ($r = .349$, $p = .009$ for the yellower direction, $r = .378$, $p = .005$ for the bluer direction, $r = .610$, $p < .001$ for the redder direction, and $r = .572$, $p < .001$ for the greener direction; Figure 4.13.B).

4.4.3 Interim Discussion

Having shown in Experiments 1 and 2 that there are inter-individual differences in performance on the IDT, we asked in Experiment 3 whether these differences are explained at the level of chromatic contrast detection ability. In particular, such a comparison allows us to address the question of whether the "blue bias" in IDT thresholds is explained by a reduced sensitivity to bluer global changes across a scene.

Firstly, we do not see the same pattern of thresholds across the chromatic directions of



$$P(correct) = P(detect)(1-P(lapse)) + (1-P(detect))P(guess)$$

$$P(detect) = (P(correct)-P(guess))/(1-P(guess)-P(lapse))$$

Figure 4.12: Calculating the probability of detection from a response probability in a 2AFC task. First, there is a detection phase where the participant detects the stimulus with probability $P(detect)$. Then, there is a response phase where the participant must issue one of two responses. If the participant detected the stimulus, they will respond correctly if they do not lapse with probability $1-P(lapse)$. If the participant did not detect the stimulus, they will guess and respond correctly with probability $P(guess) = 0.5$ in a 2AFC task. There are two possible branches that lead to a correct response. The sum of the probabilities of traversing each of these branches is the probability of a correct response $P(correct)$. The resulting formula can be rearranged to find the probability of detecting a stimulus.

change in the CCDT as we do in the IDT - there is no "blue bias". In addition, we do not see consistent correlations between the corresponding chromatic directions in the two tasks (plots across the diagonal in Figure 4.10). However, we acknowledge that there are differences between the chromatic axes of change used in the IDT and CCDT. The axes of change used in the CCDT are specified to fall along the cardinal axes of DKL colour space, generally assumed to represent the second stage of visual processing. By making the assumptions that these axes accurately and exhaustively represent the second stage of visual processing and that they act independently, we use the CCDT thresholds to predict thresholds on the IDT. We find that while CCDT thresholds can explain an average of 24.10% of the variation in the IDT thresholds (averaged over chromatic directions of

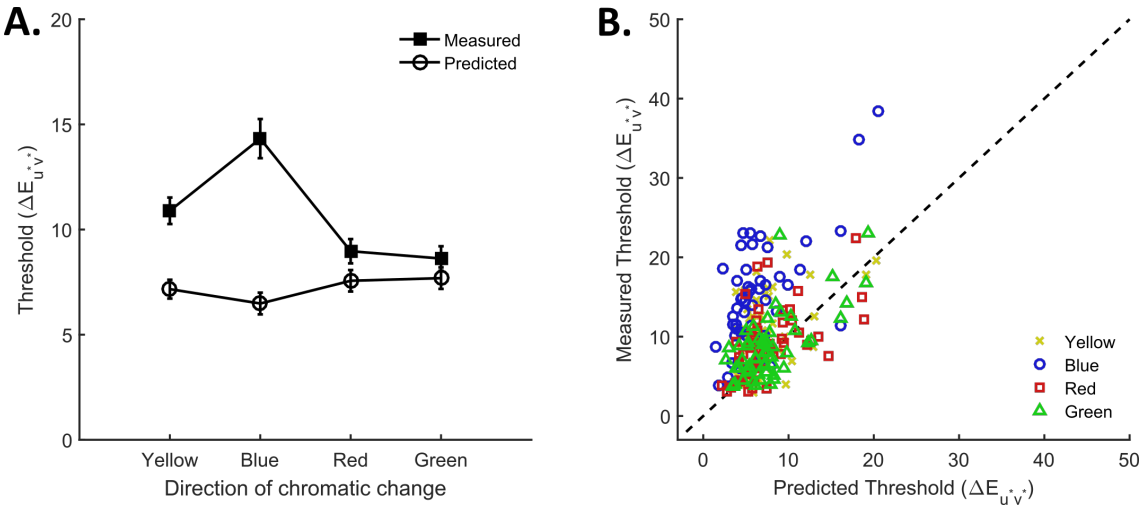


Figure 4.13: Predicting IDT thresholds from CCDT data. A. Comparing predicted and measured IDT thresholds. B. A scatter plot of measured against predicted thresholds for each direction of chromatic change in the IDT. Error bars show ± 1 SEM.

change), they do not predict the "blue bias". In fact, the CCDT thresholds predict more of the variation along the redder and greener directions of chromatic change in the IDT than they do for the bluer and yellower directions of change (32.49% compared to 19.32%). This suggests that illumination discrimination thresholds along these two pairs of axes are influenced differently by visual mechanisms. As the bluer and yellower illuminations fall along the daylight axis, representing the chromaticities of natural daylight illuminations, and as chromatic contrast detection thresholds do not show a similar bias, we take the results of Experiment 3 as further evidence for the influence of an illumination prior, representing the statistics of natural illuminations, on behavioural results in the IDT.

There are multiple reasons, however, why the comparison between IDT and CCDT data may not be valid. Firstly, there is disagreement in the colour vision community about the number of chromatic mechanisms that exist in the visual system, whether they act independently, and what the weightings of the different cone types are in the cardinal mechanisms (see Sections 1.4.1 and 1.4.2). In order to make predictions of IDT thresholds from CCDT data, we made the assumption that there are only two second stage mechanisms, that they are well-represented by the cardinal axes of DKL colour space, and that they act independently to detect chromatic changes in a stimulus. Secondly, there are

several important differences between the two tasks. The CCDT is a contrast detection task where the participant must detect the appearance of a target against a background. The participant is not required to make comparisons temporally across time as they are in the IDT, a task that involves a memory component. In the IDT, the participant must hold a representation of the reference and two comparison lights in memory in order to make a decision about which comparison was most similar to the reference. These issues mean that the predictions we make must be interpreted with caution. However, the fact that we do not see a "blue bias" in the CCDT data but do in the IDT data is still strong evidence that the bias is not caused solely by a reduction in sensitivity to bluer changes in a scene. If this was the case we would expect to see a manifestation of this reduced sensitivity in the data from both tasks.

4.5 Experiment 4: The validity of IDT predictions

In Experiment 3, we asked if individual differences in illumination discrimination ability could be explained by individual differences in chromatic contrast detection ability. By assuming that the cardinal mechanisms of DKL colour space accurately and exhaustively describe the second stage of visual processing, and that they independently detect chromatic changes in a stimulus, we used detection ability along these axes of DKL colour space from a chromatic contrast detection task (the CCDT) to predict illumination discrimination thresholds (in the IDT). We highlighted several reasons why results from the CCDT may not be predictive of results in the IDT. One such reason is that it is still debated whether the chromatic cardinal mechanisms (the $S-(L+M)$ and $L-M$, or the blue-yellow and red-green channels) provide an accurate description of the second stage of visual processing and, if so, whether these mechanisms act independently to detect chromatic changes in a stimulus. We explore these questions using the CCDT with this Experiment.

Here, we make predictions of thresholds within the same experimental paradigm. Namely, we remove the differences between the two tasks to ask if chromatic detection ability along

the cardinal axes of DKL colour space in the CCDT can be used to predict thresholds along different axes of change (intermediate axes) within the same experimental paradigm. If we find this to be the case, then our assumption that there are only two second stage mechanisms that act independently to detect chromatic changes in a stimulus is supported.

Furthermore, this could increase confidence in our illumination discrimination predictions, i.e., the predictions of the level of chromatic discriminability between two scenes in the IDT prior to any higher-level colour constancy processes taking effect. Conversely, if we find this not to be the case, we must interpret the predictions that we made of illumination discrimination thresholds from chromatic contrast detection thresholds more cautiously. In addition, the latter finding would suggest that either the chromatic cardinal mechanisms do not act independently, or that there are in fact additional, higher-order chromatic mechanisms.

We define two axes that are intermediate (at 45 degree angles) to the chromatic cardinal mechanisms in an equiluminant plane in DKL colour space. We refer to these as the lime-magenta and orange-cyan axes and hypothesise that chromatic contrast detection along each of the half axes is governed by the following pairs of cardinal chromatic half axes: green $(M - L)$ and yellow $((L + M) - S)$ for the lime half-axis, red $(L - M)$ and blue $(S - (L + M))$ for the magenta half-axis, red $(L - M)$ and yellow $((L + M) - S)$ for the orange half-axis, green $(M - L)$ and blue $(S - (L + M))$ for the cyan half-axis. Using the same prediction procedure that was used to predict IDT thresholds, we find that lime and cyan thresholds are systematically under-predicted while orange and magenta thresholds are systematically over-predicted. This suggests that if the chromatic cardinal axes of DKL colour space provide an accurate representation of the second stage of visual processing they do not act independently, but rather that the stimulation of one axis affects the sensitivity of the other.

4.5.1 Methods

4.5.1.1 Overview

A group of observers who contributed to the data set discussed in Experiment 3 were invited back to the laboratory to complete a new version of the CCDT. In this version, the chromaticity of the arrow was modulated along intermediate rather than cardinal axes in DKL colour space. We then asked if chromatic contrast detection ability along the chromatic cardinal axes of DKL colour space (used as the chromatic axes of change in the original CCDT task) predict chromatic contrast detection ability along these intermediate axes.

4.5.1.2 Participants

Twenty of the 55 participants discussed in Experiment 3 returned to the laboratory for further testing (10 male, 10 female, mean age of 21.89 ± 2.91 years).

4.5.1.3 The iCCDT

We modified the CCDT so that the chromaticity of the arrow was modulated along intermediate (rather than cardinal) axes in DKL colour space (the iCCDT). Two intermediate axes were defined and we refer to them as the orange-cyan and lime-magenta axes. The orange-cyan axes is defined such that the orange half-axis forms a 45 degree angle at the origin of DKL colour space in an equiluminant plane with both the red ($L-M$) and yellow ($(L+M)-S$) half axes and the cyan half-axis forms a 45 degree angle with both the green ($M-L$) and blue ($S-(L+M)$) half axes. The lime-magenta axes is defined such that the lime half-axis forms a 45 degree angle with both the green ($M-L$) and yellow ($(L+M)-S$) half axes and the magenta half-axis forms a 45 degree angle with both the red ($L-M$) and blue ($S-(L+M)$) half axes (Figure 4.14). All other aspects of the task remained the same.

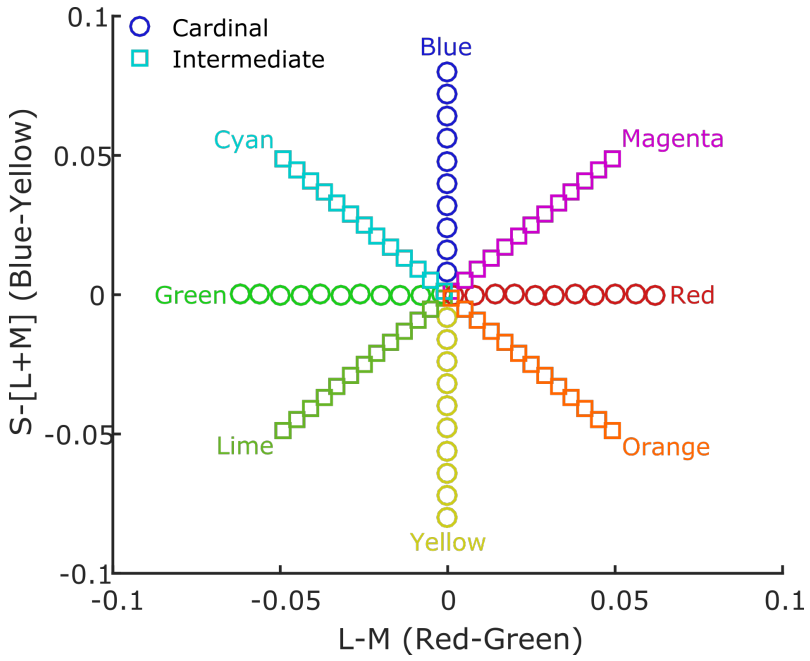


Figure 4.14: The axes of chromatic change used in the CCDT and iCCDT in DKL colour space.

4.5.1.4 Procedure and Data Analysis

All participants returned to the laboratory to complete the iCCDT. The iCCDT procedure was identical to the CCDT procedure detailed in Experiment 3.

Thresholds for the iCCDT were calculated from the mean of the last two reversals from each staircase, the same as in the CCDT, using a look up table to map nominal staircase values to $\Delta E_{u^*v^*}$ values between measured stimulus chromaticities. To predict iCCDT thresholds from CCDT data we used the same algorithm that was used to predict IDT thresholds from CCDT data in Experiment 3.

4.5.2 Results

Pooling data from the CDDT and iCCDT, there is a significant main effect of chromatic direction of change on thresholds ($F(3.77, 71.71) = 21.06$, $p < .001$, with a Greenhouse-Geisser correction; Figure 4.15). Orange and magenta thresholds did not differ significantly from one another, but both were significantly lower than thresholds for all other chromatic

directions of change ($p < .034$, in all cases). In addition, thresholds for the yellower direction of change were significantly higher than for the redder, lime, and cyan directions ($p < .002$ in all cases).

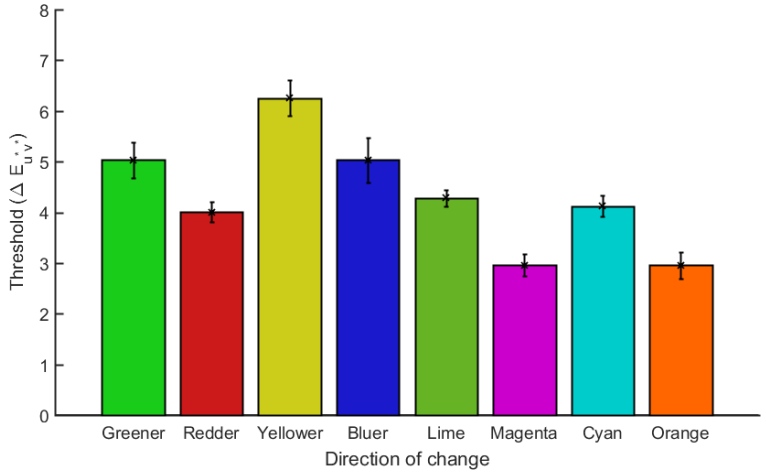


Figure 4.15: CCDT and iCCDT thresholds along the different chromatic axes of change. Error bars show ± 1 SEM.

Our motivation for this experiment was to assess whether CCDT thresholds can be used to predict iCCDT thresholds. Before we use the modelling procedure to investigate this, we can look at correlations between the thresholds for the different directions of change. The assumption we are testing is that chromatic contrast detection along the intermediate axes in the iCCDT is governed by the opponent colour mechanisms described by the cardinal axes of DKL colour space used in the original CCDT task. Thus, we hypothesise that chromatic contrast detection along each intermediate half-axis is mediated by chromatic contrast detection along two cardinal half-axes, namely, green and yellow for the lime half-axis, red and blue for the magenta half-axis, red and yellow for the orange half-axis, and green and blue for the cyan half-axis. Indeed, if we compare thresholds for the intermediate half axes to average thresholds over the two cardinal half-axes that we hypothesise are responsible for detecting such chromatic contrast changes (average predictions), we find that intermediate half axes thresholds are always over-predicted. Intermediate half-axis thresholds are always significantly lower than the average thresholds along the two cardinal half-axes that it falls between (Figure 4.16).

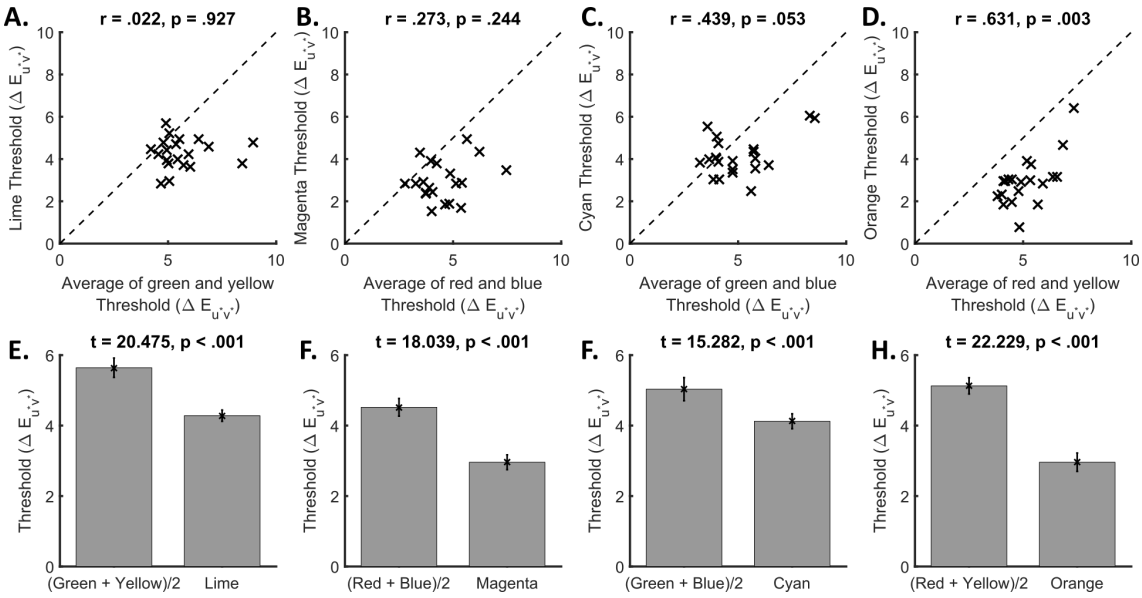


Figure 4.16: Comparing thresholds along intermediate half-axes to thresholds along the two cardinal half-axes that it falls between. A-D. Scatter plots between intermediate half-axis thresholds and average thresholds for the two half-axis hypothesised to mediate detection along said half-axis. Pearson's correlation coefficient between the two variables is shown above each plot. E-H. Average thresholds for the intermediate half-axes and over the two corresponding cardinal half-axes. We conducted paired t-tests between each pair of variables, shown above each bar plot. Error bars show ± 1 SEM.

These results, lower thresholds along intermediate half-axes than the two corresponding cardinal half-axes, is further motivation to model intermediate half-axis thresholds according to the modelling procedure used in Experiment 3. A reduction in thresholds along intermediate half-axes compared to the cardinal half-axes is supportive of our hypothesis that chromatic contrast detection along intermediate half-axes is governed by two cardinal half-axis mechanisms. If two mechanisms are responsible for detecting changes along the intermediate half-axes, with probabilities of detecting the stimulus equal to p_1 and p_2, respectively, then the combined probability of detecting the change with either mechanism is

$$p_{combined} = p_1 + p_2 - p_1 p_2. \tag{4.4}$$

Note that as $0 \leq p_1, p_2 \leq 1$, $p_{combined} \geq p_1, p_2$, and so it must be the case that detection thresholds reduce along the intermediate half-axes relative to the two corresponding cardinal half-axes.

When we employ the modelling technique detailed in Experiment 3 we predict thresholds (probability predictions) for the iCCDT that are closer in magnitude to the measured thresholds than those predicted from the average thresholds along the two corresponding cardinal half-axes (Figure 4.17.A). However, predicted thresholds for each chromatic direction of change still differ significantly for each half-axis, although now predicted thresholds are only significantly higher for the magenta and orange half-axes and significantly lower for the lime and cyan half-axes ($t = 2.46$, $p = .024$, $t = 3.83$, $p = .001$, $t = -3.49$, $p = .003$, and $t = -3.74$, $p = .001$, respectively. Note that p - values are not corrected for multiple comparisons here. If we were to accept significance at the $\alpha = .05/4 = .013$ level there is no difference between thresholds for the lime half-axis). This result can be visualised in a scatter plot where is it clear that the model over-predicts thresholds for the magenta and orange half-axes but under-predicts thresholds for the lime and cyan half-axes (Figure 4.17.B).

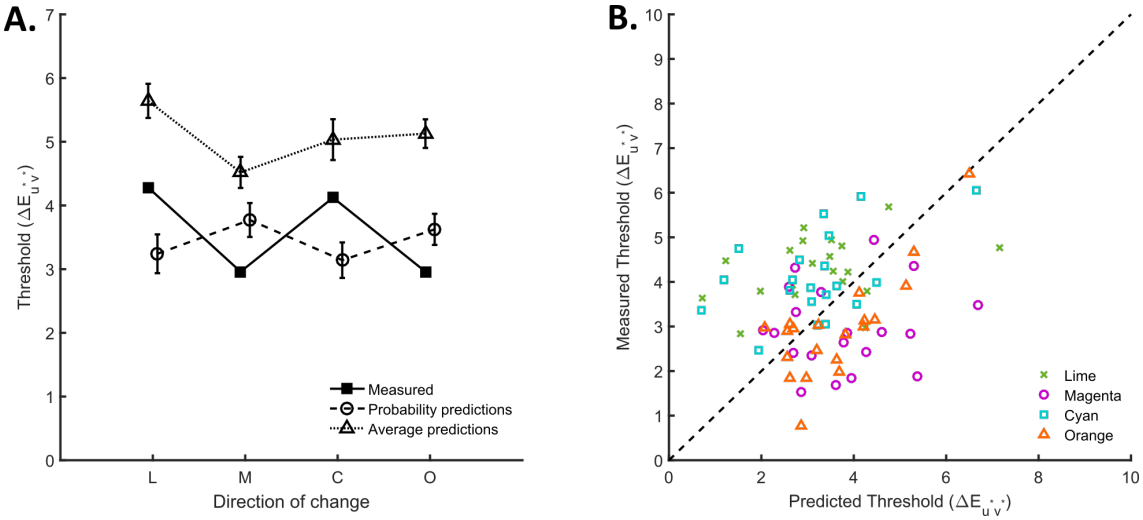


Figure 4.17: Predicting iCCDT thresholds from chromatic contrast detection data along the cardinal axes of change used in the CCDT. A. Comparing predicted and measured iCCDT thresholds. The probability predictions are the predictions made according to the modelling procedure in Chapter 4. The average predictions are the predictions made according to average thresholds along the two corresponding cardinal half-axes. B. A scatter plot of measured against probability predicted iCCDT thresholds for each intermediate half-axis. Error bars show ± 1 SEM.

In addition, the relative ordering of thresholds is not predicted by the model (Pearson's correlation between measured and predicted thresholds for each half-axis: lime $r = .342$,

$p = .140$, magenta $r = .088$, $p = .711$, cyan $r = .479$, $p = .033$, and orange $r = .639$, $p = .003$). While there is a consistent pattern in the under and over prediction of thresholds for the different half-axes, the model fails to predict the inter-individual differences in performance. This is also true of the average predictions except for the orange half axes where the average of thresholds along the red and yellow half axes explains a significant proportion of the variation in the data ($R^2 = .398$, 39.8% of the variation in orange thresholds is explained by variation in the average thresholds). However, it could be the case that inter-individual variability on the CCDT (and iCCDT) is not greater than the level of intra-individual variability on the task. As part of a separate study, 10 observers repeated the CCDT on three separate occasions. The average standard deviation of their thresholds was 1.67 ± 2.19, 2.46 ± 3.65, 2.05 ± 2.05, and 1.76 ± 1.59 $\Delta E_{u^*v^*}$ for the green, red, yellow, and blue half-axes, respectively. In comparison, the standard deviations in thresholds for the corresponding four half-axes for the 20 observers in this experiment are 1.58, 0.89, 1.57, and 1.99 $\Delta E_{u^*v^*}$, respectively. Thus, we see higher levels of intra-individual variability than inter-individual variability in the CCDT, except for the blue half-axis. Another way to quantify the level of intra-individual variability is to correlate thresholds for the different directions of change across repeated runs of the task (Figure 4.18). We only find evidence of inter-individual variation for the blue half axis.

4.5.3 Interim Discussion

We modified the CCDT such that the chromaticity of the arrow was modulated along two intermediate axes in an equiluminant DKL plane (the iCCDT). We hypothesised that chromatic contrast detection along these axes is governed by the cardinal chromatic mechanisms - the cone-opponent, red-green ($L - M$) and blue-yellow ($S - (L + M)$) axes of DKL colour space.

Thresholds for each intermediate half-axis were always significantly smaller than the average of those for each pair of cardinal half-axes (average predictions) hypothesised to be responsible for chromatic contrast detection in the relevant quadrant of the equiluminant

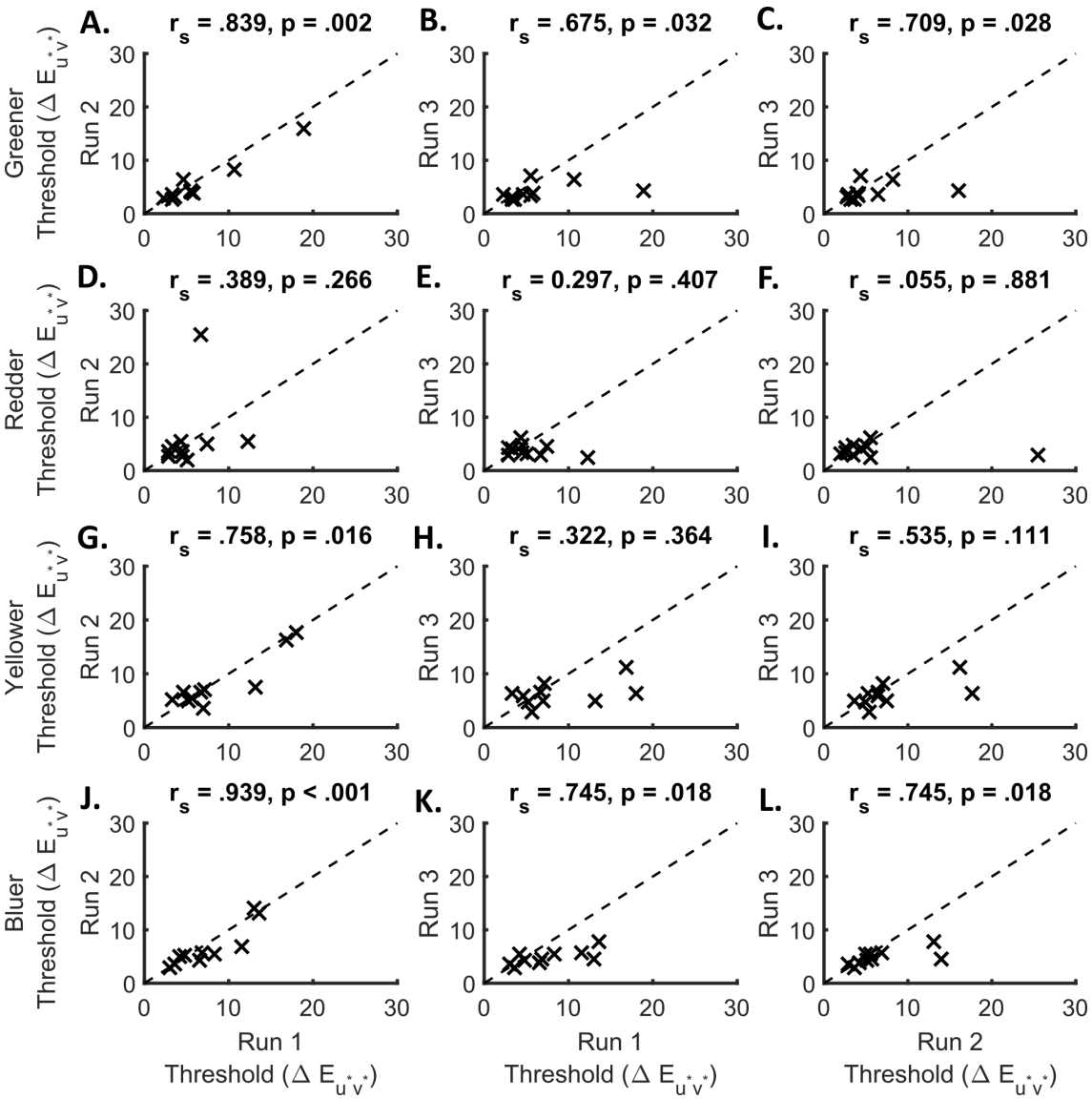


Figure 4.18: Correlations between thresholds for the different directions of change across multiple runs of the CCDT. Spearman's correlation coefficient between the two variables is shown above each plot.

DKL plane. This fits with our assumption that the cardinal cone-opponent mechanisms act in tandem to detect chromatic contrast changes in a stimulus. In addition, we employed the modelling procedure detailed in Experiment 3 to predict thresholds along the intermediate axes from chromatic contrast detection data along the cardinal axes (probability predictions). On average, the probability predictions are of the same magnitude as the measured threshold values (unlike the average predictions which are significantly

higher). However, we find systematic deviations in our probability predictions that are half-axis dependent. Thresholds for the lime and cyan half-axes are under predicted while thresholds for the magenta and orange thresholds are over predicted.

The commonality between the lime and cyan half-axes is the green ($M - L$) half-axis while the commonality between the magenta and orange half-axis is the red ($L - M$) half-axis. As thresholds are under predicted for the lime and cyan half-axes but over predicted for the magenta and orange half-axes, it seems that chromatic contrast sensitivity along the red-green ($L - M$) chromatic cardinal mechanism is modulated when the blue-yellow ($S - (L + M)$) half-axes are also stimulated. This suggests that when the blue-yellow channel is also active, sensitivity to increments along the red-green axes is decreased while sensitivity to decrements is increased.

Our original goal was to establish if the cardinal axes of DKL colour space may be considered as accurate and independent descriptions of second-order colour mechanisms and if detection of stimuli within different quadrants of DKL colour space can be predicted from detection data along these axes. We find evidence that the cardinal axes either do not act independently (they may show dependent adaptation, see Section 1.4.1) or that there are higher-order mechanisms (see Section 1.4.2). In this case, we must be cautious about how much weight we assign to our conclusions from Experiment 3.

4.6 Experiment 5: An ISETBIO model

In Experiments 1-4, we explored inter-individual differences in illumination discrimination thresholds obtained using the IDT. By taking a look back at the data presented in Chapter 3, we showed that inter-individual differences in illumination discrimination thresholds are not only consistent for different chromatic directions of change in the illumination but also for different reference illumination conditions. In addition, we collected a larger data set from a group of participants who also completed a task that established chromatic contrast detection thresholds (the CCDT). Using this data set, we showed that there is

large variability in illumination discrimination ability and that a moderate amount of the variability is explained by differences in chromatic contrast detection ability; however, we fail to predict the "blue bias". We investigated the validity of our extension of performance on the CCDT to performance on the IDT. We concluded that our predictions must be interpreted with caution, as we failed to accurately predict thresholds for intermediate axes in the CCDT. In this Experiment, we use a different approach to modelling inter-individual differences in performance on the IDT. A possible source of variability in human trichromatic colour processing mechanisms is differences in $L:M$ cone ratios. This source of variability occurs at the initial stage of visual processing and may be a limiting factor in chromatic detection and discrimination ability at later stages. Here, we model the initial stages of visual processing to assess what impact, if any, differences in $L:M$ cone rations may have on illumination discrimination ability in the IDT. We can also consider this an analysis of whether the statistics of the distal stimulus or the limiting factors of the human optical system account for the asymmetries in IDT thresholds.

A classic way to approach this problem is to develop an ideal observer. Ideal observer analysis is often used in the analysis of psychophysical data to establish an upper limit on performance by designing an algorithm for optimal performance of the task given the information available on each trial (Geisler, 2011). We take this approach, forming our ideal observer at the level of the photoreceptor mosaic after simulating the passage of light through the human optics, accounting for optical effects such as chromatic aberration and noise during phototransduction. The ideal observer vectorises matrices of isomerisations that represent the photon catches of each cone in a pre-specified cone mosaic in response to the reference and two comparison illuminations in the IDT. By calculating the vector norm between the vectorised isomerisations for the reference and two comparison stimuli, the ideal observer decides which comparison illumination is the best match to the target. Finally, we assign a lapse rate to the ideal observer such that they respond randomly on a set percentage of trials. Hence, our ideal observer is only optimal up to the point allowed by the human optics, noise at the level of the photoreceptor mosaic and attentional lapses.

We develop five categories of ideal observers, each category having one of five cone density triplets taken from measurements of cone densities in trichromatic human observers (Hofer, Carroll, Neitz, Neitz & Williams, 2005). Within each category, there are three ideal observers who share a cone density triplet but are distinguished by their cone mosaics. We simulated the performance of each ideal observer on the five different versions of the IDT used in Chapter 3 three times each (five reference illumination conditions: neutral, blue, yellow, red, and green).

4.6.1 Methods

4.6.1.1 Using ISETBIO to simulate photoreceptor isomerisations for an IDT scene

The image system engineering toolbox for biology (ISETBIO) is a freely available MATLAB toolbox that can be used to simulate the processing of light through the front end of biological visual systems (ISETBIO.org, Brainard, Jiang, Cottaris, Rieke, Chichilnisky, Farrell, Joyce & Wandell (2015)). In ISETBIO, an optical system (the eye), a sensor (the photoreceptor mosaic), and a scene are represented by different components. The toolbox provides functions to combine these components, simulating the response of each photoreceptor in a mosaic upon viewing a scene. The algorithms for combining the different components account for imperfections in human visual processing such as optical blurring, filtering by the lens and macular pigment, chromatic aberration and photoreceptor noise. We will use ISETBIO to simulate photoreceptor isomerisations, for specified cone mosaics, to different illumination views in the IDT.

First, we must form an ISETBIO scene structure (Figure 4.19 - Step 1). In Chapter 3, Section 3.2.7, we explained how we obtained a hyperspectral image of the Mondrian used in the IDT that we then used to find a representation of the spectral reflectance distribution across the part of the Mondrian that lined the back wall of the stimulus box - the reflectance image. Using the reflectance image, an intensity map (that represents the

illumination gradient across the scene and is found from measurements of the calibration tile), and a spectral radiance measure of an illumination used in the experiment (Figure 3.4), we can define an ISETBIO scene structure. Using different illumination spectra, we can define a scene that represents any given view of the back wall of the stimulus box during the experiment. As part of the ISETBIO scene structure, we also specify the viewing distance (81 *cm*) and the size of the stimulus. We will compute photoreceptor isomerisations for a cone mosaic that spans the whole field of view. For that reason, we restrict the field of view to a central square of the back wall. We repeated the simulations multiple times for different side lengths of the square. These were 38.75, 19.38, 9.69, 4.84, and 0.61 *cm* (25.57, 12.78, 6.39, 3.20, and 0.40 degrees of visual angle, Figure 4.20).

Secondly, we must define an optical system and its sensing device (Figure 4.19 - Step 2). We use the default human optics structure in ISETBIO that adopts the Marimont & Wandell (1994) shift-invariant, wavelength-dependent model of the human optics. Other default parameters specify the lens optical density function (taken from CVRL.org) and the pupil radius (1.5 *mm*). The sensor is a human trichromatic cone mosaic. We use the default human cone mosaic structure in ISTEBIO (rectangular packing) but vary the triplet of cone densities. We take five cone density triplets from Hofer *et al.* (2005) who used an adaptive optics system to image human cone mosaics. The five that we use come from observers HS, AP, MD, RS and BS (the five observer types, Table 4.2). The default cone spectral sensitivities are from Stockman & Sharpe (2000). Each cone covers an area of 2 μm in the mosaic and the size of the mosaic was set to the field of view for the scene. The arrangement of cones in the mosaic is determined at random while ensuring that the proportions are as specified. We generated three cone mosaics for each observer type, 15 observers in total.

Finally, we use ISETBIO routines to first combine the optics structure with the scene structure. This models the passage of light through the optical system to produce a retinal irradiance image. Then we combine the retinal irradiance image with the sensor structure to obtain a matrix of photoreceptor isomerisations. The integration time for

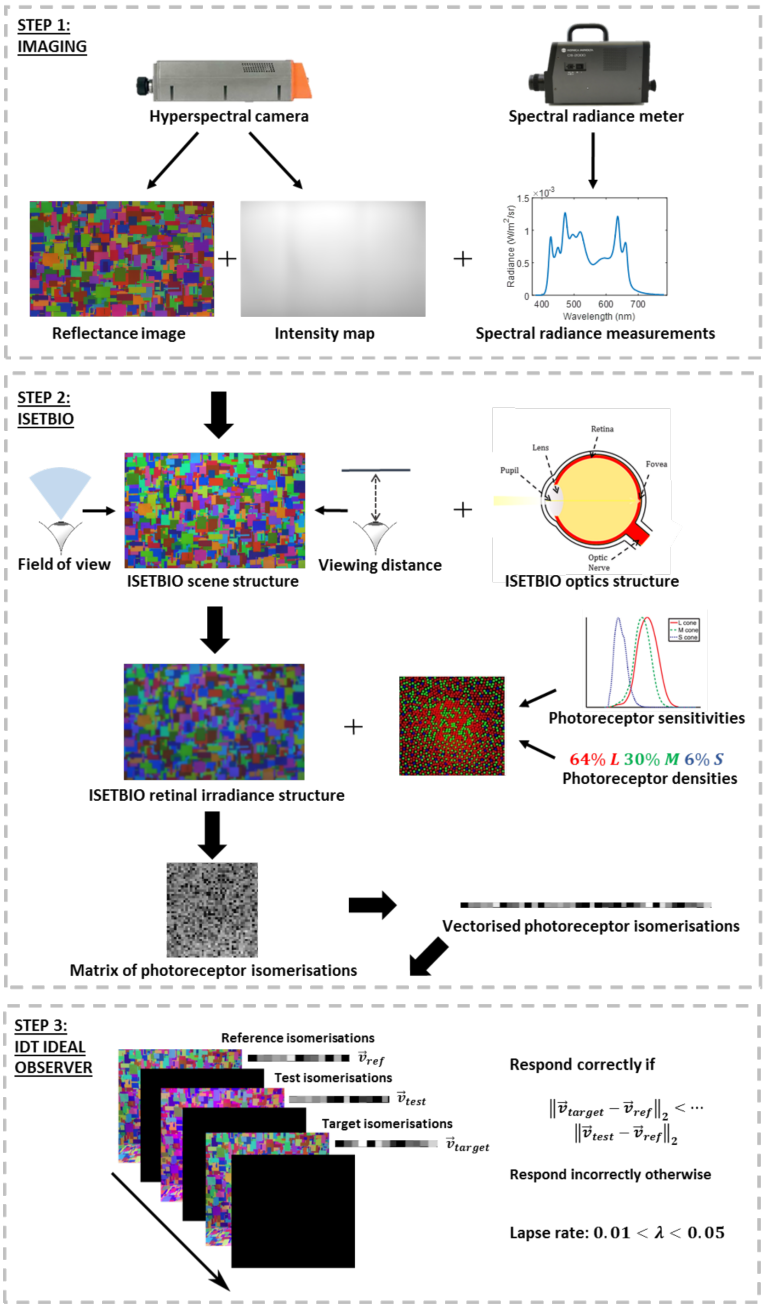


Figure 4.19: An overview of the ISETBIO modelling procedure. The different steps are explained in detail in the text.

each cone was set at 500 ms to control exposure time to the stimulus, including the scene under a reference illumination (exposure time is 2000 ms during the experiment). As isomerisation values are cumulative, this ensured that the isomerisation values for the reference illumination scene were in a similar range to those for the test and target

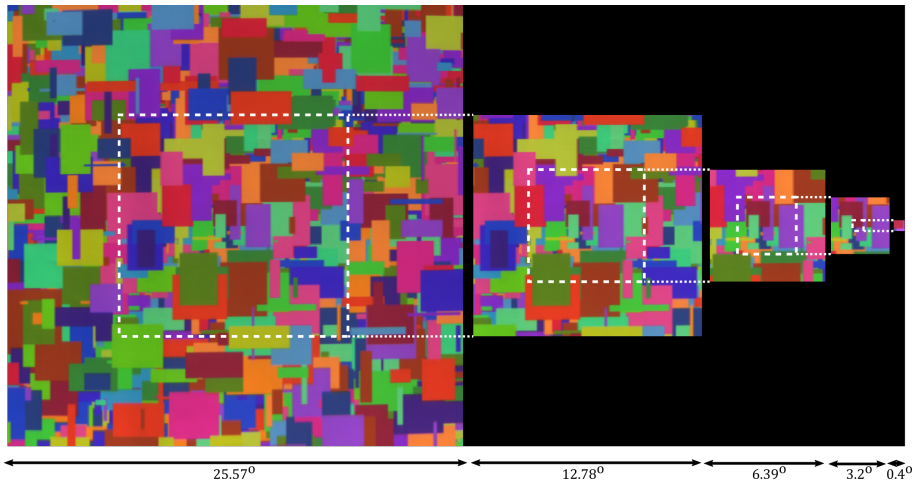


Figure 4.20: The stimuli used in the different simulations. The white dashed lines show which section of the scene was cut out to form the smaller fields of view.

Table 4.2: The cone density triplets used in the ideal observer models.

Observer	L-Cone (%)	M-Cone (%)	S-Cone (%)
HS	25.7	67.9	6.4
AP	52.4	42	5.6
MD	61.7	32.5	5.8
RS	67.2	28.2	4.6
BS	89.1	5.4	5.5

illumination scenes. Poisson noise is added to all isomerisation values with mean equal to the mean over all values.

4.6.1.2 Ideal observers at the level of the photoreceptor mosaic

For each of the 15 observers we created, we developed an ideal observer algorithm at the level of their photoreceptor isomerisations. The algorithm governs responses to each trial in the IDT. To simulate a response for a single trial, we compute two sets of isomerisations for 500 ms exposure to the reference/target illumination (they are equivalent) and one set of isomerisations for 500 ms exposure to the test illumination. We vectorise the isomerisation matrices to form $\vec{v}_{ref}$, $\vec{v}_{tar}$, and $\vec{v}_{test}$, the vectorised responses to the reference, target and

test illumination scenes, respectively. The observer responds correctly if:

$$||\vec{v}_{tar} - \vec{v}_{ref}||_2 < ||\vec{v}_{test} - \vec{v}_{ref}||_2, \tag{4.5}$$

otherwise, they respond incorrectly. Each observer is also assigned a lapse rate between 1% and 5%. On that percentage of trials, they respond randomly. By generating responses to individual trials in this way we are able to run the staircase procedure from the experiment to establish thresholds for each observer. We simulated thresholds for all 15 ideal observers for all four directions of chromatic change in all five reference illumination conditions for the experiment discussed in Chapter 3. There were three staircases for each axis of change and thresholds were calculated as the mean of the last two reversals from each staircase. Each simulation was repeated three times (three runs for each of the 15 observers).

4.6.1.3 Modelling adaptation effects

In ISETBIO there is an option to implement a biophysical model of the phototransduction cascade, accounting for adaptation of the cones. The model was proposed by Rieke and colleagues (Angueyra & Rieke, 2013; Rieke, 2014; Soo, Detwiler & Rieke, 2008) who specified a set of differential equations that would transform a stream of computed isomerisations (over a time interval) into a current. This is a computationally intensive task. For each cone in a mosaic, isomerisation counts must be computed at small time intervals (every millisecond in our calculations). The biophysical model component of ISETBIO takes this stream of isomerisation counts as input and outputs the computed current over time. We did not use this option in our ideal observer model due to computational requirements. However, to establish whether adaptation may cause us to make different conclusions from our model, we computed the current from single cones over the duration of an IDT trial as if they were observing a neutral surface (flat reflectance) illuminated by the experimental illuminations (Figure 4.21). We computed the current 5000 times for each sequence of trials (target in first or second comparison position) for a single cone of each type (L, M or S-Cone). We restricted the test illuminations to those 5, 10, 15 or 20

$\Delta E_{u^*v^*}$ away from each of the five reference illuminations along each chromatic direction of change. We simulated a 2 second dark adaptation period at the start of each trial. To establish whether an ideal observer that incorporated cone adaptation would predict a different pattern of thresholds in the IDT we compared the contrast between the test and target illuminations calculated from the current model to the contrast between raw isomerisations and, as a baseline comparison, to standard cone contrast values.

To calculate the standard cone contrast (sCC^j for $j = L, M, S$) between the target and test light, we first computed the L, M or S cone tristimulus values for both the target and test illuminations (V^j_{tar} and V^j_{test} for $j = L, M, S$, respectively) and then calculated standard cone contrast using the equation:

$$sCC^j = \frac{V^j_{test} - V^j_{tar}}{V^j_{tar}}. \tag{4.6}$$

To calculate the isomerisation cone contrast (iCC^j for $j = L, M, S$) between the target and test illumination, we first computed the mean isomerisations during the final 300 ms of the illumination presentation (Figure 4.21) for each of the 5000 simulations for each trial ordering. We dropped the first 200 ms of the simulated values to allow stabilisation of the current values. We then took the mean over all 5000 simulations for each trial ordering separately to get the mean isomerisation count for each cone type for both the target and test illuminations (I^j_{tar} and I^j_{test} for $j = L, M, S$, respectively). The isomerisation cone contrast was then calculated as:

$$iCC^j = \frac{I^j_{test} - I^j_{tar}}{I^j_{tar}}. \tag{4.7}$$

Finally, to calculate the current cone contrast (cCC^j for $j = L, M, S$) between the target and test illumination, we first computed the mean current during the final 300 ms of the illumination presentation (Figure 4.21) over the 5000 simulations for each trial ordering to form an average current trace during the final 300 ms ($\vec{c}_{tar}$ and $\vec{c}_{test}$). We took the minimum value from the average traces ($\min(\vec{c}_{tar}, \vec{c}_{test})$) and baseline corrected both

current traces to this value. We found the mean current for each of the target and test illuminations by averaging over all time points in the baseline corrected traces (C^j_{tar} and C^j_{test} for $j = L, M, S$, respectively). The current cone contrast was then calculated as:

$$cCC^j = \frac{C^j_{test} - C^j_{tar}}{C^j_{tar}}. \tag{4.8}$$

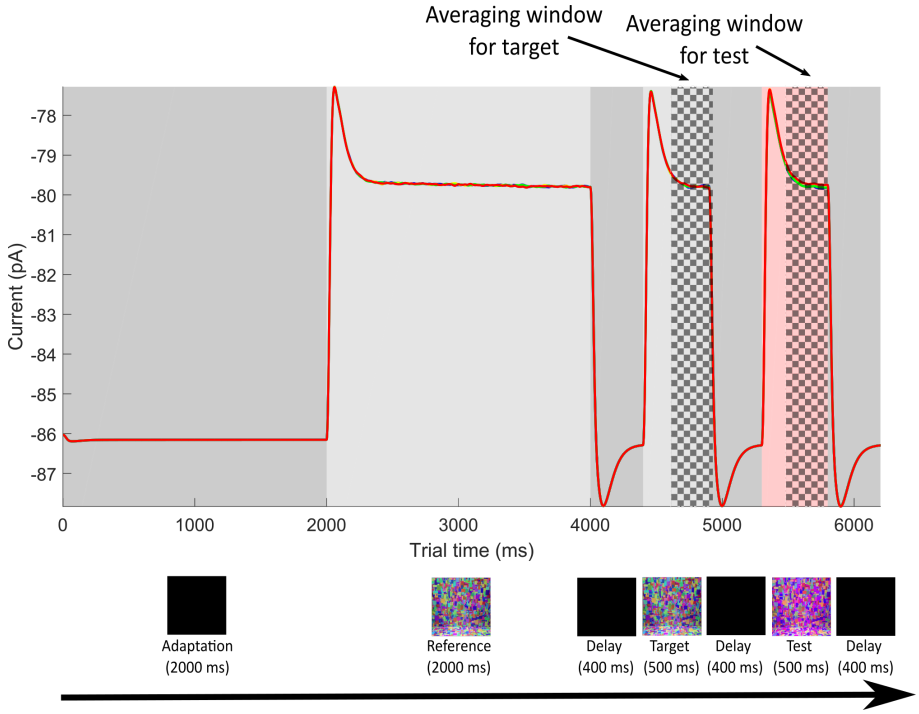


Figure 4.21: Modelling adaptation effects. Example L-cone current traces in the neutral reference illumination condition when the test illumination is 10 $\Delta E_{u^*v^*}$ away along each chromatic direction of change and the target is in the first comparison position. While the images below the plot are representative of the illumination presentation sequence during the trial, they do not represent what the single cone was viewing during the simulation. All simulations were of the experimental illuminations reflected from a hypothetical neutral surface with flat reflectance.

4.6.2 Results

4.6.2.1 Ideal observers

Simulated IDT thresholds do not match the behavioural data reported in Chapter 3 (Figure 4.22). Firstly, regardless of stimulus size (25.57, 12.78, 6.39, 3.20, and 0.40 degrees of visual

angle) simulated thresholds do not show the same pattern of bias as the behavioural data. Secondly, simulated performance of the ideal observers is generally better than of the observers tested in Chapter 3, and in some cases, the ideal observers produce the lowest possible threshold on the task (1 $\Delta E_{u^*v^*}$, effectively perfect performance). Performance of the ideal observers worsens (higher discrimination thresholds) as the size of the stimulus decreases. This is not as clear for the yellow reference illumination condition compared to the others, but this is also the condition in which the human observers display the lowest average thresholds.

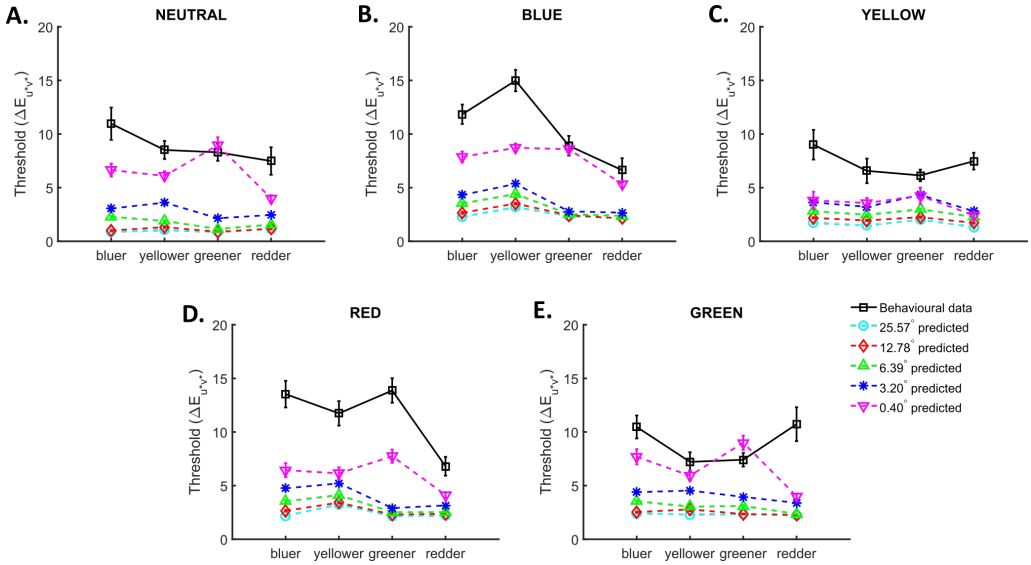


Figure 4.22: Behavioural and simulated IDT data. A-E: Behavioural and simulated thresholds for the four directions of chromatic change in each reference illumination condition. Error bars show ± 1 SEM.

We were also interested in whether incorporating different cone density triplets in our ideal observer models would predict inter-individual differences in performance on the IDT (Figures 4.23-4.27). Considering ideal observer performance at an individual level, we see that all five ideal observer types are at ceiling task performance in the neutral reference illumination condition when the stimulus size is 25.57° (Figure 4.23.A). However, performance is not at ceiling for the other reference illumination conditions with this stimulus size and we see individual differences beginning to emerge, especially in the blue and red reference illumination conditions (Figure 4.23.B and D). Ideal observers of type

BS consistently have the highest thresholds in these conditions while ideal observers of type HS consistently have the lowest.

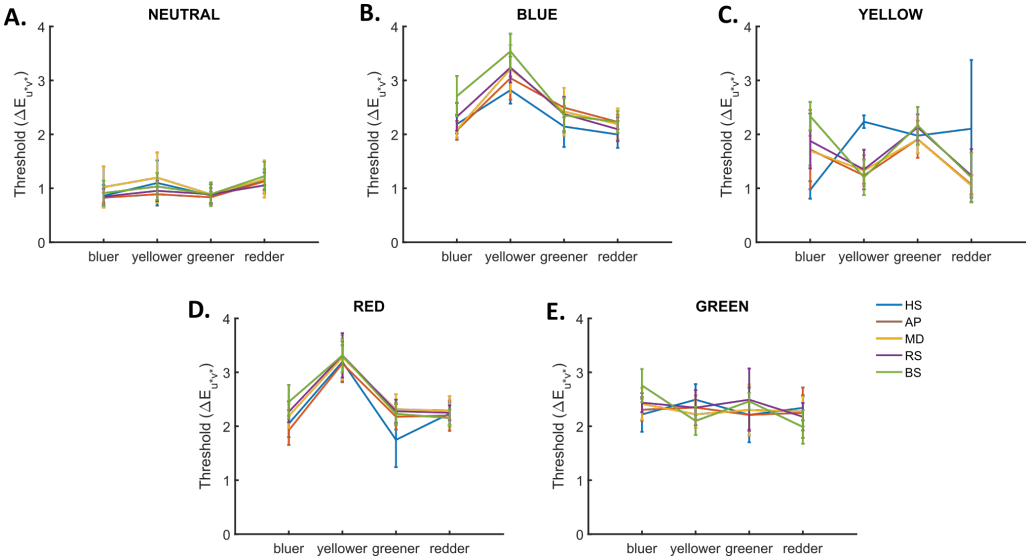


Figure 4.23: Simulated IDT data for a stimulus of size 25.57°. A-E: Individual simulated threshold traces for the four directions of chromatic change in each reference illumination condition for each of the five ideal observer types. For each observer type, an average is taken over the three cone mosaics and the three runs for each ideal observer. Error bars show ± 1 SEM.

As the stimulus size is decreased (Figures 4.24-4.27), we see the individual differences begin to emerge more. They are clearly visible by the smallest stimulus size condition (Figure 4.27) for all reference illumination conditions, although overall performance in the yellow reference illumination condition does not decline at the same rate. On average, HS observer types have the lowest thresholds, followed by AP types, MD types, and RS types, with BS observer types having the highest thresholds. The pattern in overall thresholds reflects the $L:M$ cone ratios of the five observer types. Higher $L:M$ cone ratios lead to higher than average thresholds (worse discrimination) in the IDT.

4.6.2.2 Adaptation effects

We computed current, isomerisation and standard cone contrast values (cCC, iCC, and sCC, respectively) between the target and test illuminations for four different test lights along each chromatic axis of change, for each reference illumination condition and for both

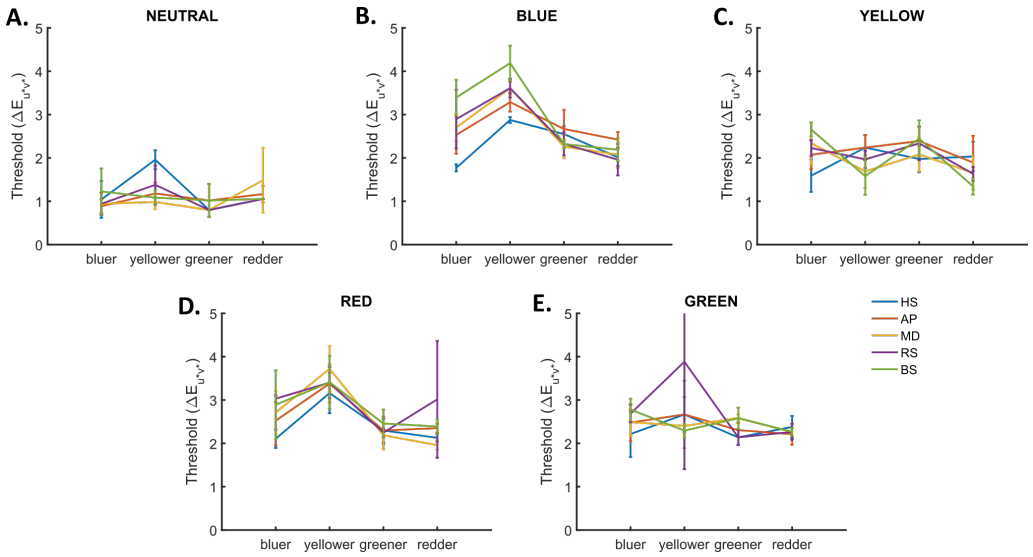


Figure 4.24: Simulated IDT data for a stimulus of size 12.78°. A-E: Individual simulated threshold traces for the four directions of chromatic change in each reference illumination condition for each of the five ideal observer types. For each observer type, an average is taken over the three cone mosaics and the three runs for each ideal observer. Error bars show ± 1 SEM.

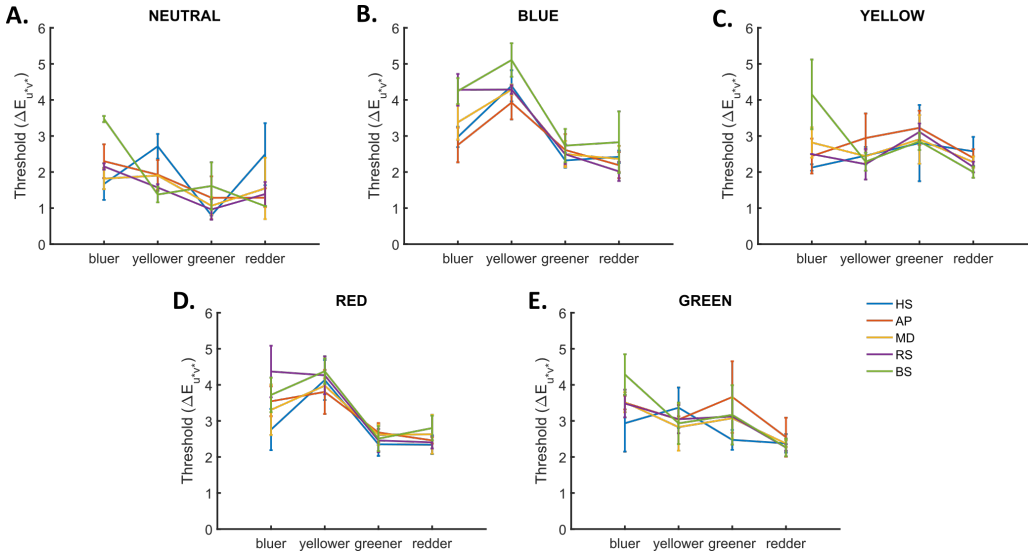


Figure 4.25: Simulated IDT data for a stimulus of size 6.39°. A-E: Individual simulated threshold traces for the four directions of chromatic change in each reference illumination condition for each of the five ideal observer types. For each observer type, an average is taken over the three cone mosaics and the three runs for each ideal observer. Error bars show ± 1 SEM.

possible orderings (target first or second) of the comparison illuminations on a trial. Here, we show the contrast values only for the neutral reference illumination condition when the target was in the first comparison position (Figure 4.28), all other results are in the

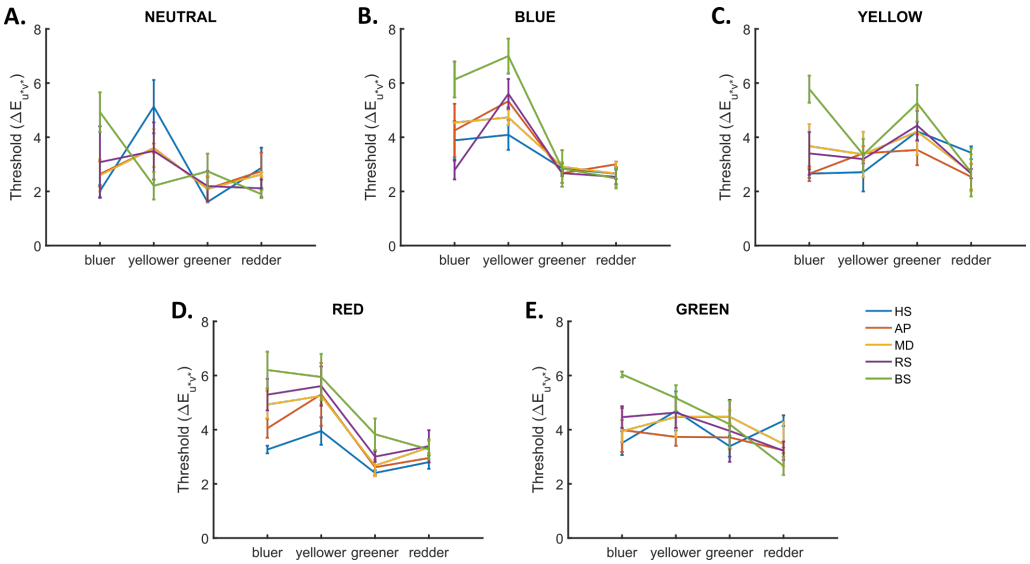


Figure 4.26: Simulated IDT data for a stimulus of size 3.20°. A-E: Individual simulated threshold traces for the four directions of chromatic change in each reference illumination condition for each of the five ideal observer types. For each observer type, an average is taken over the three cone mosaics and the three runs for each ideal observer. Error bars show ± 1 SEM.

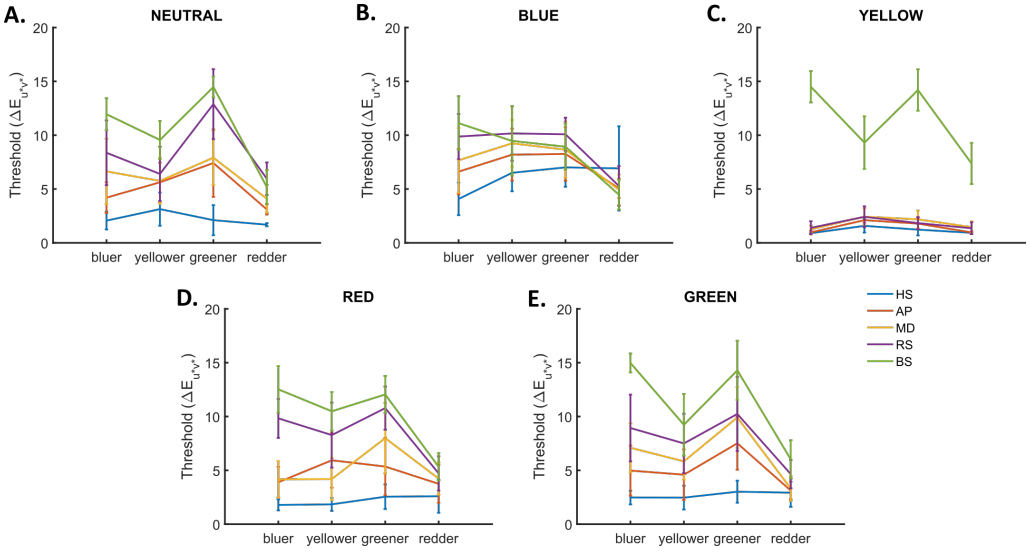


Figure 4.27: Simulated IDT data for a stimulus of size 0.40°. A-E: Individual simulated threshold traces for the four directions of chromatic change in each reference illumination condition for each of the five ideal observer types. For each observer type, an average is taken over the three cone mosaics and the three runs for each ideal observer. Error bars show ± 1 SEM.

appendix (Figures B.4-B.12). As the three types of values are of a different magnitude all contrast values were normalised to the 20 $\Delta E_{u^*v^*}$ contrast value separately for each type of contrast measure. The first point of note is that the relative change in iCC values

follows that of sCC values well. This serves as a good sanity check, as computed L,M, and S tristimulus values from the cone fundamentals should be representative of the number of photon catches by each cone type, albeit on a different scale. Secondly, in the large majority of cases, cCC values do not tell a different story to iCC values, suggesting that adaptation does not affect discriminability of the test and target illuminations and hence that adaptation cannot explain the biases in the behavioural data. However, if there are differences between the between cCC and iCC values, these tend to occur in the S-cone contrast values; in particular for the yellower direction of chromatic change regardless of reference illumination condition and for the greener direction of chromatic change in the red, yellow and green reference illumination conditions, although this observation is less clear than the former. The difference, in all cases, is that the current values predict larger contrast (better discriminability) in S-cone contrasts in these conditions than the isomerisations do. It is not the case, however, that thresholds for yellower chromatic changes in the illumination are consistently lower (better discrimination) in our behavioural data; the same is true for greener chromatic changes in the illumination. In fact, in the behavioural data, thresholds for redder chromatic changes in the illumination are lowest overall. A final feature of these simulations is that the order of the comparison illuminations (target first or second) does not affect the contrast values. In summary, it seems we can rule out the adaptation effect as an explanation of the biases in the behavioural data.

4.7 Discussion

We investigated inter and intra-individual differences in illumination discrimination ability. To do so, we first took a second look at the data presented in Chapter 3, this time at an individual level. These data revealed inter-individual differences in illumination discrimination thresholds that were consistent across the different directions of change and different reference illumination conditions. In addition, we collected two new data sets, one in which a group of observers completed the original (neutral) version of the IDT

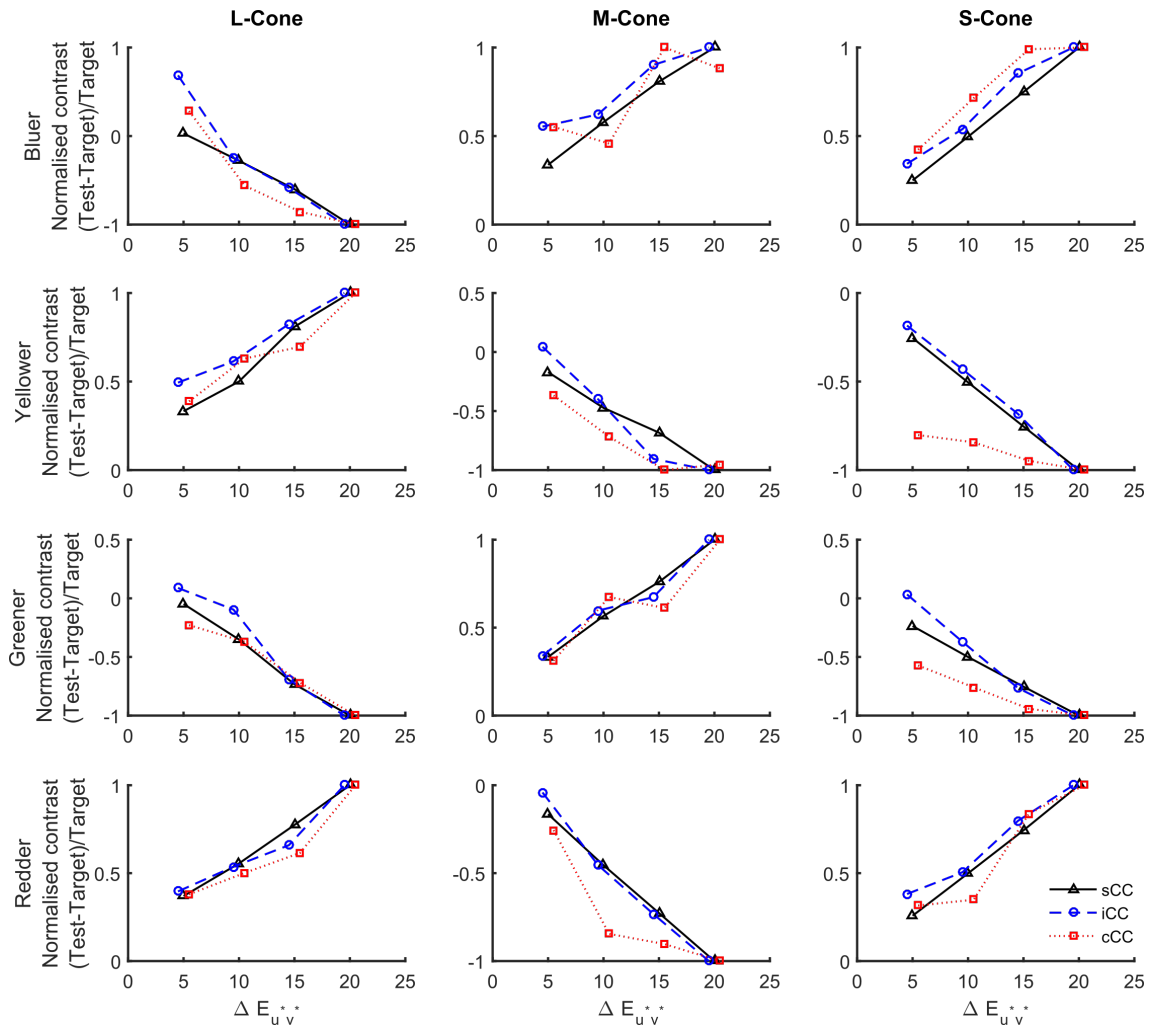


Figure 4.28: Simulated current (cCC), isomerisation (iCC) and standard (sCC) cone contrast values for the neutral reference illumination condition when the target is in the first comparison position. Each column of plots shows the contrast values for a different cone type and each row of plots for a different direction of chromatic change in the illumination. Along the horizontal axis is plotted the distance in CIELUV between the test and target illuminations. The vertical axes show the normalised contrast values.

three times and one in which the observers completed the original (neutral) IDT and a test of chromatic contrast detection ability (the CCDT). The former of these two data sets allowed us to show that while we do see intra-individual differences across multiple runs of the IDT, these differences are similar in all observers and appear to be due to a learning effect. The latter data set illustrated that the "blue bias" in illumination discrimination thresholds is not explained by chromatic contrast detection thresholds. That said, some of the inter-individual differences in illumination discrimination ability are captured by

differences in chromatic contrast detection ability.

In Experiment 1, we found evidence for inter-individual differences in illumination discrimination ability and introduced a metric to capture the amount of "blue bias" displayed by each observer. Our motivation for introducing such a metric was to establish if the "blue bias" in illumination discrimination ability was the same or different in all observers. If the "blue bias" was similar in all observers, this would lead us to conclude that the bias is brought about by a reduced physiological sensitivity to bluer chromatic changes in the illumination, possibly established through evolution to optimise colour constancy mechanisms for natural illumination changes (the nature hypothesis). However, we find that the "blue bias" is not the same in all observers, some displaying a stronger "blue bias" than others. In addition, we find that while the "blue bias" is present on average across all reference illumination conditions, inter-individual differences in the level of "blue bias" are not preserved across the different reference illumination conditions. This is suggestive of multiple colour constancy mechanisms, both a reduced physiological sensitivity to global bluer chromatic changes in a scene and the influence of a learnt illumination prior for daylight chromaticities that differs across individuals (both hypotheses: nature and nurture). A reduction in sensitivity to bluer chromatic changes in a scene predicts a "blue bias" in all reference illumination conditions. Conversely, a learnt illumination prior for daylight chromaticities predicts inter-individual differences in the level of "blue bias" (depending on the parameters of the prior). In addition, an illumination prior mechanism would differentially affect thresholds for the different reference illumination conditions. Rather than expecting the level of "blue bias" to be preserved across different reference illumination conditions, we may instead expect the level of bias in the chromatic direction opposite to the chromatic bias in the scene illuminations to be preserved across reference illumination conditions; a property that would be associated to the strength of the prior or to the amount of weight that an observer's visual system assigns to it during computations. However, this is not a feature of our data (Figure B.1).

In Experiment 2, we showed that the level of "blue bias" displayed by a given observer

is consistent across multiple runs of the task. In addition, observers with a higher "blue bias" maintain a higher level of "blue bias" across multiple runs of the task relative to other observers and vice versa. This is evidence of a mechanism that governs illumination discrimination ability (a colour constancy mechanism) that is present in all observers but differs between them, supportive of the idea of a learnt illumination prior that would vary with experience rather than evolutionary colour constancy mechanisms.

Experiment 2 also highlights a learning effect on the IDT, overall illumination discrimination thresholds decreasing from the first to second, but not second to third, repeats of the task. In Appendix A, we show that while the CIE 3-component model of daylights explains an average of 64.95% of the variation in the spectra of the illuminations used in the IDT (75.12% for the neutral condition that we used in the current experiment), a moderate percentage of the variation in the experimental spectra is not explained. This suggests that while the use of an illumination prior formed over natural daylight illuminations may lead to improved colour constancy in the IDT through improved illumination estimates, this is not the optimal prior for the task. If observers were to begin to learn the statistics of the illuminations that are used in the IDT, forming a prior specific for the task, then we may expect the opposite results to what we see - colour constancy should improve leading to increased discrimination thresholds on subsequent runs of the task. The fact that we see decreased discrimination thresholds (worse colour constancy) on subsequent runs suggests that observers are not learning the illumination statistics, but are learning a strategy for performing the task. Indeed, the aim and instructions for the task are to correctly discern the target from the test illumination. Remaining colour constant during the task does not lead to optimal task performance. In fact, to perform optimally, colour constancy should be overcome. We will return to this point in the general discussion.

Having shown in Experiments 1 and 2 that there are inter-individual differences in performance on the IDT, we asked in Experiment 3 whether these differences are explained at the level of chromatic contrast detection ability. In particular, such a comparison allows us to address the question of whether the "blue bias" in IDT thresholds is explained by

a reduced sensitivity to S-cone increments (bluer chromatic changes). Firstly, we do not see the same pattern of thresholds across the chromatic directions of change in the CCDT as we do in the IDT - there is no "blue bias". In addition, we do not see consistent correlations between the corresponding chromatic directions in the two tasks (plots across the diagonal in Figure 4.10). However, we acknowledge that there are differences between the chromatic axes of change used in the IDT and CCDT. The axes of change used in the CCDT are specified to fall along the cardinal axes of DKL colour space, generally assumed to represent the second stage of visual processing. By making the assumptions that these axes accurately and exhaustively represent the second stage of visual processing and that they act independently, we used the CCDT thresholds to predict thresholds on the IDT. We find that while CCDT thresholds can explain an average of 24.10% of the variation in the IDT thresholds (averaged over chromatic directions of change), they do not predict the "blue bias". In fact, the CCDT thresholds predict more of the variation along the redder and greener directions of chromatic change in the IDT than they do for the bluer and yellower directions of change (32.49% compared to 19.32%). This suggests that illumination discrimination thresholds along these two pairs of axes are influenced differently by visual mechanisms. As the bluer and yellower illuminations fall along the daylight axis, representing the chromaticities of natural daylight illuminations, and as chromatic contrast detection thresholds do not show a similar bias, we take the results of Experiment 3 as further evidence for the influence of an illumination prior in IDT performance.

There are multiple reasons, however, why the comparison between IDT and CCDT discrimination ability may not be valid. Firstly, there is disagreement in the colour vision community about the number of chromatic mechanisms that exist in the visual system, whether they act independently, and what the weightings of the different cone types are in the cardinal mechanisms (see Section 1.4). In order to make predictions of IDT thresholds from CCDT data, we made the assumption that there are only two second stage mechanisms, that they are well-represented by the cardinal axes of DKL space, and that they act independently to detect chromatic changes in a stimulus. Secondly, there are several

important differences between the two tasks. The CCDT is a contrast detection task where the participant must detect the appearance of a target against a background. The participant is not required to make comparisons temporally across time as they are in the IDT, a task that involves a memory component. In the IDT, the participant must hold a representation of the reference and two comparison lights in memory in order to make a decision about which comparison was most similar to the reference. These issues mean that the predictions we make must be interpreted with caution. However, the fact that we do not see a "blue bias" in the CCDT data but do in the IDT data is still strong evidence that the bias is not caused solely by a reduction in sensitivity to S-cone increments. If this was the case we would expect to see a manifestation of this reduced sensitivity in the data from both tasks.

In Experiment 4, we assessed the validity of our predictions of IDT thresholds from CCDT data by making predictions of further CCDT thresholds along intermediate axes of change in DKL colour space (the iCCDT task). Neither our average or probability predictions of iCCDT thresholds were sufficiently accurate, suggesting a more complicated model of second-order visual mechanisms in required or at least of how the mechanisms work in tandem. The fact that lime and cyan half-axes thresholds are under predicted while magenta and orange half-axes thresholds are over predicted suggests that the two mechanisms interact, the sensitivity of the red-green axis altered when the blue-yellow axis is also stimulated. The results of this experiment are cause for further caution when interpreting the predictions of IDT thresholds made in Experiment 3.

Finally, in Experiment 5, we simulated IDT thresholds using ISETBIO to ask whether the statistics of the distal stimulus or the limiting factors of the human optical system account for the asymmetries in IDT thresholds we saw in Chapter 3, and if differences in $L : M$ cone ratios account for individual differences in performance on the task. The "blue bias" and other biases seen in Chapter 3 are not an obvious feature of the simulated data, suggesting that higher-level processes are responsible for the effect. However, the simulations using the different cone density triplets did differ, mimicking the individual

differences that we observed in the real data. We found that higher $L : M$ cone ratios predict worse performance (better colour constancy) in the IDT. This effect was much clearer for smaller scenes. Decreasing the scene size decreases the size of the cone mosaic and hence the number of samples obtained from the image. Therefore, decreasing the stimulus size can be equated to ramping up the level of noise in the data that is fed to the decision rule. Hence, while the ideal observer performance is near perfect for the largest stimulus size, this is a prediction that performance on the IDT would be perfect if observers had direct access to isomerisations across the cone mosaic. It is likely, however, that the decision is made at a higher-level in the visual hierarchy after multiple processing stages have taken effect, each introducing further noise. In this sense, it seems logical that the smaller the stimulus size is, the similar the magnitude of simulated thresholds to the real data, especially since we ignore the memory component of the task in our simulations.

There are factors that could explain inter-individual differences in performance on the IDT that are independent of colour processing mechanisms, for example, age and cognitive factors (due to the memory requirement of the IDT). Age could not account for the inter-individual differences in our data (Figure B.2). We did not perform any cognitive measures on our observers to assess differences in cognitive processes such as working memory that may affect performance on the task. The closet correlate of such a measure that we have is an attentional measure, the lapse rate collected in the IDT via the catch trials. Unsurprisingly, lapse rate correlated significantly with thresholds for all chromatic directions of change (Figure B.3). In future studies, it would be beneficial to assess participants who perform the IDT with a battery of cognitive tests to see whether such differences can explain differences in performance. It is unlikely, however, that such measures will fully explain inter-individual differences IDT performance. While we do see evidence of global inter-individual differences in illumination discrimination ability (increased or decreased thresholds across all directions of chromatic change), we also see inter-individual differences in the level of "blue bias" displayed by our observers. Global inter-individual differences may be explained by differences in performance on cognitive

tasks. Inter-individual differences in the level of "blue bias", however, are more suggestive of underlying differences in perceptual mechanisms.

Experiment 4 could be subject to the effects of self-selection bias as the observers who participated in this study were invited back from the previous experiment (Experiment 3). All observers were invited to return and take part in Experiment 4, however, only the first 20 respondents were used for the study. It could be that the first 20 respondents were more motivated to take part in the study and hence that their responses in the task were more accurate that than the responses of others would have been. However, firstly, one could argue that all psychology studies where participants volunteer or sign-up to take part in a study are subject to the same issue, as only the most motivated will apply to participate. Secondly, the analyses for this experiment are within subject, therefore, if these participants are more motivated to take part in the study they will have displayed the same level of motivation when completing both the CCDT and iCCDT.

Chapter 5

What #theDress reveals about the role of illumination priors in colour perception and colour constancy

This Chapter has appeared almost in its entirety as a peer reviewed Journal of Vision publication:

Aston, S., & Hurlbert, A. C. (2017) What #theDress reveals about the role of illumination priors in color perception and color constancy. *Journal of Vision.*

5.1 Introduction

The division between people who named #theDress (the dress photograph that first appeared on social media in February 2015; Figure 5.1) "blue and black" vs. "white and gold" illustrates the subjectivity and individuality of colour perception. While there are many examples of illusions in which an individual observer sees a coloured object differently in different viewing conditions (e.g. colour contrast and colour assimilation; Brainard & Hurlbert (2015)), #thedress phenomenon differs from these in eliciting striking inter-individual differences under identical viewing conditions.

Figure 5.1: The photo of the dress taken from Wikipedia. Photo credit Cecilia Bleasdale.

When #thedress first appeared people immediately fell into two groups: one group reported a blue dress with black lace and the opposing group a white dress with gold lace.

As controlled studies have since shown, these naming differences are not due to different viewing devices (Gegenfurtner, Bloj & Toscani, 2015; Lafer-Sousa, Hermann & Conway, 2015). In addition, the stark division of people into only two naming groups seems to have occurred only due to the way the question was posed on social media. When participants were allowed to free-name the colours of the dress, a continuum of colour names emerged, but with three modal groups: blue and black (B/K), white and gold (W/G) and blue and gold (B/G) (Lafer-Sousa *et al.*, 2015).

Previous studies have also shown that observers differ not only in how they name the dress but also in the colours to which they match it (Chetverikov & Ivanchei, 2016; Gegenfurtner *et al.*, 2015; Lafer-Sousa *et al.*, 2015), leading to the conclusion that the phenomenon is a perceptual one and therefore not explained solely by differences in colour naming and/or colour categorization. However, previous studies differ somewhat in their matching results. In one study, where 53 laboratory participants made matches to the dress body and lace using a colour picker tool, matches to both regions differed between B/K and W/G observers in both lightness and chromaticity (Lafer-Sousa *et al.*, 2015). In another, Gegenfurtner *et al.* (2015) asked participants to make colour matches to both the dress body and lace as well as to select the best matching chip from the glossy version of the Munsell Book of Colour. Contrary to the results of Lafer-Sousa *et al.* (2015), both sets of matching data did not differ in chromaticity between the two groups of observers but did differ in luminance (or value in the case of the Munsell chips). A later online survey found differences in the matches for the B/K and W/G observer groups along all axes of CIELAB colour space (Chetverikov & Ivanchei, 2016). It is clear from these results that representing an individual's perception of the photograph only by the colour names they assign is not enough to capture the variability in perception across the population. Clearly, there are substantial individual differences in the processes of colour perception that are responsible for the phenomenon, but the detailed characteristics of these underlying factors remains somewhat elusive.

Brainard & Hurlbert (2015) suggested the colour constancy hypothesis to account for the

differences. The colour constancy hypothesis for #thedress phenomenon implies that individual differences in perception of the photograph arise because observers make different inferences about the incident illumination spectrum in the scene. The information that the photograph itself provides about the illumination spectrum is ambiguous, because the colours in the photograph could be produced by illuminating a blue dress with yellow light or white dress with blue light (see Figure 2 in Brainard & Hurlbert (2015)). If a difference in inferred illumination is responsible for the difference in perception then we must ask, what causes different observers to infer different illuminations? It might be that observers rely on previous experience - unconsciously embedded in the perceptual process as illumination priors - to determine the most likely illumination and overcome the uncertainty in the visual information.

If illumination priors do play a role in how the visual system calibrates itself for environmental illumination changes, then observers with different experiences of illumination changes may exhibit different perceptions. To test this reasoning, one needs a measure that predicts behavioural exposure to particular illuminations. Given that daylight illumination chromaticities are known to vary throughout the day (Hernández-Andrés *et al.*, 2001; Wyszecki & Stiles, 1967, see also Section 1.5.3) and traditional indoor lighting is much yellower than daylight (in particular for incandescent bulbs, see figure 2 in Webb (2006)), it is plausible that the observers illumination prior may be conveyed by the observers chronotype (Lafer-Sousa *et al.*, 2015), where chronotype refers to whether an individual is a morning or evening type (more colloquially, whether an individual is a lark or an owl). Morning types might be more likely to experience the blues of daylight illuminations while evening types might be more likely to experience the yellows of artificial illuminations.

The colour constancy explanation of the dress phenomenon is partially supported by the recent finding that the illumination inferred by observers in the photograph is negatively correlated with their dress colour matches (yellower illumination implies bluer dress match and vice versa) (Witzel, Racey & O'Regan, 2016). However, Witzel *et al.* (2016) conclude that differences in inferred illumination chromaticity are not due to illumination priors,

but rather because each observer implicitly estimates the illumination on a scene in an ad hoc manner. This interpretation implies that two observers who fall into different colour naming categories for the dress (e.g. blue/black vs. white/gold) will not display other distinct and predictable behavioural characteristics due to different inherent illumination priors. The results we present here allow for the possibility that illumination priors are responsible for differences in inferred illumination chromaticity, and hence bias dress colour perception, given that we find a correlation between chronotype and colour matching data. This correlation is weak and non-significant at the 5% level. On the other hand, we show that colour constancy thresholds obtained using the IDT are not related to dress colour perception, implying that generic biases in colour constancy do not explain variations in dress colour perception.

In addition, we show a dependence on luminance of subjective achromatic settings, which leads to the conclusion that subjective white point settings may be predictive of dress colour perception in the photograph but only if the settings are made at a luminance level that represents each individual's perceived level of brightness in the scene. Moreover, we present results that suggest perception and naming are disconnected, by showing that the colour names observers report for the dress photograph differ to those that they give to their dress matches when shown them in isolation, with the latter best capturing the variation in the matches.

5.2 Methods

5.2.1 Overview

Participants completed a set of computer-based matching tasks in which they provided colour appearance matches to the dress body and lace, matches to the perceived illumination in the image and achromatic settings for an isolated disk. After completing the matching tasks, participants were shown their matches to the dress body and lace and asked to name them (disk colour names). They were also asked to report (without re-

strictions) the colours that they named the dress the first time they saw it (dress colour names). Observers who had not seen the dress before participating in the study were shown the dress photo and asked to name the colours of the dress. Each participant also completed the morningness-eveningness questionnaire (MEQ; Horne & Ostberg, 1976) as a measure of chronotype. In addition, all participants completed the IDT to measure colour constancy thresholds. The study was conducted between the months of January and July 2016. (Note that hereafter, the phrase "the dress" refers to the image of the dress in the original photograph).

5.2.2 Participants

Participants were recruited from the Institute of Neuroscience (Newcastle University) volunteer pool, undergraduate courses and by word of mouth. Thirty-five participants were recruited in total. Two participants were unable to complete the matching task without the aid of an experimenter who used the Xbox controller to adjust the patch according to the participants instructions. Experimenter bias could not be ignored in these cases and the data from these participants were removed from the analyses. In addition, one participant gave identical matches to dress body and lace, neither of which matched the colour names they reported for their first view of the dress photo. It was assumed that this participant misunderstood the task and their data were also removed from the analyses. All data for the remaining 32 participants were included in the following analyses (20 female, mean age: 29.3 years, age range: 18.7 to 60.5 years).

All participants had normal or corrected to normal visual acuity and no colour vision deficiencies, assessed using Ishihara Colour Plates and the Farnsworth-Munsell 100 Hue Test. Each participant received cash compensation for their time.

5.2.3 Stimuli and apparatus

In the matching task (Figure 5.2), participants were instructed to match the colour appearance of the dress body and lace (one match each), by adjusting the colour of a disk

presented adjacent to the image of the dress. For each match, a black arrow appeared for three seconds indicating the region of the dress (body or lace) to be matched. Once the arrow disappeared, the matching disk became adjustable. Participants used an Xbox controller to adjust the disk's hue, chroma and luminance (in steps of 0.1 radians in hue, 1 in chroma and 2 cd/m^2 in luminance in HCL colour space). For the illumination matches (5 matches each), the disk was overlaid on the image and participants were instructed to adjust the disk to appear as if it were a white piece of card present in the scene. Lastly, two conditions of achromatic matches (5 of each type) were collected. For both types, a disk was presented in the center of an otherwise black screen and participants adjusted the disk's chromaticity until it appeared neutral; specifically, neither blue, yellow, red or green. For one condition, the luminance of the disk was fixed at 24 cd/m^2. For the other condition, luminance was fixed at the luminance of that particular participant's dress body match. Hence, for the achromatic settings, only hue and chroma were adjustable by the participant. Dress body and lace matches were always completed first and the order of all remaining matches (illumination and achromatic) was randomised. The chromaticity (and luminance for all but the achromatic matches) was set to a random value at the start of each trial. After all matches were complete, participants were shown two disks on the screen, on the left their dress body match and the right their dress lace match, and asked to name the colours of the disks. The matching disk was the same size on all trials (diameter 4.58 degrees of visual angle). The photograph of the dress was presented in 8-bit-colour and subsumed 20.41 by 26.99 degrees of visual angle. Stimuli were shown on a 10-bit ASUS Proart LCD screen and matches made with 10-bit resolution. A head rest was used in all tasks to ensure participants maintained a distance of 50 cm from this screen. The monitor was controlled using a 64-bit Windows machine, equipped with an NVIDIA Quadro K600 10-bit graphics card, running MATLAB scripts that used Psychtoolbox routines (Brainard, 1997; Kleiner *et al.*, 2007; Pelli, 1997). The stimuli were colorimetrically calibrated using a linearised calibration table based on measurements of the monitor primaries made with a Konica Minolta CS2000 spectroradiometer (Konica Minolta, Nieuwegein, Netherlands). Calibration checks were performed regularly throughout

the study period and the calibration table was updated when needed. All matches were converted to CIELUV for analysis according to the measured white point of the monitor having coordinates $Yxy = [180.23, 0.32, 0.33]$ in $CIEYxy$ colour space.

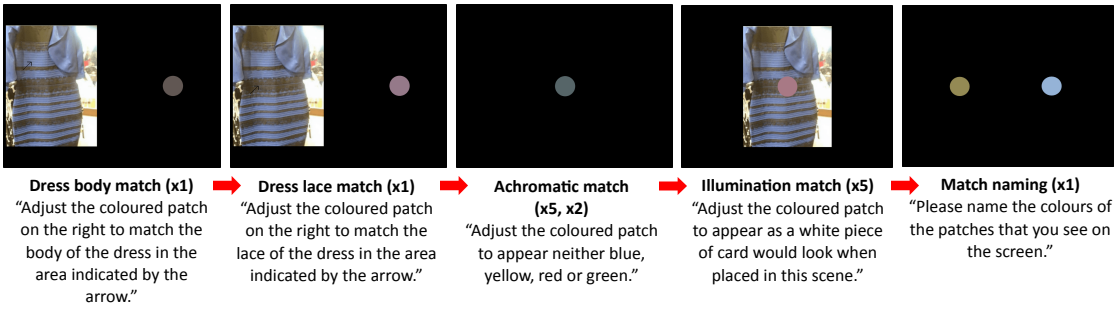


Figure 5.2: The matching task.

The different types of matches specified above are motivated by the following. First, the purpose of the achromatic matches at fixed luminance is to measure any overall bias in the observer's representation of neutral chromaticity at a luminance level that is held constant across observers. Second, collecting matches made at the luminance setting of each individual's dress body match has the purpose of assessing any specific bias due to this particular scene. As the variable that is manipulated here is the fixed luminance setting, it is assumed that bias specific to the scene may be caused by differences in perceived brightness. In both types of adjustments above, the whole scene, and hence the global image statistics, are not available to observers. Under the assumption that computations on global image statistics are necessary to infer the chromaticity and/or brightness of the incident illumination for a particular scene, observers make an illumination match in the context of the scene. In effect, all types of matches can be considered a form of achromatic adjustment. It is expected, however, that the different methods will yield different data sets.

Thresholds for colour constancy were obtained using the neutral (original) version of the illumination discrimination task described in Chapters 3 and 4.

The morningness-eveningness questionnaire (MEQ; Horne & Ostberg, 1976) consists of 19 multiple choice questions. The answers to each question are summed to form a total score,

ranging from 16 to 86. Chronotype categories are defined by score sub-ranges as follows: $16-30$ definite evening; $31-41$ moderate evening; $42-58$ intermediate; $59-69$ moderate morning; $70-86$ definite morning.

5.3 Results

5.3.0.1 Colour names for the dress differ from those assigned to the matched disk colours

The colour names that participants reported for the dress body and lace on first view divided into three groups: blue and black (B/K), white and gold (W/G) or blue and gold (B/G), in line with previous findings (see Appendix C for further details on categorisation). Each naming category has similar numbers of participants (B/K = 13, W/G = 11 and B/G = 8). Three observers had never seen the photo of the dress before participating in the experiment.

The colour names that participants gave to their disk colour matches for the body and lace divided into six groups: the original three groups, B/K, W/G and B/G, and an additional three, purple and gold (P/G), blue and green (B/Gr), and purple and blue (P/B) (see Appendix C for further details). The number of participants in each group is highly variable. B/G is now the dominant category, with most of the participants who originally named the dress B/G remaining in this group and more than a third of those who originally named the dress B/K and W/G switching to the B/G disk colour names group. Thus, participants assign different colour names to their matches, when shown them in isolation compared to the colour names they use for the dress itself.

In the following analyses (MANOVA and ANOVA analysis across disk colour name groups), only the four largest naming groups are considered, excluding the two individuals who named the discs B/Gr and B/P, in order to keep the analyses the same as for the dress colour names groups (MANOVA requires the number in each group to be at least the number of variables).

5.3.0.2 Matches to the dress lace but not dress body differ between dress colour naming groups in 3D colour space

To determine whether dress body and lace matches vary across the dress naming groups, we used a three-way MANOVA. This analysis differs from previous tests (Gegenfurtner *et al.*, 2015; Lafer-Sousa *et al.*, 2015; Witzel *et al.*, 2016) in allowing for assessment of differences in a three dimensional colour space (CIELUV) by combining the three colour co-ordinates (the three dependent variables) into a composite variable that best represents the differences in the centroids of the three matching groups (B/K, W/G, B/G, the independent variable). For comparison to earlier studies where differences are considered with respect to each colour co-ordinate separately, we include univariate ANOVA analyses of group differences in Appendix C.

Colour matches to the dress body did not differ significantly across the dress colour name groups (MANOVA with dependent variables L^*, u^* and v^*: $F(6, 56) = 1.95$, $\Gamma_p = .35$, $p = .089$), and so no composite variable representing the maximal difference between the group centroids was found. Conversely, colour matches to the dress lace show a significant multivariate difference across dress colour name groups ($F(6, 56) = 4.03$, $\Gamma_p = .60$, $p = .002$) (Figure 5.3.A-C). The MANOVA analysis yields the composite variable $d_1 = 0.48L^* + 1.31u^* - 0.43v^*$. The mean scores for d_1 differ significantly between the B/K and B/G categories ($p < .001$, Bonferroni corrected) and B/K and W/G categories ($p = .01$, Bonferroni corrected) (Figure 5.4.A), with lace matches for the B/G and W/G naming groups brighter (increased L^*) and more red (higher u^*) than those of the B/K group. This is confirmed by pairwise comparisons. The B/K group match the dress lace to be darker (decreased L^*), greener (decreased u^*) and bluer (decreased v^*) than the W/G and B/G groups ($p < .018$ in all cases with a Bonferroni correction). The W/G and B/G dress colour name groups did not differ in their matches to the dress lace along any axis of CIELUV colour space.

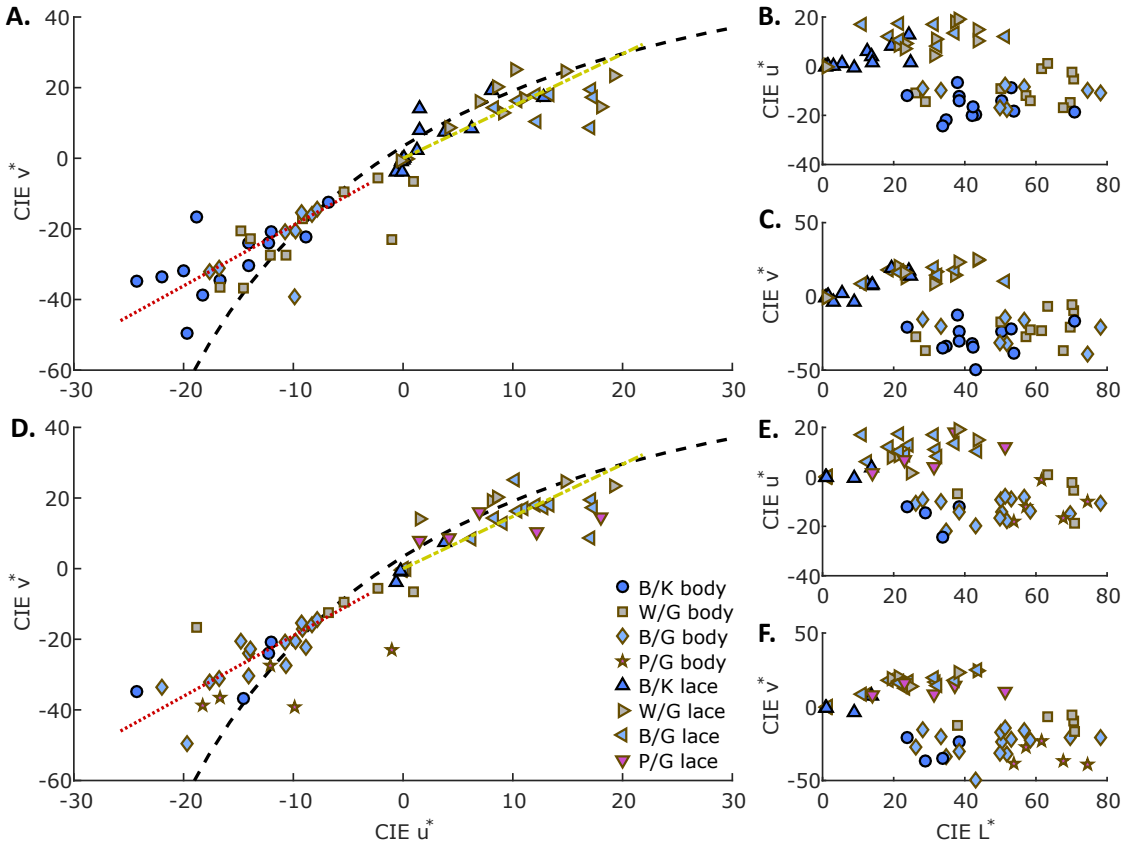


Figure 5.3: Dress body and lace colour matches. A-C. The dress body and lace matching data (two points per observer) labelled according to dress colour names. Dress body matches are square, circles and diamonds. Dress lace matches are differently-oriented triangles. D-F. Same data as in A-C labelled according to disk colour names, but note that two observers (who named the disks blue/green and blue/purple) are omitted from this plot as they were excluded from the corresponding analyses. Black dashed line indicates the Planckian locus. Red dotted line is the first principal component of the dress body matches. Yellow dot-dash line is the first principal component of the dress lace matches.

5.3.0.3 Disk colour names better represent dress body and lace match variability than dress colour names

When participants are grouped according to the colour names they assign to their disk matches to the dress body and lace, matches differ between groupings more than between the dress colour name groups, the former better representing the variation in the matches. With the disk colour name grouping employed (Figure 5.3.D-F), there is a significant multivariate difference in the matches to the dress body in CIELUV ($F(9, 78) = 4.41$, $\Gamma_p = 1.01$, $p < .001$). Here, the composite variable that best separates the groups is

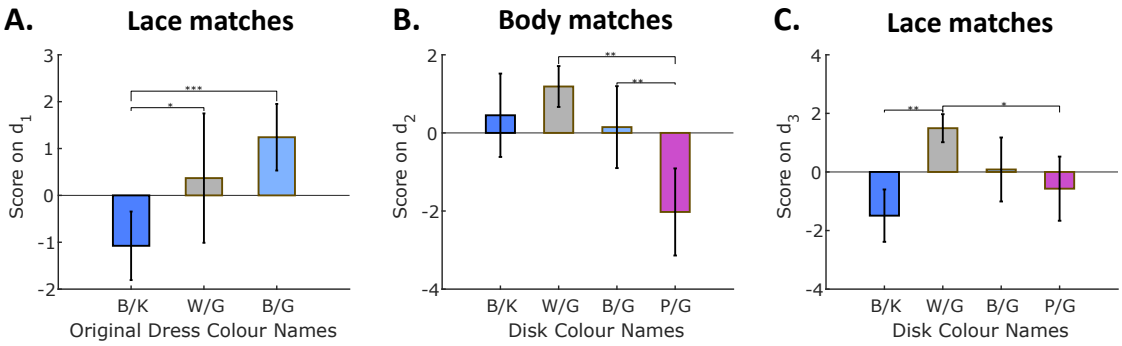


Figure 5.4: Composite variables of the matches. A. Scores on the composite variable d_1 obtained from the MANOVA analysis on the dress lace colour matches categorised by dress colour names. B. Scores on the composite variable d_2 obtained from the MANOVA analysis on the dress body colour matches categorised by disk colour names. C. Scores on the composite variable d_3 obtained from the MANOVA analysis on the dress lace colour matches categorised by disk colour names. Error bars are ± 1 SD. $*p < .05$, $**p < .01$, $***p < .001$.

$d_2 = -0.53L^* - 0.96u^* + 1.82v^*$, with the P/G group having significantly lower scores on d_2 than the B/G and W/G groups ($p = .005$ and $p = .002$, Bonferroni corrected) (Figure 5.4.B). Post-hoc pairwise comparisons of the L^* settings of the matches shows that the B/K group gave a significantly darker dress body match than the W/G and P/G groups (mean differences of 31.30, $p = .007$ and 31.70, $p = .006$, Bonferroni corrected). The W/G group matched the dress body to significantly higher v^* values (less blue), than all other groups ($p < .01$, in all cases with Bonferroni correction).

Matches to the dress lace also differ significantly across disk colour name groups along a multivariate axis defined by the composite variable $d_3 = -0.68L^* - 0.34u^* + 1.96v^*$ ($F(9, 78) = 3.26$, $\Gamma_p = 0.82$, $p = .002$). The B/K group's scores on d_3 indicate that their dress lace matches are darker (decreased L^*) and more achromatic (decreased v^*) than the matches of the other groups, significantly so in comparison to the W/G group ($p = .002$, Bonferroni corrected) (Figure 5.4.C). The W/G and P/G group's d_3 scores also differ significantly ($p = .029$, Bonferrroni corrected). The differences between the B/K and W/G group are further highlighted by the post-hoc pairwise comparisons of the v^* settings, showing that the B/K group gave significantly more achromatic lace matches (v^* closer to zero) than the W/G group (mean difference of 19.71, $p = .001$, Bonferroni corrected). Moreover, the B/K groups v^* settings were also more achromatic than the

B/G group (mean difference of 11.41, $p = .037$, Bonferroni corrected).

5.3.0.4 Examining individual variations within and between naming groups: principal components of the dress body and lace matching data

The matching data (above and in previous studies) show clearly that individuals' colour matches to the dress vary along a continuum in colour space. Our analyses show that splitting the continuum into groups using the names individuals give to their disk colour matches, rather than the colour names they give to the dress itself, better represents the variation in the matches. This suggests that the variation in the matches is more suited to being considered on an individual rather than a group level. The MANOVA analyses show that the dress body and lace matches differ across naming groups in a multivariate manner, and the scores on the composite variables retrieved from those analyses might be considered candidates for quantifying the variability. But the composite variables correspond to variation along an axis that optimally represents the differences in the naming category centroids. To capture individual variability, we use principal components analysis (PCA) instead to quantify individual variation.

The first principal component (PC) of the dress body matches is $PC_b = 0.94L^* + 0.18u^* + 0.3v^*$, explaining 67.12% of the variation (Figure 5.3, red dotted line). The first PC of the dress lace matches is $PC_l = 0.83L^* + 0.31u^* + 0.46v^*$, explaining 88.92% of the variation (Figure 5.3, yellow dot-dash line). The two PCs (PC_b and PC_l) are strongly correlated ($r = .783$, $p < .001$, Figure 5.5.A). The lightness settings of the dress body and lace matches load highly on to the respective components and are themselves strongly correlated ($r = .687$, $p < .001$, Figure 5.5.B). Thus, lightness seems to drive the association between these matches, with the lightness of one predicting the other. In the following analyses, we characterise how other measures relate to individual variations along these principal components.

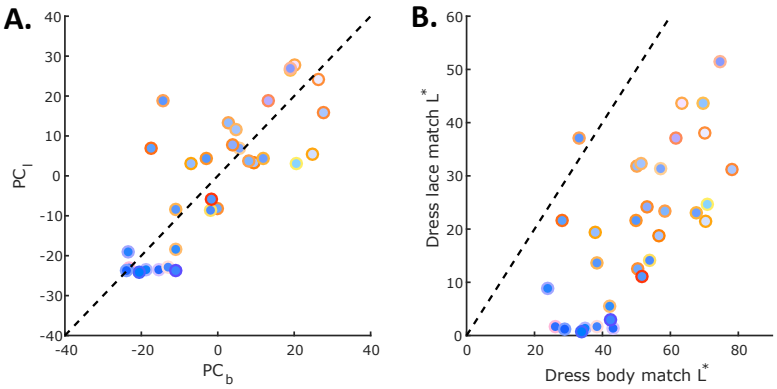


Figure 5.5: Comparing dress body and lace matches. A. The first principal component of the dress lace matches (PC_l) plotted against the first principal component of the dress body matches (PC_b). B. $CIE\ L^*$ setting of the dress lace match plotted against $CIE\ L^*$ setting of the dress body match. All markers are pseudo-coloured according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the $CIE\ Yxy$ values of the matches to $sRGB$ for display.

5.3.0.5 Variation of internal white point explains variation in dress body matches but only when made at the luminance setting of the dress body match

Achromatic settings at the fixed luminance setting of 24 cd/m^2 (Figure 5.6.A) do not differ between dress or disk colour name groups, nor do the achromatic settings produced at the luminance settings of the individual dress body matches (Figure 5.6.D; $p > .332$ in all cases).

However, the achromatic settings made at the individual body match luminance levels do explain some of the variation in the dress body and dress lace matches, as revealed by further principal component analysis. The first PC of the achromatic settings at fixed luminance of 24 cd/m^2 is $PC_f = 0.71u^* + 0.71v^*$, explaining 67.70% of the variance in achromatic settings; and the first PC of the settings at the dress body match luminance is $PC_m = 0.42u^* + 0.9v^*$, explaining 74.63% of the variance. PC_m in turn explains a small amount of the variation in the dress body and dress lace matches (17.56% and 21.34% respectively).

While the first PC of the achromatic settings at a fixed luminance of 24 cd/m^2 (PC_f)

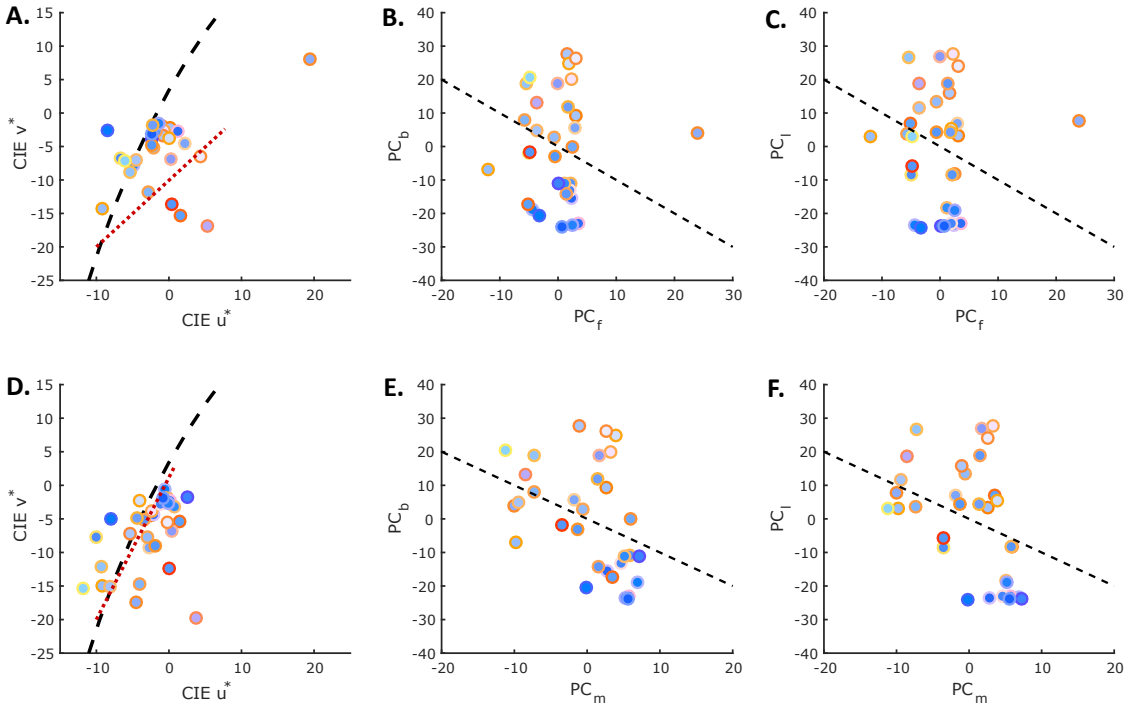


Figure 5.6: Comparing dress and achromatic matches. A. Achromatic settings with luminance fixed at 24 cd/m^2. B. The first principal component of the dress body matches (PC_b) plotted against the first principal component of the achromatic settings at fixed luminance of 24 cd/m^2 (PC_f). C. The first principal component of the dress lace matches (PC_l) plotted against the first principal component of the achromatic settings at fixed luminance of 24 cd/m^2 (PC_f). D. Achromatic settings with luminance set at luminance of dress body match (and therefore varying across participants). E. The first principal component of the dress body matches (PC_b) plotted against the first principal component of the achromatic settings at the luminance of the dress body matches (PC_m). F. The first principal component of the dress lace matches (PC_l) plotted against the first principal component of the achromatic settings at the luminance of the dress body matches (PC_m). Red dotted lines are the first principal components of the data. Black dashed line in A and D is the Planckian locus. Black dashed lines in remaining plots are lines of perfect negative correlation. All markers are pseudo-coloured according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the $CIE\ Yxy$ values of the matches to $sRGB$ for display.

does not correlate with either of the first PCs of the dress body or lace matches (PC_b and PC_l) (Figure 5.6.B-C; $r = .045$, $p = .806$ and $r = -.002$, $p = .990$, respectively), the first PC of the achromatic settings made at the luminance setting of the body match (PC_m) does (Figure 5.6.E-F; $r = -.419$, $p = .017$ and $r = -.462$, $p = .008$, respectively). This result suggests that the two sets of achromatic matches are unrelated and indeed they are (Figure 5.7; $r = .180$, $p = .326$).

The differences in the two sets of achromatic adjustments may be explained by the lu-

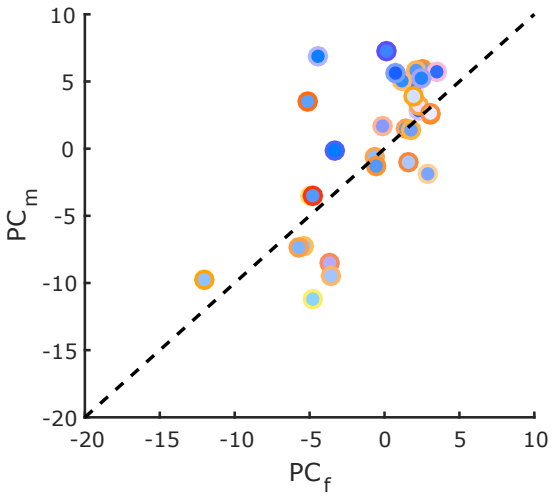


Figure 5.7: Comparing the achromatic matches. The first principal component of the achromatic settings made at the dress body match luminance (PC_m) plotted against the first principal component of the achromatic settings made at a fixed luminance of 24 cd/m^2 (PC_f). Black dashed line is line of perfect correlation. All markers are pseudo-coloured according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the $CIE\ Yxy$ values of the matches to $sRGB$ for display.

minance settings. For one set of adjustments, all participants were required to make the disk appear achromatic at the same fixed luminance (24 cd/m^2). For the other, the luminance setting was unique to each individual, and specifically related to their perception of the dress. In the latter case, when luminance levels are high, achromatic settings are bluer than when luminance levels are low (negative correlation between L^* and adjusted v^* settings, $r = -.379$, $p = .032$).

5.3.0.6 Variation in illumination estimates for the dress photo explain some of the variation in dress colour matches

PCA of the illumination matches in CIELUV colour space uncovers a PC (red dotted line in Figure 5.8.A) that explains 80.21% of the variance defined by $PC_i = 0.57L^* + 0.357u^* + 0.73v^*$. While illumination matches do not differ across dress or disk colour name groupings ($p > .177$ in all cases), the first PC of the illumination matches (PC_i) correlates significantly with the first PC of the dress body matches (PC_b; Figure 5.8.B; $r = -.395$, $p = .026$). There is a similar trend for the first PC of the dress lace matches

(PC_l; Figure 5.8.C; $r = -.343$, $p = .055$). These relationships are not predicted by correlations between any of the univariate variables (L^*, u^* or v^*), except in the case of the lace matches where the v^* setting of the dress lace match is negatively correlated with the v^* setting of the illumination match, with a bluer illumination match implying a yellower lace match. The fact that L^* and v^* load highly on to all PCs considered (PC_b, PC_l and PC_i) highlights a relationship between brightness and blue-yellowness perception. Perceiving illuminations as bluer is linked to perceiving illuminations as darker, and in turn to a brighter, more achromatic dress body match, while yellower is linked to brighter in illumination perception, and to a darker, bluer body match. This relationship between perceived illumination colour and brightness is clearly visible in the relationship between CIE L^* and correlated colour temperature (CCT) of the illumination match (Figure 5.8.D, $r = -.778$, $p < .001$). There is a similar relationship between $CIE\ L^*$ and CCT of the dress body matches ($r = -.502$, $p = .003$).

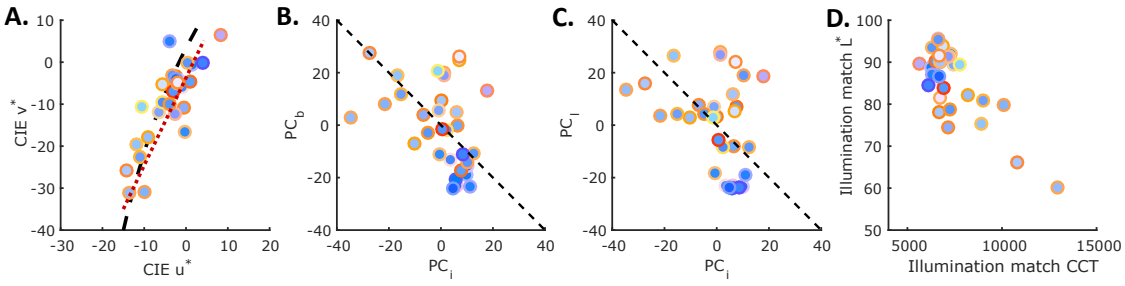


Figure 5.8: Comparing dress and illumination matches. A. Illumination matches plotted in the $CIE\ u^*v^*$ chromaticity plane. Black dashed line is Planckian locus. Red dotted line is first principal component. B. The first principal component of the dress body matches (PC_b) plotted against the first principal component of the illumination matches (PC_i). C. The first principal component of the dress lace matches (PC_l) plotted against the first principal component of the illumination matches (PC_i). D. Illumination match $CIE\ L^*$ setting plotted against illumination match correlated colour temperature value (CCT). Black dashed lines in B, C are lines of perfect correlation. All markers are pseudo-coloured according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the $CIE\ Yxy$ values of the matches to $sRGB$ for display.

The variation in the illumination matches is not related to the variation in the achromatic settings made at a fixed luminance of 24 cd/m^2 (Figure 5.9.A; $r = .051$, $p = .782$), but is related to the variation in the achromatic settings made at the luminance setting of the dress body match (Figure 5.9.B; $r = .367$, $p = .039$). The PC loadings suggest that

lighter and whiter/yellower (higher L^* and v^* settings implying a higher PC_i component score) illumination matches are related to whiter/yellower (higher v^* setting implying higher PC_m component score) achromatic settings when the luminance is fixed at the participant's body match luminance setting. This result is difficult to interpret as the relationship is not seen between the underlying values (no relationship between the L^* or v^* settings of the two adjustments and no relationship between L^* settings of the illumination matches and v^* settings of the achromatic settings, $|r| < .280$, $p > .120$, in all cases). However, there is a relationship between the u^* settings of the two adjustments ($r = .651$, $p < .001$).

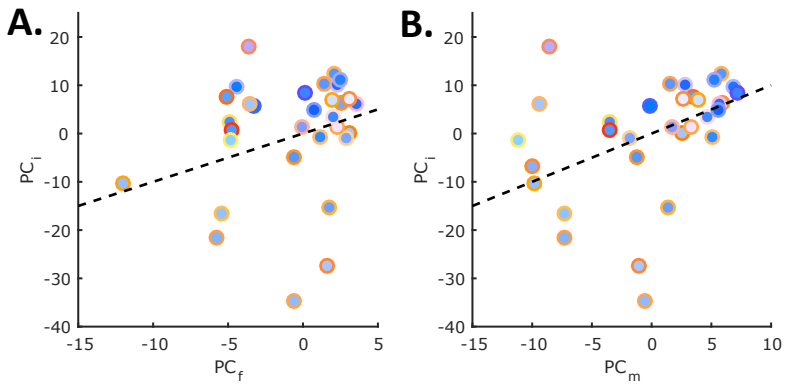


Figure 5.9: Comparing achromatic and illumination matches. A. The first principal component of the illumination matches (PC_i) plotted against the first principal component of the achromatic settings at fixed luminance of 24 cd/m^2 (PC_f). B. The first principal component of the illumination matches (PC_i) plotted against the first principal component of the achromatic settings at the dress body match luminance (PC_m). Black dashed lines are lines of perfect correlation. All markers are pseudo-coloured according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the $CIE\ Yxy$ values of the matches to $sRGB$ for display.

5.3.0.7 Individual differences in colour constancy thresholds assessed via a generic illumination discrimination task do not predict dress colour perception

With IDT thresholds grouped according to dress colour names (Figure 5.10.A), there is a significant main effect of direction of illumination change on thresholds ($F(1.97, 57.21) = 17.6$, $p < .001$, with a Greenhouse-Geisser correction). Thresholds for bluer illumination

changes are larger than those for yellower, redder or greener illumination changes ($p < .013$ in all cases with Bonferroni correction) and thresholds for yellower illumination changes are significantly larger than for greener illumination changes ($p = .029$, with Bonferroni correction). However, there is no interaction effect of group and direction of illumination change on thresholds ($F(3.95, 57.21) = 0.72$, $p = .578$, with a Greenhouse-Geisser correction). There is also no main effect of group ($F(2, 29) = 1.16$, $p = .329$).

With participants grouped according to disk colour name (Figure 5.10.B), the significant main effect of direction of illumination change remains ($F(1.97, 51.12) = 10.78$, $p < .001$, with a Greenhouse-Geisser correction). Thresholds for bluer illumination changes are still larger than for redder and greener illumination changes ($p < .005$ in both cases with Bonferroni correction), but are not larger than those for yellower illumination changes ($p = .172$ with Bonferroni correction). The difference between yellower and greener thresholds still holds ($p = .049$ with Bonferroni correction). Again though, there is no interaction effect ($F(5.90, 51.12) = 0.48$, $p = .820$, with a Greenhouse-Geisser correction) or main effect of disk colour name ($F(3, 26) = 2.41$, $p = .090$).

There are also no significant correlations between the dress body and lace match PCs (PC_b and PC_l) and IDT thresholds for any direction of illumination change ($p > 0.1$ in all cases), nor was there a relationship between the PC of the illumination matches (PC_i) and any direction of illumination change ($p > 0.1$ in all cases).

5.3.0.8 Does chronotype explain some of the variability in dress colour perception?

Scores on the MEQ do not differ significantly across either the dress or disk colour name groupings ($F(2, 29) = 1.14$, $p = .335$ and $F(3, 26) = 0.73$, $p = .544$, respectively). In addition, the correlation between MEQ scores and the first PC of the dress body matches was not significant at the 5% level (PC_b; Figure 5.11.A; $r = .344$, $p = .054$). However, as the effect size is medium (MEQ scores explain 11.83% of the variation in the matching data) and we may have lacked enough power for significance, we consider the implications

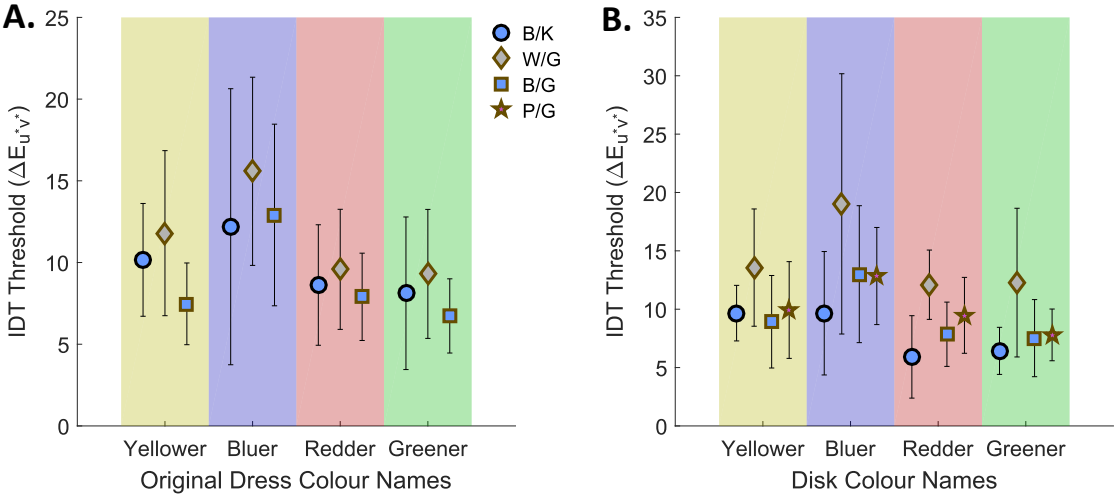


Figure 5.10: IDT thresholds. A. Mean illumination discrimination task (IDT) thresholds of the dress colour naming groups. B. Mean IDT thresholds of the disk colour naming groups. Coloured underlays indicate the direction of illumination change. Error bars are ±1 SD.

of such a correlation. L^* loads highly onto PC_b and v^* loads moderately. This implies that a high score on PC_b leads to a bright, white/yellow dress body match (i.e., morning types, with high MEQ scores, make brighter, whiter/yellower dress body matches than evening types, with low MEQ score, as we hypothesised). There is also a relationship between MEQ scores and the first PC of the illumination matches (PC_i; Figure 5.11.B; $r = -.323$, $p = .072$), although this correlation is also not significant at the 5% level. Similarly, L^* and v^* load highly onto PC_i implying that a high score on PC_i indicates a bright, white/yellow illumination match; in other words, evening types make brighter, whiter/yellower illumination matches than morning types.

5.3.0.9 Age explains some variability in dress colour perception but is also related to MEQ score

While age does not differ significantly across dress colour name groups ($F(2, 29) = 1.12$, $p = .340$), there is a significant difference in age across disk colour name groups ($F(3, 26) = 5.30$, $p = .005$). The W/G disk colour name group are significantly older than both the B/K and B/G disk colour name groups ($p = .014$ and $p = .006$ with Bonferroni correction, respectively). Moreover, age is related to: MEQ scores (Figure 5.12.A; $r = .316$, $p = .078$),

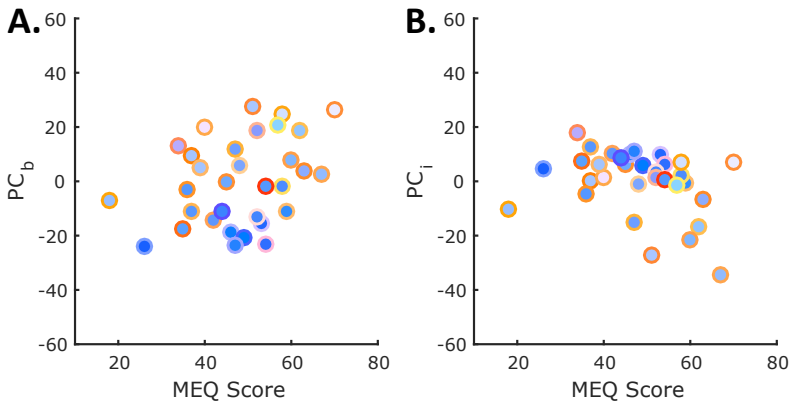


Figure 5.11: MEQ Scores I. A. The first principal component of the dress body matches (PC_b) plotted against MEQ score. B. The first principal component of the illumination matches (PC_i) plotted against MEQ score. All markers are pseudo coloured according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the $CIE\ Yxy$ values of the matches to $sRGB$ for display.

with older individuals having higher MEQ scores (more morning type); the first PC of the dress body matches (PC_b; Figure 5.12.B; $r = .534$, $p = .002$), with older individuals giving brighter (increased L^*) and whiter (increased v^*) dress body matches (age positively correlates with L^* and v^* settings of dress body matches, $r > .427$, $p < .015$ in both cases); the first PC of the dress lace matches (PC_l; Figure 5.12.C; $r = .398$, $p = .024$), with older individuals giving brighter (increased L^*) and yellower (increased v^*) dress lace matches (age positively correlates with L^* and v^* settings of dress lace matches, $r > .381$, $p < .030$ in both cases); and the first PC of the achromatic settings made at the luminance value of the dress body match (PC_m; Figure 5.12.E; $r = -.343$, $p = .055$). Yet, age is not related to the latter achromatic settings along any single axis of CIELUV colour space, nor to the first PC of the achromatic settings made at a fixed luminance of 24 cd/m^2 (PC_f) or of the illumination matches (PC_i) (Figure 5.12.D; $r = -.191$, $p = .296$ and Figure 5.12.F; $r = -.052$, $p = .777$, respectively).

5.4 Discussion

We set out to test the colour constancy explanation of the dress phenomenon. First, we further investigated whether differences in perceived colour of the dress are explained by

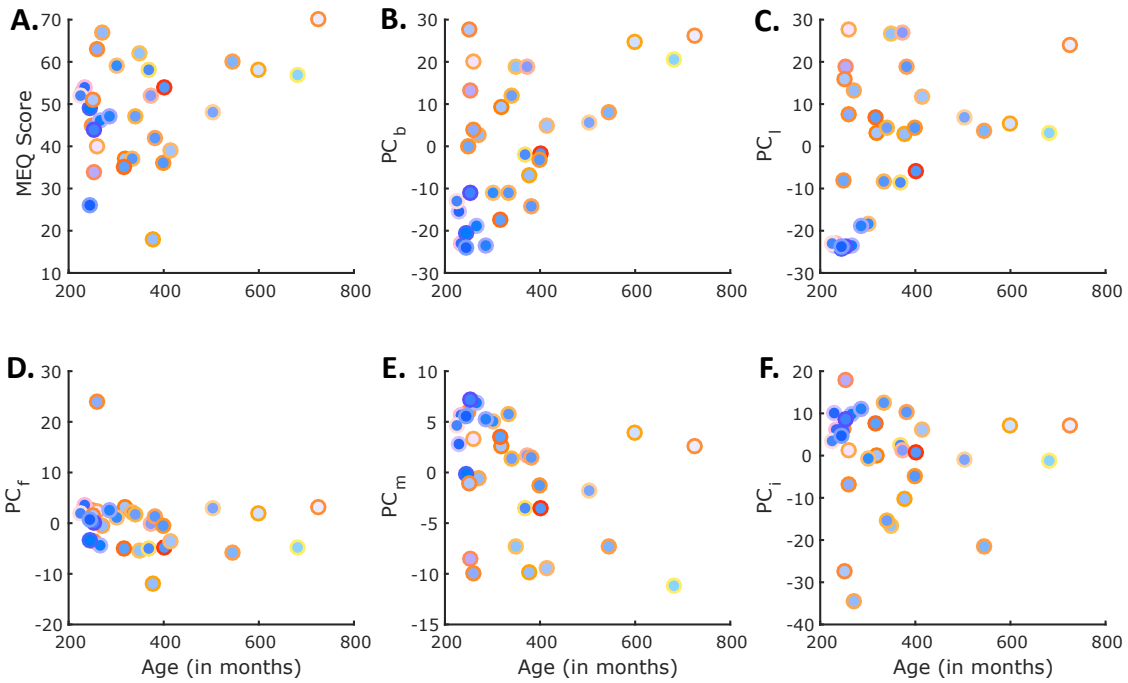


Figure 5.12: MEQ scores II. A. MEQ scores plotted against age. B. The first principal component of the dress body matches (PC_b) plotted against age. C. The first principal component of the dress lace matches (PC_l) plotted against age. D. The first principal component of the achromatic settings made at a fixed luminance of 24 cd/m^2 (PC_f) plotted against age. E. The first principal component of the achromatic settings at the luminance setting of the dress body match (PC_m) plotted against age. F. The first principal component of the illumination matches (PC_i) plotted against age. All markers are pseudo coloured according to the corresponding dress body (marker face) and dress lace (marker edge) matches for each participant by converting the $CIE\ Yxy$ values of the matches to $sRGB$ for display.

differences in inferred illumination chromaticity, asking whether different prior assumptions were responsible for variations in inferred illumination by using chronotype as a proxy for experience (on the assumption that experience governs the formation of an individual's prior assumptions). Second, we used an established illumination discrimination task to measure colour constancy thresholds for illumination changes in our observers, assessing whether individual differences in generic colour constancy may explain individual differences in perception. The results of the study support the colour constancy hypothesis of the dress phenomenon in demonstrating a relationship between the inferred illumination chromaticity in the image and colour matches to the dress (replicating previous results). Correlations, although non-significant, between chronotype and dress body matches and between chronotype and inferred illumination chromaticity suggest that unconsciously

embedded expectations of illumination characteristics ("illumination priors"), shaped by experience, may act by biasing perception under uncertainty and be responsible for the differences in colour names assigned to the dress. In particular, we find that observers differ not only in the colour names they give to the dress, but also in the chromaticity and luminance of the illumination they estimate to be incident on the dress, with the variation in these illumination matches related to both dress body and lace colour matches. Further, illumination and dress body matches both show some relationship to MEQ scores, suggesting that chronotype may prove useful as a marker for the chromatic bias of an observer's illumination prior. In addition, our results demonstrate a disconnect between perception and naming. Our observers report different colour names for the dress photograph and their isolated colour matches, the latter best capturing the variation in the matches.

We find, in agreement with the previous report of Lafer-Sousa *et al.* (2015), that observers fall into three main groups on the basis of how they named the dress colours on first view: the blue/black (B/K), white/gold (W/G) and blue/gold (B/G) dress colour names groups. We find that colour matches to the dress lace differ significantly between the three groups, in both lightness and chromaticity (along both blue/yellow and red/green axes), while colour matches to the dress body differ only along a red/green chromatic axis. These results partially replicate the results of two previous studies. First, Lafer-Sousa *et al.* (2015) report significant differences in both dress body and lace matches along a lightness and chromatic axis between B/K and W/G naming groups, but do not report any differences between the matches of the B/G group (which they call blue/brown) or their "other" category and any other group. Secondly, Gegenfurtner *et al.* (2015) report differences in dress body colour matches between B/K and W/G observers but only in luminance settings, a result that has since been replicated by the same group (Toscani, Gegenfurtner & Doerschner, 2017). The differences in results between studies may be due to differences in sample sizes (53, 15, 38 and 32 for the Lafer-Sousa *et al.* (2015), Gegenfurtner *et al.* (2015), Toscani *et al.* (2017) and current study, respectively), and/or to differences in the luminance and chromaticity of the displayed image and in the chromatic

resolution of the matches.

Observers' colour matches to the illumination on the dress vary in both chromaticity and luminance and correlate negatively with their colour matches to the body of the dress, explaining 15.52% of the variation in the dress body matching data; these results concur with the findings of Witzel *et al.* (2016) and Toscani *et al.* (2017) and are related to the findings of Chetverikov & Ivanchei (2016). That is, observers who make illumination matches that are bluer (than the mean) tend to make dress body matches that are less blue, whereas observers who make illumination matches that are less blue tend to make dress body matches that are bluer. A similar trend exists for the dress lace matching data. The fact that darker illumination matches tend to be more blue than lighter illumination matches fits with the observation that, in nature, indirect lighting and shadows tend to be bluish, being due to an absence of direct (yellowish) sunlight (Churma, 1994). Certain visual illusions suggest that the human visual system might have incorporated this relationship, perceiving dimmer areas of images as being indirectly lit or in shadow, and therefore attributing their bluish tints to the illumination rather than the object (Winkler, Spillmann, Werner & Webster, 2015). It remains an open question whether the bluish tints in brighter areas of the image are also more easily attributed to the illumination than bright yellowish tints. Uchikawa, Morimoto & Matsumoto (2016) have shown that the optimal colour hypothesis (Uchikawa, Fukuda, Kitazawa & Macleod, 2012) predicts a discrepancy in the colour temperature of the inferred illumination in the dress image dependent on the inferred illumination intensity. The peak of the optimal colour distribution for the dress image varies with intensity: high intensity implies a more neutral/yellower illumination (light at 5000 K), while low intensity implies a bluer illumination (light at 20000 K). It remains to be shown though that the optimal colour hypothesis is the algorithm adopted by observers to estimate scene illumination.

Achromatic settings made at the luminance of each observer's dress body match explain more of the variation in the dress body and lace matching data than the illumination matches (34.30% and 31.78%, respectively), with the achromatic settings at variable lu-

minance levels showing the opposite trend to the illumination matches: lighter achromatic settings are more blue than darker achromatic settings, whereas lighter illumination matches are less blue than darker illumination matches. Achromatic settings are commonly used to capture an observer's internal white point, typically assumed to be a measure of the chromaticity of the observer's default neutral illumination (Brainard, 1998). These results suggest that traditional achromatic settings (adjustments of a small patch of fixed luminance to appear achromatic, set against a uniform background) do not reflect the chromaticity of illumination that an observer will estimate for a different and more complex scene; the achromatic settings collected in the present study at a fixed luminance level of 24 cd/m^2 were not associated to dress colour perception or illumination matches. This lack of association agrees with the findings of Witzel *et al.* (2016), who also find that classical achromatic settings (same fixed luminance across all observers, but with luminance texture), do not predict dress body or lace colour matches.

However, the achromatic settings made at the luminance setting of each participant's dress body match suggest that an observer's internal white point is influenced by luminance, with settings made at higher luminance levels bluer than those at lower luminance levels. It therefore seems that achromatic settings reveal an observer's bias only when an appropriate fixed luminance is chosen for the achromatic adjustment. It is plausible then that when processing a scene radiance image, an observer initially estimates the irradiance level of the illumination, effectively parsing the overall scene radiance (proportional to image irradiance) into material reflectance and illumination irradiance components. Only after this initial parsing does the observer infer the illumination chromaticity, calling on their prior knowledge of the relationship between illumination irradiance and chromaticity to do so. The bias lies in the balance of that parsing. Indeed, most of the variation in illumination estimates lie in the luminance settings of the illumination matches.

If we assume that prior knowledge of the relationship between illumination irradiance and chromaticity is a commonality across observers and that different fixed luminance levels are the cause of the variation in v^* settings seen in the set of achromatic adjustments made

at the luminance setting of each observer's dress body match, then we expect a higher fixed luminance to lead to bluer achromatic settings than lower fixed luminance across all observers. In a control experiment (see control experiment in Appendix C) a subset of the participants ($n = 7$) made achromatic settings at each of 5 different luminance levels, spanning the range from minimum to maximum dress body match luminance in the main experiment. There was a clear trend towards increasing blueness (lower v^*) with increasing luminance in all participants.

The discrepancy between the blueness-brightness relationship of these variable-luminance achromatic settings and the illumination estimation matches reinforces the conclusion that measurements of subjective white points do not necessarily reflect the chromaticity of the observer's internal default illumination. The illumination matches imply that illuminations that are perceived as brighter are perceived as whiter/yellower, while darker illuminations are perceived as bluer. Similar conclusions from different methods of illumination estimation (Uchikawa *et al.* (2016) and Witzel *et al.* (2016)) support this interpretation. The fact that achromatic settings vary in the opposite way - being bluer at higher luminance levels and yellower at lower luminance levels - suggests that they are measuring something other than the default illumination estimation. Of course, as we pointed out in the methods section, these differences are to be expected considering the different methods used to obtain the matches.

The observation that achromatic settings vary with luminance has been made before, but the underlying cause of this variation remains unclear (Chauhan, Perales, Hird & Wuerger, 2014; Kuriki, 2015). Conversely, Brainard (1998) found that achromatic settings made at fixed luminance levels lie along a straight line in a three-dimensional cone space, concluding from this that changing luminance did not affect the chromaticity setting of an observer's achromatic point.

The fact that the principal variation in achromatic settings falls roughly on the daylight locus (as also reported by Witzel *et al.* (2016)), which we find to be more so for the variable-luminance achromatic settings than the 24 cd/m^2-fixed-luminance settings, may seem to

argue that achromatic settings do reflect illumination estimation. Yet if the settings indicate instead the surface chromaticity that appears neutral under natural illumination of that luminance, they would be expected to evince the same variation. The fact that observers require more blue to make the isolated disk appear neutral at higher luminance levels may indicate that they implicitly assign more yellow to the illumination. In the fixed luminance achromatic setting task, the only cue to the irradiance of the illumination - the brightness of the isolated matching disk against a black background - does not vary between observers, and hence there is less cause for variation in the achromatic point between observers. The fact that the variation in achromatic setting at fixed luminance occurs only along an axis roughly orthogonal to the blue-yellow axis also supports the notion that the coupling between assumed irradiance and chromaticity of the illumination falls mainly along the blue-yellow axis. Whatever the correct interpretation of this difference between achromatic settings and illumination estimation, we suggest that care must be taken in future studies to ensure that fixed luminance levels are not influencing behaviour in tasks where achromatic settings are taken as a measure of illumination chromaticity. Under the assumption that our illumination matches do capture the variation in illumination estimation across observers, these results suggest that conventional achromatic settings are not indicative of the illumination chromaticity that an observer will estimate on a particular scene.

The main question we address in this study is whether the observer's tendency to infer a particular illumination on the dress is underpinned by a generic bias in illumination priors, as suggested by Lafer-Sousa *et al.* (2015). Witzel *et al.* (2016) conclude from similar measurements that the bias is specific to the image, not indicative of a general underlying unconscious expectation of illumination chromaticities. Our additional results suggest that the bias in people who tend to see the dress as lighter and the illumination as darker, is at least partly generic. Scores on the MEQ, a questionnaire that quantifies chronotype, partially predict dress body and illumination matches. We speculate that the MEQ score provides a quantification of an individual's internal illumination priors

as it may predict the illumination chromaticities to which an observer is most frequently exposed due to their interaction with specific environments engendered by their internal body clock. In other words, MEQ scores are indicative of an observer's perceptual biases due to illumination chromaticity estimation. Indeed, we find that observers who score highly on the MEQ and hence are considered more morning-type score higher on the first PC of the dress body match (PC_b) than the more evening-type individuals and lower on the first PC of the illumination matches (PC_i). This finding supports the hypothesis that an observer's visual system is calibrated for the illuminations in which they find themselves most often, with morning types making bluer illumination matches compared to evening types. It is plausible that morning types are more often exposed to bluer illuminations than evening types as they are more likely to spend time outdoors during the day in bluish daylights (Hernández-Andrés *et al.*, 2001) and less time in yellowish artificial lighting at night (Lafer-Sousa *et al.*, 2015).

In a recent study, Wallisch (2017) also found that morning types (strong larks in that study) are more likely to name the dress W/G than B/K compared to evening types (strong owls). The results are stronger than the relationship we find here. The difference might be due to how Wallisch (2017) parcels observers into chronotype groups. Firstly, the observers are self-categorised and are asked to assign themselves to one of 4 groups ("strong lark", "lark", "strong owl" and "owl"). In our study we use an established questionnaire (the MEQ) that is widely used to assess chronotype. The MEQ places observers on a scale rather than into distinct groups allowing for more variability that might reduce our ability to show the effect. The same is true for how we represent perception. In our study, observers make colour matches to the dress body and lace and the PCs of these matches are correlated with scores on the questionnaire. Again, Wallisch (2017) allowed observers to place themselves into a perceptual category (e.g. W/G observer type).

Age is a confounding factor in our results, with MEQ scores showing a relationship to the PCs of the colour matches, but age also differing significantly across the disk colour names groups and correlating with the MEQ scores. From this, one may infer that chronotype

is not at all related to individual differences in colour perception (by being indicative of chromatic bias in illumination priors), but that age is the underlying variable that drives the relationship. The correlation between age and MEQ scores is not a surprise as it is well known that chronotype varies with age (Adan, Archer, Hidalgo, Milia, Natale & Randler, 2012). It would be premature, though, to conclude that age and not experience is the driving factor here as an individual's illumination priors may change dynamically during their lifetime as their daily experiences also change. On the other hand, aging is known to affect lower level visual factors such as lens optical density (Pokorny, Smith & Lutze, 1987). To separate these two effects, a study is required that controls for age while sampling from a population of varying chronotypes.

Individual differences in generic colour constancy measurements - via the IDT - did not predict individual differences in colour matching or naming of the dress. The lack of relationship between dress and illumination matches and IDT thresholds might be because the blue bias is present in all individuals and that the main driver underlying the individual differences in perception of the dress photograph is at a higher level than the illumination discrimination task probes, such as the interpretation of the illumination in the scene. Different types of colour constancy tasks (e.g. naming the colours of objects under different illuminations or retrieving the same object under multiple illuminations) may reveal a relationship between colour constancy measurements and dress colour naming and matching. Alternatively, it might be that the colour constancy tasks commonly used in psychophysical experiments do not invoke the use of illumination priors, as the stimuli are well controlled and lack ambiguity. For illumination priors to be used, it seems logical that the incoming sensory information must be somewhat uncertain and that weighting by the prior is necessary to keep perception stable.

Our data also reveal a dissociation between naming and matching, in agreement with our assertion that the distribution of colour names assigned to the dress do not fully capture the variety of perceptions experienced by the population. Colour appearance matches to the dress body and lace fall along a continuum, not into discrete groupings, confirming

previous results (Gegenfurtner *et al.*, 2015; Lafer-Sousa *et al.*, 2015; Witzel *et al.*, 2016). Further, when participants are asked to name their matches presented in isolation, these names differ from those they assign to the dress. The disk colour names also predict the dress colour matches better than the dress colour names, capturing more of the variation in the matching data. The effect is not explained by simple local contrast effects, as in both the naming and matching tasks, the matching disk is surrounded by the same black background; and at the time of making the match, observers seem generally satisfied that the match is representative of how they perceive the colours in the image. A speculative conclusion is that naming the dress B/K, W/G or B/G involves cognitive as well as phenomenal processes, in that observers do not simply name what they "see", but that there is a higher-order judgement about the colour and the linguistic category to which it belongs based on the object with which it is associated and its surroundings. However, further work is needed to assess this claim.

This observation is relevant to the historical debate over the level on which colour constancy exists or may be measured: is the stability of object colour maintained by low-level mechanisms which are inaccessible to conscious influence (Crick & Koch, 1995) or does colour constancy require an act of judgement or reasoning at a higher level (Hatfield & Allred, 2012)? The former would entail perfect constancy in the phenomenal sense, i.e. that colours would appear the same under an illumination change and therefore observers would make perfectly equivalent appearance matches (e.g. in the forced-choice matching paradigm of Bramwell & Hurlbert, 1996), whereas the latter would permit the observer to tolerate differences in colour appearance while still judging the two surfaces under comparison to be the same (e.g. in the object selection task of Radonjić *et al.*, 2016*a*). This difference has been brought out empirically in previous colour constancy studies, most notably in the "hue-saturation" (appearance) vs. "paper" (surface identity) matches of Arend & Reeves (1986) and more recently by Radonjić & Brainard (2016). The matching and naming results reported here suggest that higher level reasoning does play a role in everyday colour constancy, and that observers implicitly or explicitly reason about the

colour of objects in a scene and may report a colour name for an object which is not uniquely representative of how it phenomenally appears.

Another possible explanation for the dissociation between the matching and naming results is that the observer is limited in matching the colour appearance of a textured object in a complex scene to a uniformly coloured matching disk against a black background. Yet here we are asking the observer to make a single match in the same way that the observer gives a single colour name to the object. The dress photograph was a social media frenzy because different individuals named the dress different colours; e.g. "blue body with black lace" or "white body with gold lace" using a single colour term to describe each part of the dress. This implies that despite a noisy photograph in which dress colours vary on a pixel by pixel basis, observers assign a single colour to the dress body and lace. What we aim to obtain here through our colour appearance matches is exactly that, the colour that the dress appears to the observer despite the noise, texture, shading and other confounds in the image. Obtaining appearance matches in such a way is a standard method in colour science and is a method used by other research groups investigating the dress phenomenon (e.g. Gegenfurtner *et al.*, 2015) and other aspects of colour perception (e.g. Giesel & Gegenfurtner, 2010; Poirson & Wandell, 1993). Additionally, anecdotal evidence from conversations with participants after completion of the task suggests they were generally satisfied with their matches and thought their selections accurately represented what they perceived. This reasoning supports the conclusion that the same colour appearance elicits different colour names depending on its context.

5.5 Conclusions

In summary, these data are supportive of the colour constancy explanation of the dress phenomenon, in demonstrating the relationship between inferred illumination and perceived dress colour. Furthermore, the results suggest that individual differences in perception of the dress photograph may be partly explained by chromatic bias in illumination priors and that these biases are influenced by factors related to individual experiences. However, we

show that a generic measure of individual differences in colour constancy does not explain variability in perception of the colours of the dress. In addition, we show that perception and naming may in fact be disconnected. Our results suggest that the colour names observers assign to surfaces may depend more on their global perception of the scene rather than only their local surface perception, as observers name their colour matches to the dress differently when presented them in isolation from how they name the colours of the dress in the photograph.

Chapter 6

Learning illumination priors

6.1 Introduction

Having found evidence in earlier chapters that observers utilise an illumination prior (learnt during their lifetime - the nurture hypothesis) to aid colour constancy, in this chapter, we perform an experiment to investigate whether observers can learn an illumination prior over a specific illumination characteristic (hue value) within the short time frame of a psychophysics task. To do so, we borrow an experimental paradigm from the central tendency bias literature.

Central tendency biases in perception are the phenomenon whereby perceptual estimates tend to be biased towards the average of similar recently seen stimuli. The term can be traced back to Hollingworth (1910) who illustrated a central tendency bias for judgements of size. Hollingworth showed participants a square card and, after a delay, asked participants to choose the card that appeared the same size to them from a set of standards. The standard that participants chose as the match for a given reference card C was influenced by the series of reference cards within which C appeared. If C was larger than the average size of the series it was matched to a standard smaller than its actual size and closer to the average. Conversely, when C was smaller than the average, the chosen standard was larger than the actual size and again closer to the average.

Central tendency biases have appeared in the literature under a variety of names (e.g. adaptation-level theory (Helson, 1964), time-order effects (Jamieson & Petrusic, 1975), context effects (Schifferstein, 1995) and contraction bias (Ashourian & Loewenstein, 2011)) and have been illustrated using a variety of stimulus types such as judgements of line length (Ashourian & Loewenstein, 2011; Duffy, Huttenlocher, Hedges & Crawford, 2010; Huttenlocher, Hedges & Vevea, 2000), sweetness (Riskey, Parducci & Beauchamp, 1979), facial expressions (Roberson, Damjanoviv & Pilling, 2007), absolute size (Huttenlocher

et al., 2000), shades of grey (Huttenlocher *et al.*, 2000), time-interval estimation (Jamieson, 1977; Jazayeri & Shadlen, 2010; Ryan, 2011), and most recently, the hue of a coloured patch presented on a monitor referred to hereafter as a simulated surface (Olkkonen, McCarthy & Allred, 2014; Olkkonen & Allred, 2014).

In this Chapter, we are interested in central tendency biases in the domain of colour perception. Olkkonen *et al.* (2014) showed a central tendency bias for perceptual estimates of the hue (or colour) of a light-based surface. They showed that the point of subjective equality (PSE) of a particular reference hue, after a short delay, varied with respect to the stimulus series (the set of reference hues) within which it was shown. A reference hue that was bluer (larger hue value) than the mean hue of the series had a yellower PSE and vice-versa. If bias is calculated as the PSE hue value minus the true hue value then a reference with a bluer PSE than its true hue will have a positive bias and a reference with a yellower PSE will have a negative bias. The magnitude of the bias then represents the strength of the central tendency. Olkkonen *et al.* (2014) also showed that the strength of the bias can be manipulated by increasing the delay between the presentation of the reference hue and the test hue (about which a judgement is made: "is it bluer or yellower than the reference?") and by increasing the noise in the stimuli (adding chromatic heterogeneity to the stimulus). In the former case, their intention is to manipulate internal noise, effectively the decay of the stimulus representation in memory. In the latter, it is to manipulate external noise. In both cases they find that increasing the noise increases the magnitude of the central tendency bias. They interpret this as evidence for the use of prior knowledge in visual perception that shapes perception under uncertainty.

As we discussed in Chapter 1, the light signal received by the eye depends on variables such as viewing angle and lighting conditions, creating extrinsic variability. In addition, the biological constraints of the human visual system give rise to intrinsic noise. It has long been suggested that the visual system can overcome such ambiguity with the aid of visual priors. It has been shown that a Bayesian observer model accounts for the central tendency bias in visual perception (Huttenlocher *et al.*, 2000; Jazayeri & Shadlen, 2010).

Also in Chapter 1, we discussed how Bayesian models can be used to model human colour constancy (Brainard *et al.*, 2006; Brainard & Freeman, 1997). If such models of colour constancy hold, then observers should be able to form an illumination prior over repeated illumination exposures that manifests as a central tendency bias for perceptual estimates of illumination hue. We investigate this hypothesis in the work that follows.

In Experiment 1, using the same experimental procedure as Olkkonen *et al.* (2014), but changing the stimuli from light-based surfaces to illuminations presented against a grey background, we show a central tendency bias for perceptual estimates of illumination hue. This suggests that a prior for illuminations can by quickly learnt during the time frame of an experiment. However, in Experiment 2, we show that a proportion of the bias may be explained by an alteration in the observers' colour term usage, illustrated by showing how the way an observer categorises the illuminations into colour categories depends on context (the series of illuminations in which the illumination appears). This implies that asking participants to use colour descriptors to issue their responses in Experiment 1 exaggerates the central tendency bias. In Experiments 3 and 4, we explore this further by repeating Experiment 1 with different task designs that avoid the use of colour descriptors. Experiment 3 shows no evidence of a central tendency bias; however, in Experiment 4, a bias emerges but only after participants have been trained to do the task.

6.2 Experiment 1

In Experiment 1, we aimed to replicate the results of Olkkonen *et al.* (2014) but for perceptual estimates of illumination hue rather than estimates for the hue of a light-based surface. Using a custom built light box, in a series of trials, participants were exposed to a reference illumination and, after a two second delay, indicated whether a test illumination was "bluer" or "yellower" than the reference. A staircase procedure was used to find the PSE, considered a correlate of the perceptual estimate of the reference illumination held in memory. Trials were split into three blocks with five reference hues (some overlapping) in each block. The results reveal a central tendency bias for illumination hue estimates. In

addition, they show that perceptual estimates of a given reference hue depend on context as the estimates of reference hues appearing in multiple blocks differ significantly across blocks.

6.2.1 Methods

6.2.1.1 Participants

Eleven participants (seven female, mean age 22 ± 3 years) participated in Experiment 1. Six were members of the laboratory in which the study was conducted but all were unaware of the purpose of the experiment and had no prior exposure to the stimuli. The remaining five participants were recruited through word of mouth. All participants had normal or corrected to normal visual acuity and gave consent prior to participation. All were assessed for normal colour vision using the Ishihara colour plates and the Farnsworth-Munsell 100 hue test. Each participant received a voucher as compensation for their time.

6.2.1.2 Apparatus

Stimuli were shown in a custom built light box (Figure 6.1.A) with dimensions 125 *cm* (height) by 80 *cm* (width) by 53 *cm* (depth). The light box housed two 13-channel spectrally tunable LED luminaires (Section 2.3.1). By controlling the output power of each individual LED in the luminaires, it is possible to create a spectrum of light that has almost any desired chromaticity and luminance. The full procedure for generation of such lights is detailed elsewhere (Finlayson *et al.*, 2014; Pearce *et al.*, 2014; Radonjić *et al.*, 2016*b*, see also Section 2.4). During the experiment, the participant's view was restricted to a piece of uniform grey mount board attached to the back wall of the light box (Figure 6.1.B). The viewing port that restricted their view was 13.5 *cm* wide by 9 *cm* high. Participants rested their chin on a chin rest positioned 20 *cm* from the viewing port. The illuminations used in the experiment were calibrated to produce a spectral radiance of specified CIE 2006 xy chromaticity with fixed luminance of 15 cd/m^2 (Figure 6.1.C) when reflected from

the grey mount board. To that end, the basis functions used in the fitting procedure were measured from the grey mount board and all stimuli were controlled with respect to this surface.

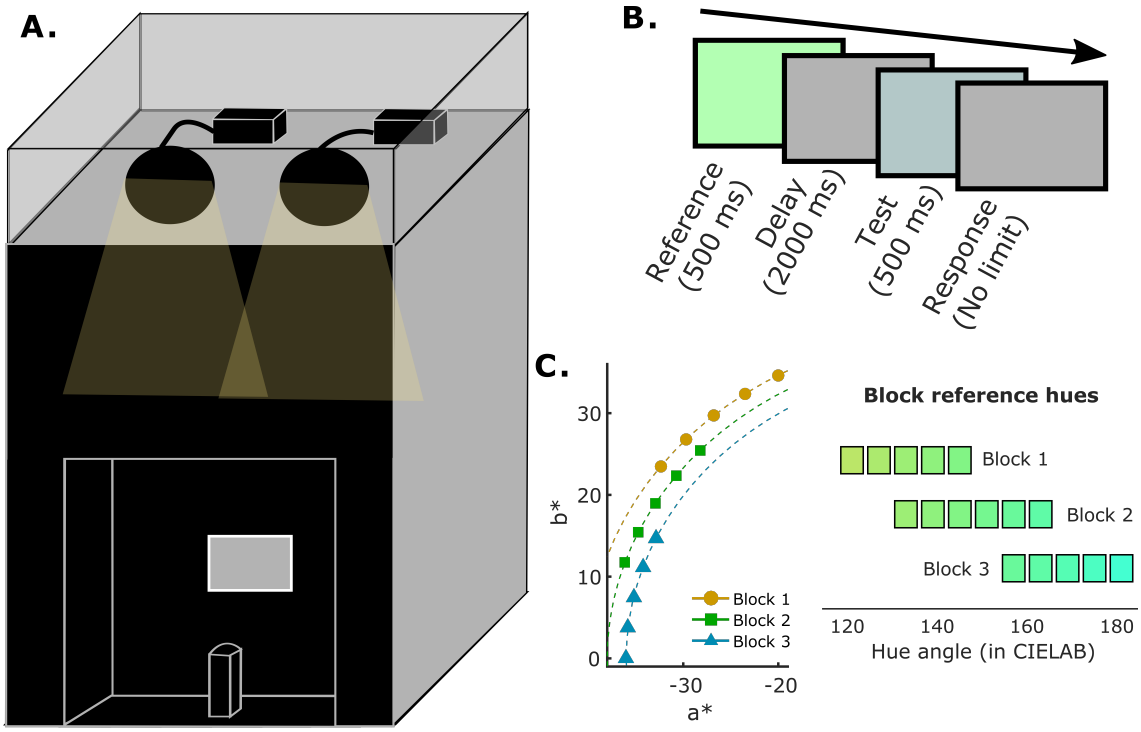


Figure 6.1: The experiment. A. Participants position their head on a chin rest and look into a custom built lightbox where the illumination is tightly controlled using spectrally tunable LED lamps (Section 2.3.1). B. The participant's view is restricted to the backwall of the stimulus box. On each trial, participants first see a reference illumination (500 *ms*) presented against the backwall. After a delay during which the adapting illumination is presented (2000 *ms*), they are presented with a test illumination (500 *ms*). Participants then indicate whether the test was "bluer" or "yellower" than the reference while under the adapting illumination. C. The chromaticities and a cartoon representation of the three blocks of illuminations used in the experiment.

6.2.1.3 Stimuli

Eleven reference illuminations were specified (Figure 6.1.C, Table 6.1). By design, these reference illuminations elicit the same response at the retina as the simulated surface stimuli used by Olkkonen *et al.* (2014). For each reference illumination, 20 comparison illuminations were generated in each direction around the hue circle in $CIE\ L^*a^*b^*$ colour space. These form the set of test illuminations and were separated by hue steps of ≈ 1.5

degrees ($\approx 1\ \Delta E_{a^*b^*}$ for constant chroma of 40). In addition, a $D65$ metamer was generated with the same luminance as the rest of the stimuli (15 cd/m^2) to serve as the adapting illumination.

Table 6.1: $CIE\ Yxy$ and $CIE\ L^*a^*b^*$ values and hue angle (in CIELAB) of the eleven reference stimuli used in the experiments along with the details of which block(s) they appeared in.

Number	Block	Hue angle	Y (cd/m^2)	x	y	L^*	a^*	b^*
1	1	120	15	0.34	0.41	100	-20.00	34.64
2	1	126	15	0.33	0.41	100	-23.51	32.36
3	1	132	15	0.33	0.40	100	-26.77	29.73
4	1,2	138	15	0.32	0.40	100	-29.73	26.77
5	1,2	144	15	0.31	0.40	100	-32.36	23.51
6	2	150	15	0.30	0.39	100	-34.64	20.00
7	2,3	156	15	0.29	0.38	100	-36.54	16.27
8	2,3	162	15	0.28	0.38	100	-38.04	12.36
9	3	168	15	0.28	0.37	100	-39.13	8.32
10	3	174	15	0.27	0.36	100	-39.78	4.18
11	3	180	15	0.26	0.35	100	-40.00	0.00

6.2.1.4 Procedure

The procedure (Figure 6.1.B) followed that of Experiment 1 in Olkkonen *et al.* (2014) (also matching the memory condition in Olkkonen & Allred (2014)). The task began with a 2 minute adaptation to the adapting illumination. On each trial, participants were presented with one of the 11 reference illuminations (500 ms) followed by a delay under the adapting illumination (2000 ms) before presentation of a test illumination (500 ms). After viewing the test illumination, participants indicated via a button press whether the test illumination appeared "bluer" or "yellower" than the reference (no time limit; note that in Olkkonen *et al.* (2014) participants were asked which stimulus is "bluer" out of the reference and test stimuli, or which stimulus is "yellower", depending on the participant). During the response period, the adapting illumination was presented. The test illumination differed from the reference only in hue (see Stimuli) and, on each trial, the amount of hue difference between the test and reference stimuli was determined by a 1-up, 1-down staircase procedure (fixed step size of 1, terminating after 30 trials) chosen to target

the PSE (or equivalently, the 50% correct point on the psychometric function). The task was completed in three blocks, with a subset of reference illuminations presented in each block (5 reference illuminations in each, see Table 6.1). In each block, the set of reference illuminations had different mean chromaticities, varying in appearance from yellow-green to green to blue-green across blocks 1 to 3. Some reference illuminations appear in multiple blocks, appearing bluer than the mean reference chromaticity in one block but yellower in another (larger/smaller hues). For each reference illumination, two staircases were completed, one starting at a yellower and one a bluer hue (between 24 and 30 degrees away clockwise/anticlockwise around the hue circle (smaller/larger hues), respectively). Importantly, staircases starting at smaller/larger hue values than the reference were free to travel to larger/smaller hue values, unlike a traditional staircase where possible values that the staircase can take are truncated at the reference value (cf. the staircase used for the discrimination task detailed later). If the participant responded by a "bluer" button press, the test stimulus hue value would always increase; if they responded with a "yellower" button press, the test stimulus hue would decrease. Within each block, staircases for all reference illuminations were interleaved (10 staircases or 300 trials per block). Block order was randomised across participants. While participants could break between the blocks for as long as they liked (with a top up adaptation required if they left the dark environment of the laboratory between blocks), they were required to complete all blocks within the same experimental session.

6.2.1.5 Data Analysis

The final value of each staircase (the test illumination hue on the 30^{th} trial) was taken as the estimated PSE. We took this value to avoid throwing away any data. If we used the mean of the last two reversals from each staircase one participant had an extremely large bias value from their yellow staircase for reference 1 in block 1 (participant 4023) as they did not reverse after the 9th trial. However, the staircase does appear to converge, ending at the same hue value as the blue staircase. Using the last value of each staircase rather

than the mean of the last two reversals maintains the general pattern in the data while reducing individual differences and ensuring all data is usable without correction. Bias was calculated as PSE hue value minus reference hue value, a positive bias indicating a shift in the PSE towards bluer (larger) hues and a negative bias indicating a shift towards yellower (smaller) hues. When calculated from the staircases that started at a test hue larger than the reference hue (the blue staircase), bias values were larger than those calculated from staircases beginning at smaller hue values (the yellow staircase) (Wilcoxen signed-ranks test: $z = 5.05$, $p < .001$). However, analysing the data from the blue and yellow staircases separately did not change the results of the experiment; therefore, all bias values presented in the main text are an average of the bias values from the respective blue and yellow staircases. Separate analyses of each set of staircases can be found in Appendix D.

6.2.2 Results

The calculated bias values for each reference hue angle are plotted with respect to block in Figure 6.2.A. Consider the data from block 1. These data show that the average PSEs for the two reference hues in this block that are the most "yellow" (lowest hue angle) are positively biased, or they are "bluer" (higher hue angle) than the true hue. Similarly, the average PSEs for the two most "blue" (highest hue angle) reference hues in block 1 are negatively biased, or are "yellower" (lower hue angle) than the true hue. However, the central reference hue in the block has a bias value close to zero, indicating that the PSE was close to the true hue value. This trend is maintained in blocks 2 and 3. As certain reference hues are used in multiple blocks, appearing "yellower" (smaller hue angle) than the mean reference in one block and "bluer" (larger hue angle) in another, the PSEs for these hues can be shown to vary with respect to block (e.g. red circles in Figure 6.2.A.). There are four such reference hues in the experiment (138, 144, 156 and 162 degrees). A 2×4 repeated measures ANOVA with direction of change from the mean reference hue angle in the block (to a larger or smaller hue, or becoming "bluer" or "yellower" in appearance; hereafter referred to as direction of change) and reference hue angle (138, 144, 156 or 162 degrees)

as independent variables and PSE as the dependent variable confirmed a significant main effect of direction of change on PSEs ($F(1, 10) = 105.73$, $p < .001$). When the reference hue angle is smaller ("yellower") than the central or mean reference hue angle in that block, the PSE hue value is always positively biased ("bluer") and vice-versa. In addition, this effect is not dependent upon the hue angle of the reference, with a non-significant direction of change and reference hue angle interaction term ($F(3, 30) = 0.06$, $p = .981$).

Following Olkkonen *et al.* (2014), we fit a regression line to the bias trend line for each individual participant in each block of the experiment. The slopes of the fitted regression lines can be used to quantify the bias, indicating the amount of central tendency bias (the amount of "pull" towards the mean reference hue angle) in each block. A slope value close to zero implies no bias, with PSEs close to the true hue value. Small negative slope values indicate increased bias, as the slope value becomes more negative bias increases. The slope of every fitted regression line is negative, suggestive of a central tendency bias in all blocks for all participants (Figure 6.2.B). Indeed, the average slope value is significantly different from zero in all three blocks ($t(10) = -7.67$, $p < .001$, $t(10) = -9.17$, $p < .001$ and $t(10) = -8.67$, $p < .001$, all Bonferroni corrected for blocks 1-3, respectively). In addition, a repeated measures ANOVA with block as the independent variable and fitted slope as the dependent variable shows that slope values did not differ across blocks ($F(2, 20) = 1.64$, $p = .219$).

6.2.3 Summary

Experiment 1 extends the results of Olkkonen *et al.* (2014) to show that hue estimates are susceptible to a central tendency bias not only for light-based surfaces but also for estimates of illumination hue. We also showed that illumination hue estimates depend on context, with estimates for a given reference hue differing significantly when embedded within a different series of reference hues.

The results cannot be explained by a response bias (observers favouring a response of yellow over blue or vice versa). If this was the case, then the bias would be the same

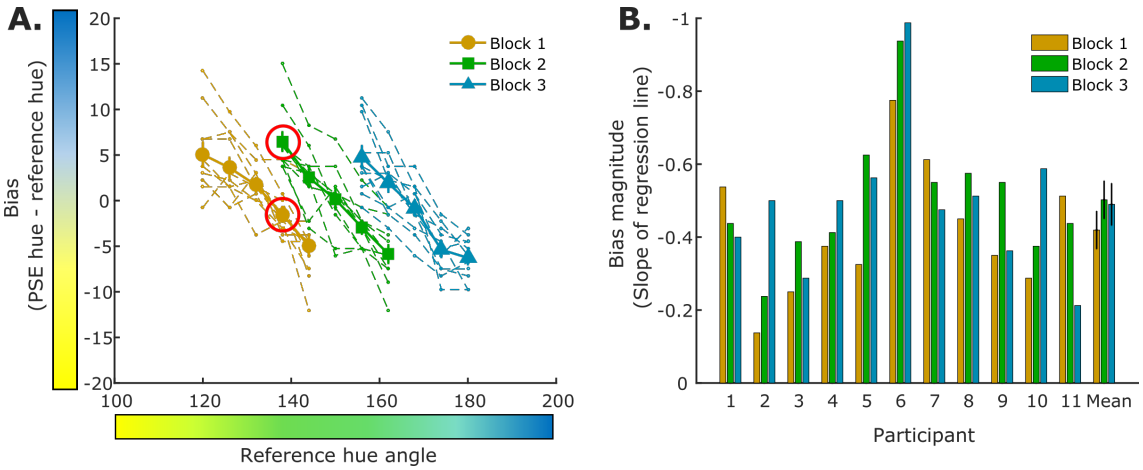


Figure 6.2: Central tendency bias in experiment 1. A. The bias for each reference hue angle with respect to block (1-3). Bias is calculated as PSE hue angle minus reference hue angle in CIE $L^*a^*b^*$, with a positive bias indicating that the PSE for a particular reference was "bluer" than the true hue and a negative bias indicating that the PSE was "yellower". Bold lines with circle, square or triangular markers are the mean bias trend lines over all participants. Thin dashed lines are individual bias trend lines for each participant. Red circles highlight an example of a case where the PSE for a given reference hue is shown to depend on block. B. The slopes of the regression lines fit to each individual bias trend line for each participant and the mean over these (error bars are ± 1 SEM). Note that as all slopes were negative, values on the y-axis have been flipped.

for all reference stimuli. Either all perceptual estimates would be bluer or all would be yellower. Our results indicate an almost 50/50 split which is characteristic of a central tendency bias.

6.2.4 Interim discussion

Despite the consistency of these results with those of Olkkonen *et al.* (2014), there is reason to believe that other factors may be responsible for the effect. During the task, movements of the staircase are dictated by the participant's responses to the test illuminations. Participants are instructed to respond according to whether they perceive the test illumination to be "bluer" or "yellower" than the reference illumination. For the staircase to converge on the PSE, it is assumed that when the participant cannot discriminate between the reference and test illuminations they are equally likely to respond with either a "bluer" or "yellower" button press. There is an alternative explanation of the participants behaviour, though. All illuminations used in the experiment are chromatic and could be

(even if somewhat ambiguously or uncertainly) assigned a colour name. Consider the scenario in which the participant cannot discriminate between the reference and test hue but is forced into a response of either "bluer" or "yellower". It is likely that the participant will issue the response that best describes the colour appearance of the test illumination under the reasoning that if the test appears blueish, then it is more likely to be bluer than the reference rather than yellower. If this is the case, then for the bluest reference hues, when the test is close to the reference in hue and discrimination is not possible, participants are more likely to issue a "bluer" response as the test appears blue in comparison to other illuminations within the current block. A string of such responses will bias the PSE towards hue values yellower than the true reference hue (a "bluer" response results in a decrease in the test illuminations hue value towards yellower hues), and will only stop when discrimination between the reference and test hue is possible. This will result in the PSE representing one extreme of the discrimination interval along the hue circle around the reference hue, the yellowest hue in the discrimination interval, manifesting as a central tendency bias. The same is true in the other direction, forcing PSEs for the yellowest reference hues to be biased towards blue. In sum, this would imply that the bias observed in the results is caused or at least exaggerated by the experimental design. In Experiment 2, we asked if the argument presented above explains the results of Experiment 1.

6.3 Experiment 2

In Experiment 2 we asked if using colour names to determine responses in Experiment 1 introduced an element of bias to the task that exaggerated the central tendency bias. We did so by collecting hue discrimination thresholds around each reference hue used in Experiment 1 for a subset of the original participants. These thresholds are found separately, without interleaving reference illumination conditions. In other words, we find typical discrimination thresholds for each reference illumination in each direction around the hue circle rather than PSEs (they should not be subject to a central tendency bias). In addition, we established the probability that participants would assign the colour names

"blue" or "yellow" (also "green" in the second version of the naming task) to each reference illumination within the context of each block of Experiment 1. We show that the colour name assigned to a given reference illumination is dependent upon context and is not a fixed property of colour perception. Combining these results with the hue discrimination data, we show that participant responses to Experiment 1 can be partially explained by an adjustment of colour categories and that our finding of a central tendency bias may be caused or at least exaggerated by the design of the task.

6.3.1 Methods

6.3.1.1 Participants

All participants who completed Experiment 1 were invited back to take part in Experiment 2. Only 7 (6 female, mean age 23 $\pm$ 3 years) returned. Each participant received further compensation for their time.

6.3.1.2 Apparatus and Stimuli

We used the same apparatus and stimuli as in Experiment 1. For the discrimination task, the reference and test hues were identical to those generated for Experiment 1. For the naming task, the 11 reference illuminations were combined with 12 intermittent hues taken from the set of test illuminations for a finer sampling of the stimulus space.

6.3.1.3 Procedure: The naming task

Two versions of the naming task were completed (the 2-naming and 3-naming tasks described below). In both versions, participants were required to name the subset of reference illuminations used for each block of Experiment 1 along with six equally spaced (hue steps of $\approx$ 3 degrees) intermittent illuminations taken from the set of test illuminations (11 illuminations per block; Figure 6.3). Accordingly, the naming tasks were also split into three blocks, completed in the same order as the participants completed the blocks in Ex-

periment 1. Prior to starting the task, participants viewed the adapting illumination for 2 minutes. On each trial, the illumination to be named was presented for 500 ms before the adapting illumination reappeared while the participant issued their response via a button press (no time limit). Each illumination in a block was named 20 times in a random order. In the 2-naming task, participants could name the illuminations either "blue" or "yellow". In the 3-naming task, illuminations could be named either "blue", "yellow" or "green". As in Experiment 1, participants could break between the blocks for as long as they liked but were required to complete all blocks within the same experimental session. Although not a requirement, all but one of the 7 participants who returned to complete this task completed both versions within one visit to the laboratory. The order of task completion (2-naming or 3-naming task first), was randomised across participants.

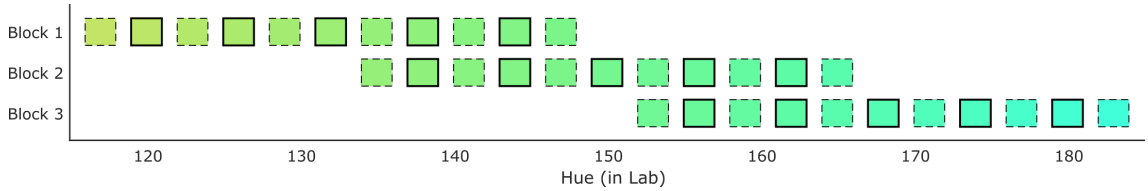


Figure 6.3: A cartoon of the stimuli used in the naming tasks.

6.3.1.4 Procedure: Discrimination task

The discrimination task (Figure 6.4) was designed to find hue discrimination thresholds in each direction (clockwise/anticlockwise) around the hue circle for each reference illumination. To achieve this, discrimination thresholds for each reference illumination were found in separate runs (only one reference illumination was presented per run to avoid eliciting a central tendency bias). Again, the task was preceded by a 2 minute adaptation period. On each trial, participants were presented with a reference illumination (500 ms), followed by a 2000 ms delay under the adapting illumination before the presentation of two comparison illuminations (each for 500 ms, separated by 400 ms of the adapting illumination). One comparison illumination would always be identical to the reference illumination (the target illumination) and one would differ in hue from the reference illu-

mination (the test illumination, taken from the appropriate set of test illuminations used in the Experiment 1) according to a 1-up, 1-down staircase procedure (fixed step size of 1, terminating after 30 trials or 6 reversals, starting between 15 and 22.5 degrees away from the reference illumination hue). Whether the target illumination appeared as the first or second comparison was randomised on each trial. After presentation of both comparisons, the adapting illumination reappeared and participants were required to press a button that indicated whether they thought the first or second comparison was the best match to the reference (no time limit). If they chose the target illumination the staircase stepped closer (in hue value) to the reference illumination, otherwise, it stepped further away. For each reference illumination, two interleaved staircases were completed, one staring from a bluer (larger) and one from a yellower (smaller) hue value. This comprised one run, with 11 runs (one for each reference illumination) in total for the task. Runs were completed in a random order. Again, while participants could break between runs, they were required to complete all runs within one session. All but one of the 7 participants who returned completed the task twice on two separate occasions (different days). One participant who completed the task twice did not finish the task in one session and had missing data for one of the reference hues (162 degrees, effectively they only completed the task once for this reference hue).

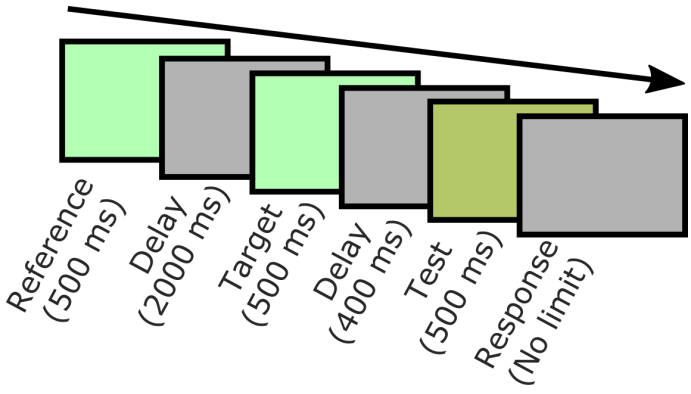


Figure 6.4: The discrimination task. The reference illumination was presented for 500 ms. After a 2000 ms delay under the adaptation illumination, two comparison illuminations were presented for 500 ms each, separated by 400 ms of the adaptation illumination. One of the comparisons was identical to the reference (the target), the other differed from the reference (the test). Participants had to indicate which comparison was most similar to the reference.

6.3.1.5 Data analysis and modelling

For the discrimination task, thresholds were calculated as the mean of the last two reversals from the staircase. Where possible, thresholds for any one staircase were averaged over the two runs of the task. Average thresholds from the two types of staircases (starting at larger or smaller hue values) for each reference illumination formed a discrimination interval for each participant, or, more accurately, a no-discrimination interval within which a change from the hue of the reference illumination could not be detected.

The naming tasks resulted in a distribution of naming probabilities, the probability that each illumination would be named either "blue" or "yellow" (or "green" in the 3-naming task), defined as the number of times a given colour term was used divided by the total number of presentations (20 in all cases). The illuminations used in the naming tasks were sampled in hue steps of 3 degrees while stimuli used to determined thresholds in the discrimination task and PSEs in Experiment 1 were sampled at a higher resolution (every 1.5 hue steps). To obtain naming probabilities for each hue value used in the latter two tasks, naming probabilities for each block in the 2-naming task were interpolated (separately for each participant) to estimate the probabilities of naming intermittent stimuli as either "blue" or "yellow".

Data from the discrimination and naming tasks were combined as follows to assess whether the central tendency bias observed in Experiment 1 can be explained by these results alone. For each participant, we simulated their response to the task used in Experiment 1. On each trial, the simulated response was decided differently depending on whether the test illumination was within the current reference illumination's discrimination interval or not. If the test illumination was outside the discrimination interval, then it is assumed that the participant can perceive the difference between the test and reference illuminations. More specifically, it is assumed that they can perceive the correct difference between them such that when the test is bluer (larger hue value) than the reference they respond "bluer" and when the test is yellower (smaller hue value) than the reference they respond "yellower". However, if the test illumination is within the discrimination interval then it is assumed

that the participant cannot discriminate and bases their response on how they perceive the test illumination. How they perceive the test illumination is represented by the 2-naming task data, so they would name it "blue" (respond "bluer") with probability p and name it "yellow" (respond "yellower") with probability $1-p$, with p taken from the data collected in the 2-naming task.

6.3.2 Results

6.3.2.1 The Naming Tasks

Figure 6.5.A shows the data from the 2-naming task and illustrates that the most likely colour name assigned to a given reference hue changes across the blocks (top plot is block 1, middle plot is block 2 and bottom plot is block 3). Consider, for example, the reference illumination with hue value 141 degrees. In block 1, participants are most likely to name this illumination "blue" on average while in block 2 they are most likely to name it "yellow".

These data suggest that participants may adjust their colour category boundaries to compensate for the small range of stimuli used in the different blocks, allowing for a more optimal use of colour terms to partition the stimulus space that they are faced with. If the colour names that participants assigned to an illumination with given hue value did not change from blocks 1 to 2, then almost all stimuli in block 2 would be named "blue" and hence colour terms would not effectively distinguish between the stimuli. Adjusting the colour category boundaries may aid discrimination in this way.

The majority of stimuli used in the task appeared green to the experimenters. If this is also the case for the participants then the claim that responses in Experiment 1 are based on the colour appearance of the test illumination may not be valid (as they always perceive it as green and neither "bluer" nor "yellower" are good descriptors of the perceived colour). To determine whether this argument might hold, we conducted a second version of the naming task for all participants, in which they were also allowed to use the colour term

"green" (the 3-naming task). Figure 5.B shows the data from the 3-naming task and illustrates that even when the colour term "green" can be used in addition to "blue" and "yellow", participants still use the latter two terms and how terms are assigned still varies in the different blocks.

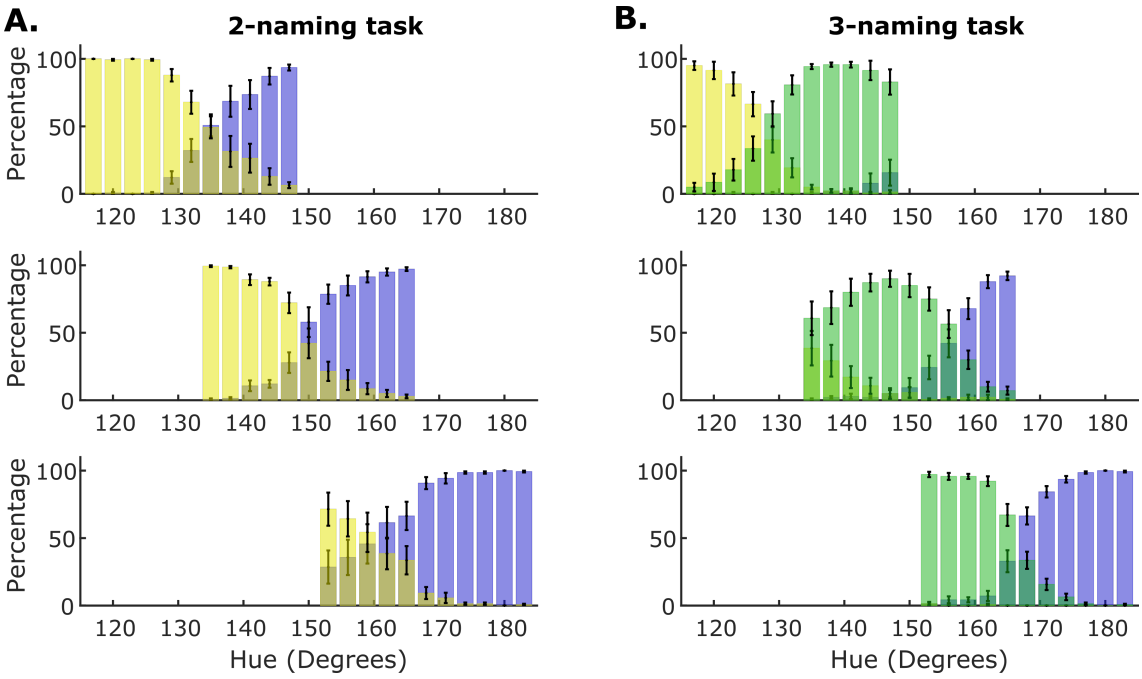


Figure 6.5: Naming tasks data. Mean naming probabilities across all participants in the 2-naming (A) and 3-naming (B) tasks. Yellow bars represent the probability of naming that reference hue yellow, blue bars represent the probability of naming that reference hue blue and, for the 3-naming task, green bars represent the probability of naming that reference hue green. The top row of figures represent colour name assignment in block 1, the second in block 2 and the third in block 3. Error bars are ± 1 SEM.

6.3.2.2 The Discrimination Task

Average discrimination thresholds (across both directions of change towards bluer and yellower hues) were 5.96±3.43 degrees (Figure 6.6). For bluer (larger) reference hues, discrimination thresholds towards bluer hues seem to increase while discrimination thresholds for yellower hues decrease relative to those for yellower (smaller) hues. However, a 2×11 repeated measures ANOVA with direction of change (bluer or yellower) and reference hue value as independent variables and discrimination threshold as the dependent vari-

able found no significant interaction effect of direction of change and reference hue value ($F(10, 60) = 1.91$, $p = .061$). Note though that this study was not designed to look for such an effect and hence is underpowered with such a small sample size ($n = 7$). Importantly, the average size of thresholds (5.96) is supportive of our hypothesis that the central tendency bias in Experiment 1 is exaggerated by responding with the colour term most representative of the perceived colour of the test illumination as the range of average bias values in Experiment 1 (Figure 6.2.A, $[-6.27, 6.48]$) suggests most PSEs fall within these non-discriminable intervals. For each participant, we checked if the PSE for each reference hue in each block in Experiment 1 fell within the discrimination interval for that reference. On average 72.38±9.76% of PSEs for each participant fell within the discrimination intervals.

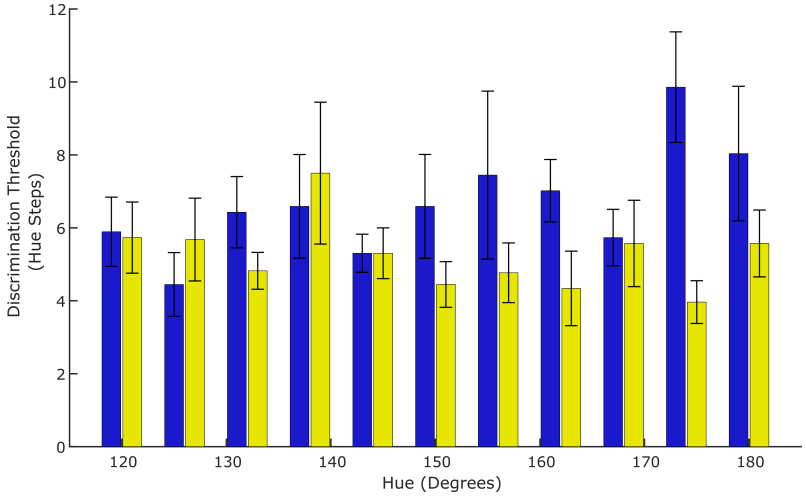


Figure 6.6: Discrimination task data. Discrimination thresholds (in hue steps, or the difference between the termination point of the staircase and the reference hue in degrees) for each reference hue used in Experiment 1 in CIELAB. Blue bars (leftmost of the pairs) represent the discrimination thresholds for bluer (larger) hue values. Yellow bars (rightmost of the pairs) represent the discrimination thresholds for yellower (smaller) hue values.

6.3.2.3 Simulating Responses to Experiment 1

We combined the data from the discrimination and 2-naming tasks to simulate responses to Experiment 1. As the simulations are probabilistic, multiple simulations will result in different predictions for Experiment 1. For this reason, we ran 10,000 simulations to

get an estimate of how well the model accounts for the data collected in Experiment 1. Figure 7 shows an example simulated data set (1 of the 10,000). The simulated data shows a similar overall pattern to the real results (compare Figure 6.7.A and Figure 6.2.A). However, the fitted slope values suggest the model only partially accounts for the data, with some slope values now positive and a difference in slope values that is block dependent (but note a non-significant main effect of block for this particular simulation, $F(2, 12) = 2.85$, $p = .097$). To quantify how well the model predicts the results of Experiment 1, we calculated Spearman's correlation coefficients between the real data collected in Experiment 1 and the 10,000 simulated data sets. The mean correlation was $r_s = .66 \pm .02$ (for the simulated data set shown in Figure 6.7 it was $r_s = .65$), suggesting that while the simulated data accounts for the real data reasonably well, there is still a proportion of variability in the real data that is unaccounted for.

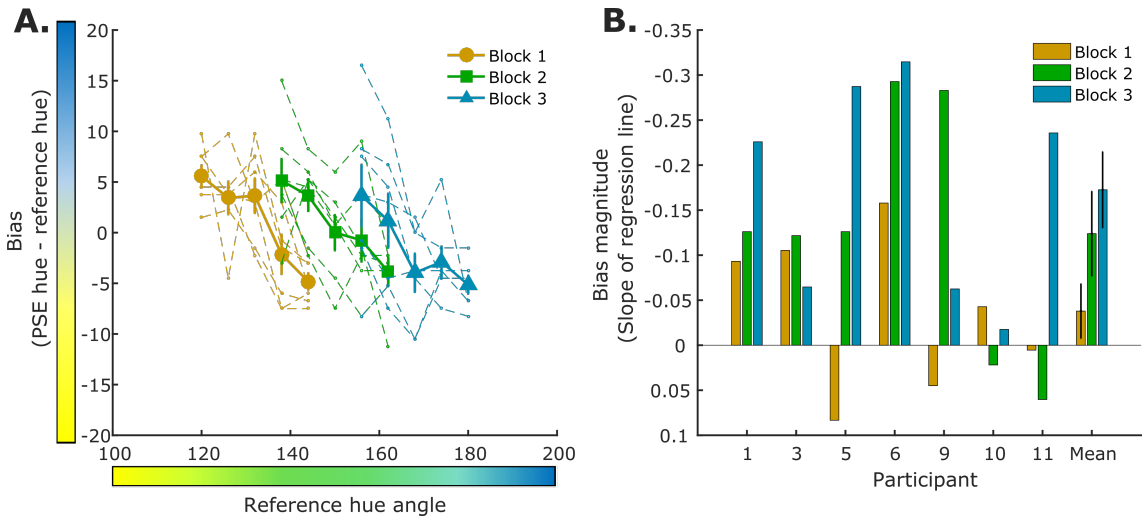


Figure 6.7: Simulated responses to Experiment 1. A. The simulated bias for each reference hue angle with respect to block (1-3). Simulated bias is calculated as simulated PSE hue angle minus reference hue angle in *CIE* $L^*a^*b^*$, with a positive bias indicating that the PSE for a particular reference was "bluer" than the true hue and a negative bias indicating that the PSE was "yellower". Bold lines with circle, square or triangular markers are the mean bias trend lines over all participants. Thin dashed lines are individual bias trend lines for each participant. B. The slopes of the regression lines fit to each individual bias trend line for each participant and the mean over these (error bars are ± 1 SEM). Note values on the y-axis have been flipped for easy comparison to Figure 6.2.

6.3.3 Summary

The results of Experiment 2 suggest that the association of colour terms to the responses in Experiment 1 may have caused or at least exaggerated the central tendency bias we observed. The naming tasks showed that the assignment of colour names to illuminations is not fixed and that the visual/cognitive system adapts to the current context in a way that may optimise colour term usage to partition the current stimulus space. By establishing a discrimination interval for each reference illumination, we also showed that while PSEs display a central tendency bias in Experiment 1, they fall inside the participant's discrimination interval 72.38% of the time, suggesting that something other than perceived colour contributes to the bias. By combining the results from the 2-naming and discrimination tasks, we were able to predict the emergence of a central tendency bias in Experiment 1, although the variability in the data is not fully explained.

6.4 Experiment 3

In Experiment 3, we take this a step further by changing the experimental design of Experiment 1 such that colour terms are no longer linked to the responses. In the new task, we separate reference stimuli into overlapping blocks as in Experiment 1. Now however, instead of determining PSEs by asking participants to respond by indicating if a test illumination is "bluer" or "yellower" than the reference, we show participants two test illuminations and ask them which is "most similar to the reference". The test illumination that is picked remains the same on the next trial of that staircase (not a staircase in the traditional sense so we refer to it hereafter as a pseudo-staircase). The test illumination not chosen steps closer in hue value to the chosen option. This pseudo-staircase is designed to target the PSE without the use of colour terms to issue responses. We hypothesise that this will suppress or at least reduce the magnitude of the central tendency bias observed in Experiment 1 (fitted regression lines will have shallower gradient).

6.4.1 Methods

6.4.1.1 Participants

Eleven participants (eight female, mean age 24±3 years) participated in Experiment 3. Three had previously participated in Experiments 1 and 2, the remaining eight participants were recruited through word of mouth. All participants had normal or corrected to normal visual acuity and gave consent prior to participation. All were assessed for normal colour vision using the Ishihara colour plates and the Farnsworth-Munsell 100 hue test. Each participant received a voucher as compensation for their time.

6.4.1.2 Apparatus and Stimuli

We used the same apparatus and stimuli as in Experiment 1.

6.4.1.3 Procedure

The design of the experiment was similar to Experiment 1 in that there were three separate blocks of the task with five reference hues interleaved in each block (Table 6.1). However, in this experiment we did not use a conventional staircase to estimate PSEs (we refer to it as a pseudo-staircase instead) and there were two comparison illuminations on every trial. On each trial participants were first presented with a reference illumination (500 ms) then, after a 2 second delay, they were presented with comparisons one (500 ms) and two (500 ms), separated by a small delay (400 ms). After viewing comparison two participants were required to respond by indicating "which of the two comparisons was most similar to the reference". During the response and delay periods, participants viewed the adapting light (Figure 6.8). It was not a necessity that either comparison illumination was equivalent to the reference. At the beginning of each staircase, one comparison was yellower (smaller hue value) than the reference and the other was bluer (larger hue value). The two comparisons were an equal number of hue steps away from the reference at the start of the staircase but how far away they were was different for each staircase (randomly

generated to start between 15 and 34 hue steps away). The comparison chosen by the participants ("one" or "two") remained the same on the next trial. The comparison not chosen stepped towards the chosen comparison (one nominal step, equivalent to a hue step of 1.5). In other words, if the most yellow (smallest hue value) stimuli was chosen, the one not chosen became yellower (hue value decreased) on the next trial and vice-versa. On trials where both comparisons were identical, the hue value of the non-chosen comparison increased or decreased with probability 0.5. All staircases terminated after 40 trials.

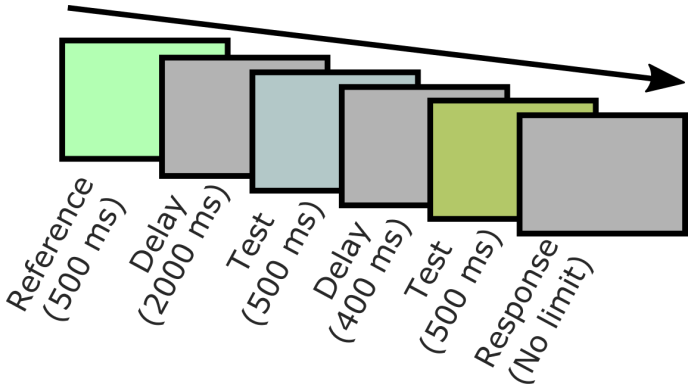


Figure 6.8: Experiment 3 overview. Participants were first shown a reference illumination (500 ms) then, after a short delay, (2000 ms) they were presented with two test illuminations (500 ms each separated by 400 ms of the adapting illumination). The participant had to chose the test that was most similar in appearance to the reference.

6.4.1.4 Data analysis

PSEs for each reference were calculated as the mean of the last 10 chosen comparison hue values in that reference's pseudo-staircase. Bias was calculated as PSE hue value minus the reference hue value.

6.4.2 Results

The calculated bias values in Experiment 3 are very different to those we collected in Experiment 1 and to those predicted in Experiment 2 (Figure 6.9.A). The pattern of bias values do not suggest a central tendency bias. Repeating the analysis of Experiment 1, a 2×4 repeated measures ANOVA with direction of change and reference hue angle as

independent variables and PSE as the dependent variable showed that direction of change did not affect PSEs for the four repeated hues ($F(1, 10) = 0.15$, $p = .707$). In addition, the fitted slope values are not significantly different from zero for any reference block ($t(10) = -1.26$, $p = .235$, $t(10) = -0.195$, $p = .849$ and $t(10) = -0.335$, $p = .744$, for blocks 1-3, respectively), nor did slope values differ across blocks ($F(1.33, 13.26) = 0.46$, $p = .563$, with a Greenhouse-Geisser correction).

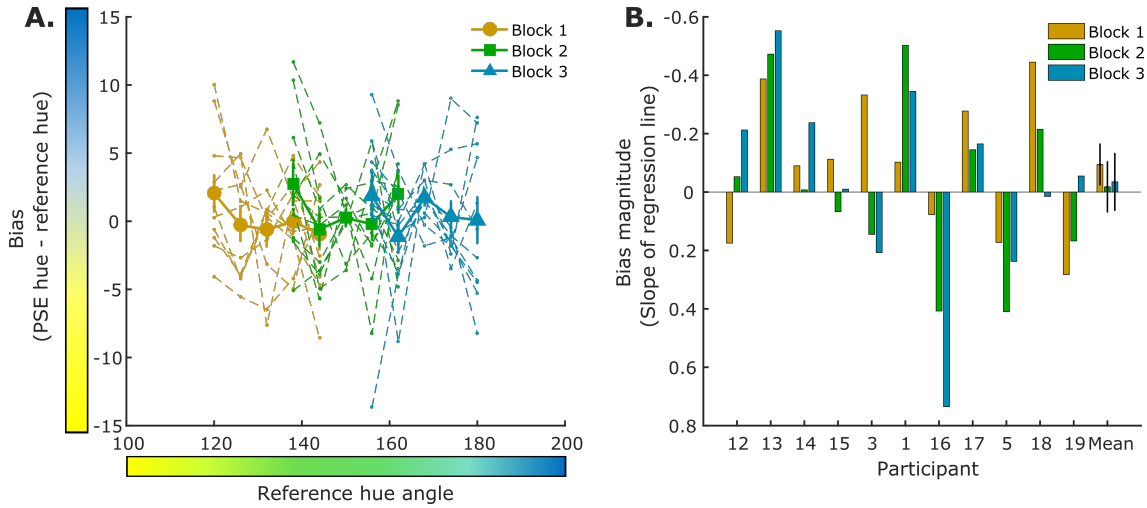


Figure 6.9: Results of Experiment 3. A. The bias for each reference hue angle with respect to block (1-3). Bias is calculated as PSE hue angle minus reference hue angle in $CIE\ L^*a^*b^*$, with a positive bias indicating that the PSE for a particular reference was "bluer" than the true hue and a negative bias indicating that the PSE was "yellower". Bold lines with circle, square or triangular markers are the mean bias trend lines over all participants. Thin dashed lines are individual bias trend lines for each participant. B. The slopes of the regression lines fit to each individual bias trend line for each participant and the mean over these (error bars are ± 1 SEM). Note values on the y-axis have been flipped for easy comparison to Figure 6.2.

6.4.3 Summary

Experiment 3 does not provide any evidence for a central tendency bias in perception of illumination hue; average PSEs are close to the true hue values of the reference illuminations. Although the individual traces in Figure 6.9.A show that PSEs are quite variable across the participants, note that the majority of PSEs fall within the discrimination interval limits established in Experiment 2, suggesting that PSEs are generally falling within a range of non-discriminable hue values.

6.4.3.1 Interim discussion

It would be rash to conclude from the results of Experiments 2 and 3 that a central tendency bias does not exist for illumination hue perception; the explanation offered here (that the central tendency is caused by categorisation or naming of stimuli) cannot account for the central tendency bias observed in different domains where different experimental paradigms were used. For example, when Huttenlocher *et al.* (2000) illustrated a central tendency bias for perceptual estimates of absolute size, line length and shades of grey they required participants to adjust the test stimulus to reproduce the reference. To be sure, we ran a third and final experiment that used a reproduction task to establish PSEs.

6.5 Experiment 4

Experiment 4 again involves the use of a task designed to ask the same questions as Experiment 1, but without assigning colour names to the responses. In this task, PSEs for the set of reference hues within block 2 are established using a reproduction task. On each trial, participants see the reference illumination but now, after the delay, the test illumination can be adjusted by the participant until they judge its appearance to match that of the reference. Pilot testing revealed this to be a hard task for participants, and for this reason, participants were trained to complete the task by completing three sessions.

6.5.1 Methods

6.5.1.1 Participants

Thirteen participants (9 female, mean age 25 $\pm$ 4 years) were recruited for Experiment 4. None of these participants had previously taken part in Experiments 1-3. Two participants (both female) did not complete the required three repeats of the task, only doing the first session. All participants had normal or corrected to normal visual acuity and gave consent prior to participation. All were assessed for normal colour vision using the Ishihara colour

plates and the Farnsworth-Munsell 100 hue test. Each participant received a voucher as compensation for their time.

6.5.1.2 Apparatus and stimuli

We used the same apparatus and stimuli as in Experiments 1 and 3.

6.5.1.3 Procedure

The design of the experiment was similar to Experiments 1 and 3. On each trial, participants were presented with the reference illumination for 500 ms. After a 2000 ms delay under the adapting illumination, participants were presented with the test illumination. The test illumination was on for 2000 ms, off for 400 ms (replaced by the adapting illumination), on for 2000 ms, off for 400 ms and so on until the participant indicated that they had matched the test to the reference illumination (Figure 6.9). During periods when the test illumination was on, participants could use two buttons to adjust its hue. One button increased the test hue (it became bluer), the other decreased it (it became yellower). Participants matched each reference hue 40 times with matches for all reference hues interleaved. The starting hue value for the test was randomly chosen on each trial. As the time demands of this task were greater than of Experiments 1 and 3, and as participants found the task difficult prior to training in pilot tests, the task was only completed for one block of reference hues from Experiments 1 and 3 (block 2; Figure 6.1.C; Table 6.1). However, participants repeated the task three times on three separate days. We expected that participants' matches would become less variable over the course of the repeats as they became more familiar with the task. Experimenters avoided the use of colour terms when explaining the task to the participants (i.e. they told participants that the two buttons would adjust the appearance of the test illumination but did not say that they would make the test illumination bluer or yellower). We believed this to be an important aspect of task design considering our conclusions from Experiment 2; we did not want to encourage participants to assign colour terms to their responses or to categorise colours

as a method for completing the task.

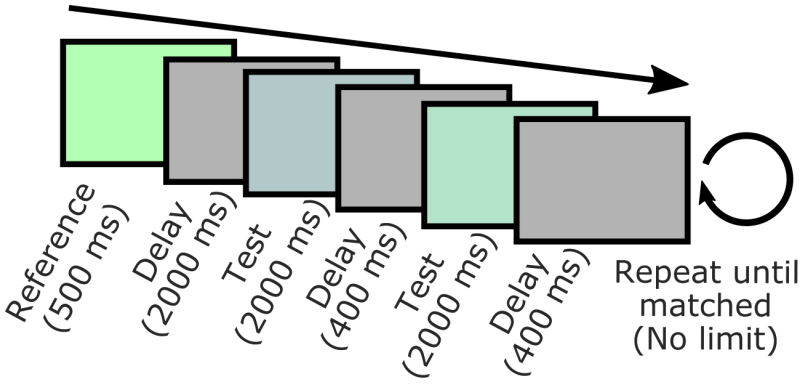


Figure 6.10: The reproduction task. On each trial, participants were presented with the reference illumination for 500 ms. After a 2000 ms delay under the adapting illumination, participants were presented with the test illumination. The test illumination was on for 2000 ms, off for 400 ms (replaced by the adapting illumination), on for 2000 ms, off for 400 ms and so on until the participant indicated that they had matched the test to the reference illumination.

6.5.1.4 Data analysis

PSEs were taken as the mean over the 40 matches for each reference hue. Bias was calculated as PSE hue minus reference hue value.

6.5.2 Results

6.5.2.1 Match variability decreases with training

To assess whether the variability of the matches made by the participants decreased with training we obtained the standard deviation of each participant's bias for each reference illumination across the repeated runs of the task (Figure 6.11). For all reference illuminations, the data show a similar trend; the standard deviation of the participant's bias seems to decrease with training (a decrease from first to third run). Indeed, a 3×5 repeated measures ANOVA with run (3 levels) and reference (5 levels) as the independent variables and standard deviation of the bias as the dependent variable finds no main effect of reference ($F(4, 40) = 1.04$, $p = .400$) or interaction effect of reference and run ($F(8.80) = 1.13$, $p = .350$), but there was a significant main effect of run ($F(2, 20) = 4.47$, $p = .025$).

However, no pairwise comparisons between the average standard deviations, regardless of reference illumination, were significant ($p > .05$ in all cases).

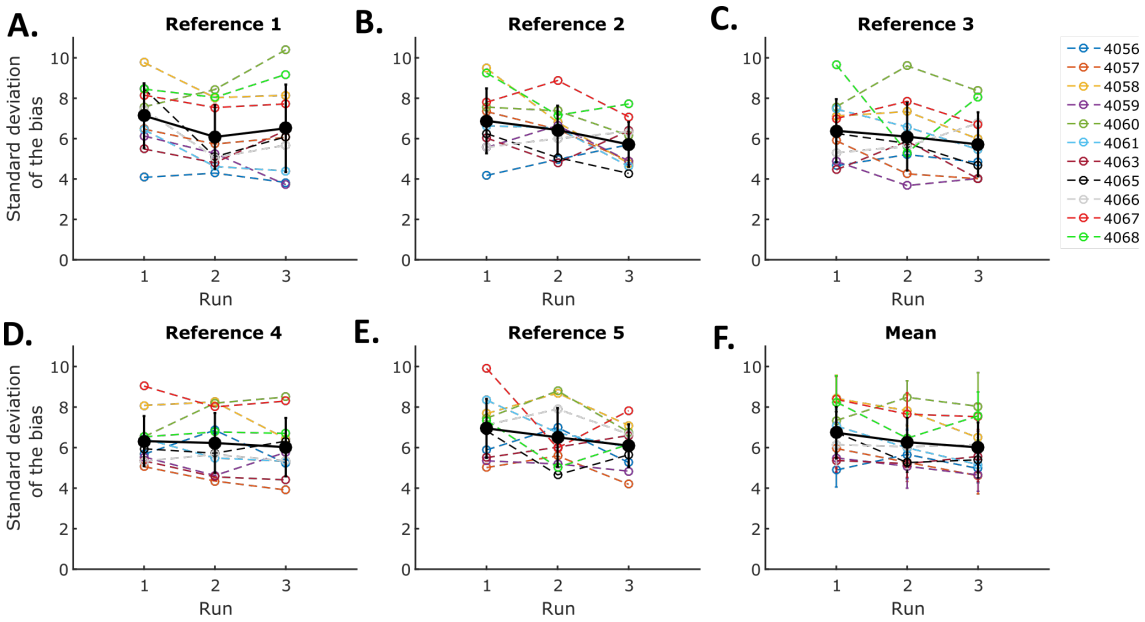


Figure 6.11: Reproduction task training effects. A-E. The standard deviation of each individual participant's matches across the repeated runs of the task for each reference illumination (coloured dashed lines) with the mean over all participants superimposed (solid black line). F. The mean standard deviation of each individual participant's matches over all reference illuminations across the repeated runs of the task (coloured dashed lines) with the mean over all participants superimposed (solid black line). Error bars are ± 1 SEM.

6.5.2.2 A bias emerges in the third run of the task

As we found a main effect of run on the variability of bias values within an individual, we analysed the data from each run of the task separately. We expect that any bias present in the results will be more prevalent in the data from the third run where the bias values are less noisy. Two extra participants completed the first run of the task (only 11 did all three runs) and we include these participants in the analysis of the first run here.

The average slope value is negative for all runs of the task (Figures 6.12-6.14). However, the average slope value is only significantly different from zero in the third run of the task ($t = -1.72$, $p = .112$, for run 1; $t = -1.92$, $p = .084$, for run 2; $t = -3.06$, $p = .012$, for run 3). In the first run of the task, only 8/13 participants have the negative slope value

associated with a central tendency bias. However, after training, in the third run of the task, slope values are consistently negative; 10/11 participants show a central tendency bias. Note though that the magnitude of this bias is much smaller than the magnitude of the bias observed in Experiment 1. The average slope value here is -0.14 ± 0.15 (M $\pm$ SD) compared to -0.50 ± 0.18 (M $\pm$ SD) in Experiment 1.

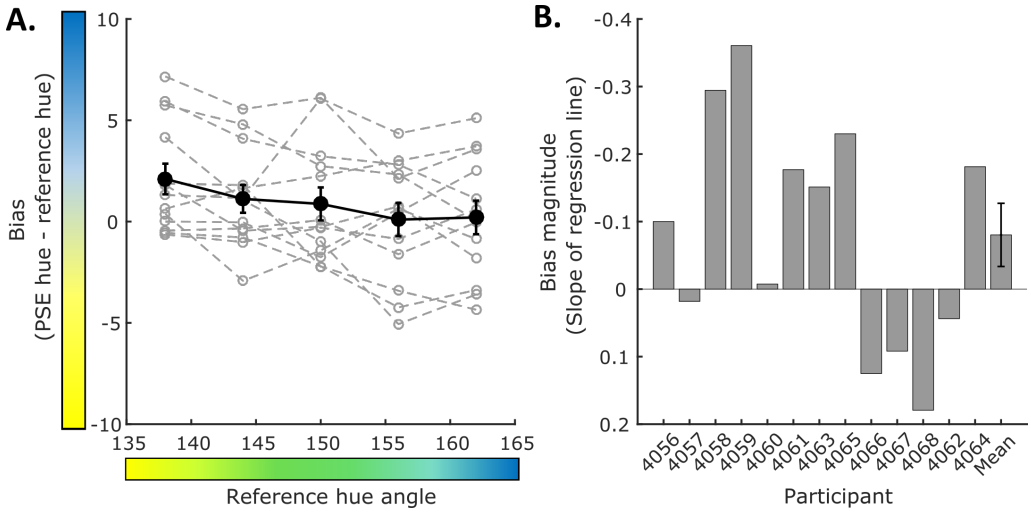


Figure 6.12: The bias in the first run of Experiment 4. A. The bias for each reference hue. Bias is calculated as PSE hue angle minus reference hue angle in $CIE\ L^*a^*b^*$, with a positive bias indicating that the PSE for a particular reference was "bluer" than the true hue and a negative bias indicating that the PSE was "yellower". Bold line is the mean bias trend line over all participants. Thin dashed lines are individual bias trend lines for each participant ($n = 13$). B. The slopes of the regression lines fit to each individual bias trend line for each participant and the mean over these (error bars are $\pm$ 1 SEM). Note values on the y-axis have been flipped for easy comparison to previous figures.

6.6 Discussion

Using the same task as Olkkonen *et al.* (2014) we showed a central tendency bias for illumination hue perception (Experiment 1). Previous demonstrations of a central tendency bias for other types of stimuli are well-accounted for by Bayesian models of perception (Huttenlocher *et al.*, 2000; Jazayeri & Shadlen, 2010). If we consider the central tendency bias observed in Experiment 1 in the same framework, then we found evidence that human observers can learn an illumination prior in a short space of time, a tool that may aid colour constancy. However, in further experiments (Experiment 2) we found evidence

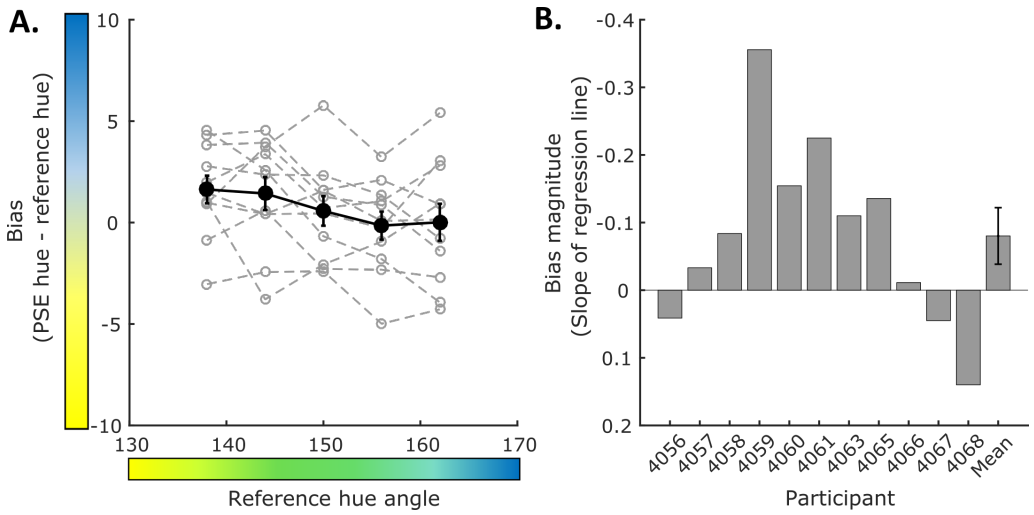


Figure 6.13: The bias in the second run of Experiment 4. A. The bias for each reference hue. Bias is calculated as PSE hue angle minus reference hue angle in $CIE\ L^*a^*b^*$, with a positive bias indicating that the PSE for a particular reference was "bluer" than the true hue and a negative bias indicating that the PSE was "yellower". Bold line is the mean bias trend line over all participants. Thin dashed lines are individual bias trend lines for each participant ($n = 11$). B. The slopes of the regression lines fit to each individual bias trend line for each participant and the mean over these (error bars are ± 1 SEM). Note values on the y-axis have been flipped for easy comparison to previous figures.

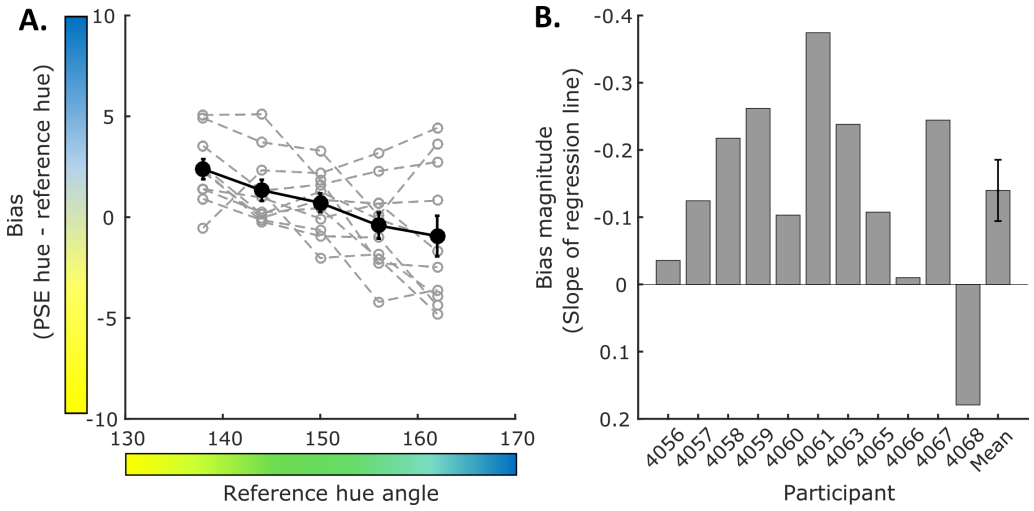


Figure 6.14: The bias in the third run of Experiment 4. A. The bias for each reference hue. Bias is calculated as PSE hue angle minus reference hue angle in $CIE\ L^*a^*b^*$, with a positive bias indicating that the PSE for a particular reference was "bluer" than the true hue and a negative bias indicating that the PSE was "yellower". Bold line is the mean bias trend line over all participants. Thin dashed lines are individual bias trend lines for each participant ($n = 11$). B. The slopes of the regression lines fit to each individual bias trend line for each participant and the mean over these (error bars are ± 1 SEM). Note values on the y-axis have been flipped for easy comparison to previous figures.

that the bias we observed may be caused or at least exaggerated by an element of the task where observers are required to link their responses to colour terms; observers were asked: is the test "bluer" or "yellower" than the reference? We showed that as the stimulus range presented within a particular block of the experiment changes, the way that participants assign colour terms to the stimuli also changes. It seems that participants may assign colour names differently in order to optimally partition the stimulus space. If we assume that participants resort to using these colour terms to issue responses on the task when the reference and test illumination become indiscriminable, then we can predict the central tendency bias. We then changed the method by which we estimated PSEs for the reference illuminations after the short delay in Experiment 1 such that colour terms were no longer linked to the responses (Experiment 3). Changing the task in this way eliminated the central tendency bias. However, the data were particularly noisy and previous central tendency bias studies find an effect using reproduction tasks that also avoid assigning categorical terms to the responses. Finally, we also used a reproduction task to estimate PSEs for the reference illuminations after the delay and found evidence of a central tendency bias but only after participants were trained on the task. In addition, the magnitude of the bias was much smaller than that observed in Experiment 1.

Considering the findings of Experiments 1-4 together, we find evidence of a central tendency bias for illumination hue perception (and evidence that observers can learn a prior over illumination hue if we adopt the Bayesian explanation for the effect), but we also find that the magnitude of the bias is likely exaggerated in Experiment 1 due to the assignment of colour names to the responses. Interestingly, the central tendency bias only emerges in Experiment 4 in the final run of the task after participants have been trained. It could be that the residual bias is still caused by participants assigning colour names to the stimuli and response options during the task. Following completion of Experiment 4, we asked participants to fill in a short questionnaire designed to assess if participants used a particular strategy to complete the task (Figure D.3). Only seven of the eleven participants who completed the three runs of the task filled in the questionnaire. Of these

seven, four self-reported the use of colour names during the task; either as a method for encoding the reference, as labels assigned to the response buttons, or both. One participant (participant 4059) said that they adopted this strategy for the second and third runs of the task but did not use it in the first run. However, whether or not participants adopt a strategy of this kind is not predictive of the level of bias they display in the third run of Experiment 4. For example, the participant with the most negative slope value (participant 4061) claimed to only take a visual snapshot of the reference and adjust the test until its appearance matched the remembered reference. Conversely, the participant whose slope remains positive throughout all runs of the task (participant 4068) reported using the colour names strategy. It seems then that the residual bias seen in the final run of Experiment 4 is independent of categorisation effects (using colour names during the task).

Why though do we only find significant evidence of a central tendency bias in the final run of Experiment 4 but not in the first two runs or in Experiment 3? In Experiment 3, we changed the temporal statistics of the experimental stimuli; two test lights are now seen per trial rather than one (as in Experiment 1) and could make formation of a prior over reference illuminations more difficult given the added interference of an extra illumination view on each trial. A similar argument may explain the lack of bias in the first two runs of Experiment 4. As participants can adjust the test, they view it for much longer than the reference on each trial. In addition, the test illumination is flashing on and off and this could be considered multiple illumination presentations. On top of that, participants are adjusting the test illumination. Again, this could hinder formation of a prior over the reference illumination parameters due to the extra interference. It may be that participants carry on learning from one repeat of the task to the next in Experiment 4, however, and that they have not formed an adequate prior until the final run of the task. Indeed, once a participant begins to show a negative slope in one of the task runs, they do not go back to a positive slope, with the exception of participant 4056.

We must also ask why we do not find evidence of a central tendency bias in Experiment

3. In Experiment 3, observers must hold not only the reference in memory, but also the first comparison illumination. Hence, when making the decision of whether comparison 1 or 2 is most similar to the reference, comparison 1 may be subject to the same memory reconstruction bias (by a learnt prior) as the reference and this could be what produces such noisy data. To test this explanation, this experimental paradigm could be used on a monitor for perceptual estimates of coloured patches when both comparisons can be presented simultaneously.

Our results can be compared directly to those of Olkkonen *et al.* (2014) as the stimuli were controlled such that they produce the same retinal excitation. Our data from Experiment 1 are suggestive of a slightly stronger central tendency bias than their data (an average slope of approximately -0.3 in Olkkonen *et al.* (2014) across the three blocks compared to -0.47 in Experiment 1). However, in the final run of Experiment 4, we find an average slope value of -0.14, suggestive of weaker central tendency bias than that of Olkkonen *et al.* (2014). It could be that the task used by Olkkonen *et al.* (2014) is not as susceptible to the effects of categorisation (like the task used in Experiment 1) due to one small difference between the tasks. In our Experiment 1, we ask participants if the test illumination was bluer or yellower than the reference. Olkkonen *et al.* (2014) asked participants whether the test or reference was bluer. While it would be easy to extend the model we present in Experiment 2 to predict data for the task used by Olkkonen *et al.* (2014), we do not have naming or discrimination data for these individuals so cannot ask how much of the bias can be explained by this model as we did for our task in Experiment 2.

An emerging stream of literature has illustrated a perceptual bias towards the most recently viewed stimulus, a one-back bias in perceptual estimates (although note that studies report a dependence on stimuli seen further back in the stimulus stream; for example, up to three-back (Fischer & Whitney, 2014)). This type of bias is treated separately in the literature from a central tendency bias (a bias towards the stimulus one trial back compared to a bias towards the mean of all stimuli within a series) and has been coined serial dependence. Serial dependence is seen in estimates for a range of stimulus characteristics

such as numerosity (Cicchini *et al.*, 2014; Corbett *et al.*, 2011), orientation (Fischer & Whitney, 2014), face identity (Liberman *et al.*, 2014), face gender (Taubert *et al.*, 2016*a*), and attractiveness (Kondo *et al.*, 2012; Taubert *et al.*, 2016*b*; Xia *et al.*, 2016).

The distinction between central tendency bias and serial dependence seems unclear. Indeed, when the effects of serial dependence were noted in the past, albeit under the guise of assimilation effects (McKenna, 1984), they were considered in the same vein as the central tendency bias (Hellström, 1985; Ryan, 2011). Indeed, some more recent studies seek to investigate how the central tendency bias evolves during an experimental session, assessing the weighting placed on stimuli seen n trials back in the running average and effectively asking when central tendency is in fact serial dependence (Hubert-Wallander & Boynton, 2015; Mattar *et al.*, 2016). To that end, it seems that if there should be - and is - a distinction between the two types of bias then the distinction comes down to the two mechanisms having different temporal integration constants. One must consider then that the central tendency bias we see in Experiments 1 and 4 is only a bias towards the mean of the most recently seen stimuli (only a few stimuli back) rather than towards the mean of the whole stimulus set. In our experiments, the stimulus series from which reference stimuli are drawn always consists of five stimuli following a uniform distribution. As the sample mean of variables drawn from such a distribution quickly converge on the expected value (the central tendency, see Appendix D), only a few stimuli are needed before an accurate estimate of the mean of the stimulus set can be formed.

In this chapter, we have considered the manifestation of a central tendency bias in perceptual estimates of illumination hue as evidence that observers have learnt a prior over a set of reference illumination hues. We draw such a conclusion due to models in the central tendency bias literature that attribute the effect to observers learning and using priors to bias perception (e.g. Huttenlocher *et al.*, 2000). The behavioural data we present cannot easily be explained by other perceptual phenomena such as adaptation or perceptual learning. Firstly, at the start of the task observers are adapted to a D65 illumination. Considering the short exposure time to the reference and test stimuli and the constant

adaptation top-up periods (to D65 again), it is fair to assume that the observers adaptation state remains consistent across the different blocks of the experiment. Secondly, in the perceptual learning literature, through repeated testing and exposure to stimuli, observers generally get better at discriminating differences between stimuli (Fine & Jacobs, 2002). This is the opposite to a central tendency bias where all reference stimuli are remembered as more similar than they actually are (all perceptual estimates are pulled towards the mean).

The results of the 2 and 3-naming tasks and the discrimination intervals collected in Experiment 2 are interesting in their own right. Firstly, we are unaware of a previous study that has illustrated how colour category boundaries may shift with the stimulus set, a result that is suggestive of range adaptation and/or the optimal usage of colour terms for information transmission (but see Appendix B in Brown *et al.* (2011)). Of course, in the naming tasks used here we restricted participants to two or three colour terms and only used a small range of stimuli from the hue circle. It remains to be seen how consistent colour category boundaries are when the number of colour terms that can be used is increased or free naming is used and if the range of stimuli shown is varied. Secondly, the discrimination interval data is relevant to the questions we raised about the "blue bias" in Chapters 3 and 4. We will return to this point in the general discussion.

All participants in Experiment 2 and three participants in Experiment 3 had taken part in previous parts on the study. For Experiment 2, this was considered an important factor as we wished to ensure that simulated data corresponded to that collected in Experiment 1. For Experiment 3, this could mean that participants carry forward learning from the previous Experiments. However, considering that there are three blocks of reference illuminations used in the Experiment, participants must relearn the the reference illumination within each block and therefore, carrying forward a learnt prior from previous runs of the task is not helpful.

6.7 Conclusion

By using the same experimental design as Olkkonen *et al.* (2014) we showed a central tendency bias for estimates of illumination hue. Further experiments showed that the magnitude of the bias may be exaggerated by the use of colour terms as response labels, but that this did not account for all of the variability in the data. In illustrating this we also showed that colour categories are not a fixed property of colour perception and category boundaries may move with the stimulus range.

Chapter 7

Implications and Future Work

7.1 Purpose of this work

Previous work using an illumination discrimination paradigm (the IDT) found that colour constancy was better for illumination changes that became nominally bluer (relative to a neutral D67 reference illumination) along the Planckian locus (an estimate of the variation in daylight chromaticities) compared to illuminations that became nominally yellower along the Planckian locus, or that varied along an axis that represented atypical illumination changes (nominally, they became either redder or greener; Pearce *et al.*, 2014; Radonjić *et al.*, 2016*b*; Alvaro *et al.*, 2017). As measurements of daylight chromaticities vary more in the region of bluer illuminations used in the experiment compared to yellower illuminations (even though both are defined to fall on the Planckian locus), and the redder/greener illumination chromaticities are not likely in the natural world (Hernández-Andrés *et al.*, 2001; Spitschan *et al.*, 2016, see also Section 1.5.3), these findings suggest that human colour constancy mechanisms are optimised for the statistics of the environment, in the sense that they have developed or evolved to cope best with natural illuminations. Other authors have found a similar optimisation for bluer changes in illumination chromaticities that correspond to those likely to occur in nature (Delahunt & Brainard, 2004; Radonjić & Brainard, 2016), although not all the literature on this topic is in agreement (Brainard, 1998; Foster *et al.*, 2003; Rüttiger *et al.*, 2001).

We focused our efforts on understanding the results of the IDT where better colour constancy for the bluer illumination changes in the task is a reliable finding, referred to as a "blue bias" for colour constancy (Pearce *et al.*, 2014; Radonjić *et al.*, 2016*b*; Alvaro *et al.*, 2017). We hypothesised two possible mechanisms that may mediate the effect and investigated which of the two is more plausible with the studies in this thesis, although it should be noted that both mechanisms could be operating in tandem.

The first of our hypotheses (referred to as the nurture hypothesis) suggests that the effect is mediated by a learnt illumination prior that observers obtain through repeated illumination exposures throughout their lifetime - an observer specific prior. This hypothesis makes multiple predictions that were tested in this thesis. Firstly, it predicts that a similar bias (better colour constancy for bluer illumination changes) will not be seen in different versions of the IDT where the reference illumination (and all comparisons) are specified to fall in a different location in colour space such that the natural variation in measured daylight chromaticities does not coincide with the bluer illuminations used in the experiment. This prediction is tested in Chapter 3. Secondly, the nurture hypothesis predicts that, as observers have their own individual prior shaped by their personal experience with illuminations, there will be inter-individual differences present in the level of "blue bias" displayed by observers. We test this prediction primarily in Chapter 4 but also in Chapter 3. As we find evidence in support of the nurture hypothesis in both Chapters 3 and 4, we ask in Chapter 5 if individual priors can explain a recent visual illusion and, finally, in Chapter 6, we look for evidence that observers can learn a prior over illumination characteristics that influences behaviour.

The second hypothesis (the nature hypothesis) suggests that the effect is mediated by a reduced sensitivity to global changes in scene chromaticity that are in a bluer direction in a chromaticity plane. A reduced sensitivity to such changes would be expected if the human visual system has become optimised for the statistics of natural illuminations through evolution and visual mechanisms less sensitive to such changes have been selected for, effectively producing a species specific illumination prior. This hypothesis also makes multiple predictions that were tested in this thesis. One prediction is that the "blue bias" we see in the IDT results will also be present in other, more classical, colour discrimination or detection tasks. We investigate this in Chapter 4. In addition, the nature hypothesis makes contrasting predictions to the nurture hypothesis. It predicts that the level of "blue bias" will be less variable across observers and that a "blue bias" will be seen for other versions of the IDT where the reference illumination (and all comparisons) are specified

to fall in a different location in colour space. The results of both Chapters 3 and 4 are relevant to this controversy.

7.2 Summary of main findings

7.2.0.3 Evidence for the nature vs. nurture hypotheses

In Chapter 3, we find that thresholds for illumination discrimination in the IDT show different patterns of asymmetry when the reference illumination (and all comparison illuminations) are shifted in chromaticity. Specifically, if we consider the shift relative the neutral reference illumination condition that was used in previous studies (Pearce *et al.*, 2014; Radonjić *et al.*, 2016*b*; Alvaro *et al.*, 2017), then thresholds are enlarged for the direction of illumination change that is chromatically opponent to the chromatic change in the illuminations; for example, when the experimental illuminations all become greener, the largest thresholds are in the redder direction of chromatic change. In other words, thresholds are enlarged along the chromatic axis of change along which the illuminations approach the distribution of daylight chromaticities. This suggests that colour constancy mechanisms are optimised for illuminations with similar chromaticities to those of daylights. This is an effect one may expect to see if observers utilise a daylight prior (either an observer specific or species prior), and we provide an explanation of why a model of this kind may predict these data in the Discussion to Chapter 3. However, we also find that thresholds for the bluer direction of change (regardless of reference illumination condition) are still the highest on average. This supports the nature hypothesis, suggesting a reduced sensitivity to chromatic changes in the illumination that are "blueish" that, importantly, is not specific to the changes typical of changes in daylight illuminations. In addition, by comparing our results to those of Radonjić *et al.* (2016*b*), we find that the explanation of their results in terms of the biased set of surfaces in the scene may not hold as it cannot be extended to our results.

In Chapter 4 we explored individual differences in illumination discrimination ability,

specifically by looking at inter-individual differences in the level of "blue bias" displayed by observers. We find that the "blue bias" is variable across observers, the magnitude of the variability exceeding that of intra-individual differences. In fact, some observers show no "blue bias" at all. These results are suggestive of individual priors (nurture hypothesis) rather than species priors (nature hypothesis). In addition, we show in Chapter 4 that illumination discrimination ability (in the IDT) is not easily predicted by chromatic contrast discrimination ability (in the CCDT); although we point out several issues in the Discussion to Chapter 4 with the comparison between the two tasks. In addition, we show that a bias in the retinal information, caused by asymmetries in cone ratios or the biological constraints of human optics, cannot account for the "blue bias". However, differences in cone ratios may begin to explain some of the individual differences in overall illumination discrimination ability; but this idea needs further exploration and development.

Chapter 5 considered the idea of daylight illumination priors in the context of #theDress (the internet phenomenon of 2015). It has previously been hypothesised that observers infer incident illuminations of different chromaticities, relying on illumination priors to overcome the ambiguity of the image, and that a cue to differences in illumination priors may be an observers' chronotype. We found a relationship between illumination matches and matches to the dress (as others before us, Witzel *et al.*, 2016), and a weak relationship between chronotype and illumination matches (morning types giving bluer illumination matches than evening types). In other words, the dress photograph may be considered as evidence of the influence of illumination priors on colour perception.

Finally, in Chapter 6 we conducted a study to assess if observers can learn a prior over illumination parameters during the time frame of a psychophysics experiment. In a first experiment, we found strong evidence of a learnt illumination prior. However, we postulated another explanation for the results that was related to the task design. When we altered the experiment to account for these effects, we still found evidence of a learnt illumination prior, but the effect was much smaller.

7.2.0.4 Contribution to the literature on second and higher-order colour mechanisms

After making certain assumptions about the second-order mechanisms in Chapter 4 to make what we called "probability predictions" of illumination discrimination thresholds using chromatic contrast discrimination data, we went on to perform a second task (the iCCDT) to assess how valid these predictions were. The assumptions that we made were firstly that the cardinal axes of DKL colour space are an accurate and exhaustive representation of the second-order colour mechanisms and, secondly, that they act independently to detect chromatic changes in a stimulus. The validation experiment that we performed suggests that these assumptions cannot be made; we found evidence of either higher-order colour mechanisms or that the cardinal mechanisms do not act independently. It Section 1.4 we briefly reviewed the literature on this topic and concluded that there is currently no consensus in the field on the number or independence of chromatic mechanisms. Our data have implications for this debate, supporting the argument that there is an interaction between the cardinal mechanisms or that higher-order colour mechanisms also exist.

7.2.0.5 Contribution to the literature on central tendency biases in perception

The results we present in Chapter 6 are also relevant to the literature on central tendency biases in perception (and possibly also the literature on serial dependencies in perception, see the Discussion in Chapter 6 where we discuss if these two phenomenon should really be considered separately). We began in that Chapter by replicating an experiment originally performed by Olkkonen *et al.* (2014) that showed a central tendency bias for perpetual estimates of simulated surface hues, although we used real illuminations rather than simulated surfaces as our stimuli. Like Olkkonen *et al.* (2014), we found a central tendency bias for perceptual estimates of illumination hue. However, we postulated that the effect may be exaggerated by the design of the task and found evidence of such in a second experiment. In further experiments, we find mixed evidence for a residual central

tendency bias after the effect of the task design is counteracted. These results should be considered in relation to other studies on central tendency biases in perception where similar task design elements may be exaggerating the central tendency bias effect.

7.2.0.6 Contribution to the literature on the Bayesian brain hypothesis

We cannot draw conclusions from this work as to whether the brain does indeed behave like a Bayesian hypothesis tester or predictive machine, rather, that is the framework that we adopted when conducting these experiments. A different stream of research is necessary to assess whether the brain is actually performing these types of computations or heuristics of them such as studies that ask whether observers show behavioural qualities that Bayesian computation would predict (e.g. Ernst & Banks, 2002), or studies that perform model comparison (e.g. asking whether a Bayesian or switching observer model best fits behavioural data; Laquitaine & Gardner, 2017).

7.3 Limitations of this work

To measure thresholds for colour constancy in this thesis, we adopted the illumination discrimination task. In the illumination discrimination task, observers must find the illumination, among two comparisons, that best matches a reference. The instructions for the task are: *"On each trial, you will see the reference illumination followed by two comparison illuminations. You will use the gaming pad to indicate which of the two comparison illuminations most closely matched the reference"*. This means that optimal performance on the task requires colour constancy mechanisms to fail; i.e. in order to always detect the changes in the proximal stimulus caused by a change in the illumination, the participant should *not* remain colour constant. Indeed, we find evidence of learning along these lines over repeated runs of the neutral (original) IDT in Experiment 2 of Chapter 4.

When an experimenter says they are measuring thresholds, it is generally considered that they are measuring an upper limit in performance. While we can say we are measuring

an upper limit of illumination discrimination ability with the IDT (the smallest change in illumination detectable to the participant), we cannot be sure that we are also measuring an upper limit of colour constancy (the largest change in the illumination under which the participant remains colour constant) as this is not the participant's goal in the task.

This highlights a second criticism of the IDT: it is yet to be established how thresholds on the IDT relate to colour constancy in the standard sense. We will return to this point when we discuss ideas for future work (Section 7.4 below).

A common question asked of the IDT results is how much the threshold asymmetries depend on the colour space that thresholds are calculated in. In a previous paper (Radonjić *et al.*, 2016*b*), we showed that using a different colour space (S-CIELAB) the asymmetries are still present.

7.4 Future work

There are possible avenues for future work that follow on from the studies in this thesis. In the Discussion to Chapter 3 we suggested that a Bayesian model that incorporates a daylight prior, or even a daylight-manmade prior, for illuminations may explain the biases that we see in the tasks. However, for such a model to be developed, one would need: (i) a database of man-made illumination spectra to combine with measurements of daylights to form the prior; (ii) to choose a method by which observers extract illumination estimates from the scene (e.g. the-grey-world or brightest-is-white hypothesis); (iii) a generative model that captures how different illuminations interact with surfaces in the scene (making assumptions about the likely surface reflectance properties of surfaces in the scene) in order to form the likelihood function on each trial. On top of that, one would need to decide whether to deal with spectral quantities or just tristimulus values (actually, a mixture of both will likely be necessary). The first of these requirements is the most easy to meet, the second requires a subjective choice by the modeller, but the third is particularly hard to achieve and is one of the reasons why such a model was not developed as part of this

work. However, a future long term project may be able to achieve this goal as we gain more of an understanding of the statistics of natural and man-made scenes.

There is also scope to extend the ideal observer models developed using ISETBIO in Chapter 4. In ISETBIO, it is possible to model the transmission of signals through the bipolar and ganglion cell layers of the retina. In addition, the ISTEBIO team intend to extend this further, taking into account processing in the LGN and visual cortex. We showed in Chapter 4 that an alteration to the modelling procedure that amounts to increasing the level of noise in the model leads to predicted thresholds of a more similar magnitude to those observed in the task. Adding these extra stages of visual processing will also add further noise to the model and could even lead to predictions of the asymmetries in thresholds. Furthermore, the output of the ISETBIO model could be used as the sensory input in the formulation of a Bayesian model of the type discussed above.

In Chapter 4, we only considered the effects that individual differences in chromatic contrast discrimination ability or $L : M$ cone ratios may have on IDT thresholds. The inter-individual differences that we observe in overall ability on the IDT, however, may be explained by factors unrelated to the mechanisms of colour vision, such as working memory or other cognitive factors. It would be useful in the future to perform a cognitive test battery on participants who complete the IDT to investigate whether inter-individual differences in cognitive abilities can account for inter-individual differences in performance on the IDT.

In addition, as mentioned in the previous section, it is unclear how thresholds on the IDT relate to classical measures of colour constancy; especially since optimal performance on the IDT encourages a lack of colour constancy. One way to assess this would be to design a task in which observers are required to find the corresponding object across changes in illumination. The IDT makes a prediction of the probability that an observer finds the correct object under each illumination change used in the IDT. For example, if we fit a psychometric function to the staircase data for the redder direction of change in the neutral (original) IDT task for one observer, then the reciprocal of probabilities read out from the

fitted function can be used to predict the probability of correct object identification under each illumination change along this axis of change (or at least it can for objects whose surface spectral reflectance functions are similar to those of the Mondrian patches used in the experiment).

Finally, it remains to be seen whether the "blue bias" will translate to experiments using more naturalistic scenes. Alvaro *et al.* (2017) provide a promising contribution along these lines that suggests the effect is not specific to Mondrian papered scenes like the one used in the IDT, rendering illuminations on hyperspectral images of natural scenes.

7.5 Overall conclusions

To conclude, in the experiments of this thesis, we found further evidence that the colour constancy mechanisms of the human visual system are optimised for the statistics of the natural world; i.e. they are optimised for natural daylight illuminations. In addition, we found evidence that this optimisation comes about through the use of a learnt prior over the characteristics of daylight illuminations (the nurture hypothesis - individual priors). However, we also find evidence of a generic reduction in sensitivity to bluer changes in an illumination (the nature hypothesis - a species prior), but this same reduction in sensitivity is not present in measures of chromatic contrast discrimination ability. In this case, at present, we must conclude that there are colour constancy mechanisms that relate to both of our hypotheses. We do not yet know how a daylight illumination prior or its influence on colour perception would be represented in the visual system but suggest that a Bayesian model may predict the results in future.

Appendix A

Chapter 3 appendix

A.1 Fitting the 3-component CIE daylight model to our illuminations

To assess how much the spectral content of the illuminations used in the experiment match that of daylights, we asked how well our illuminations could be model using the 3-component CIE daylight model (CIE, 2004). The CIE daylight model consists of three basis functions: $S0$, $S1$ and $S2$ (Figure A1.A). To fit the CIE daylight model to our illuminations we first normalised the CIE daylight basis functions by their vector norm (L^2 norm). We then took each experimental illumination spectrum (Figure A1.B), normalised it by its vector norm, and used multiple least squares regression to find the weightings of the three CIE daylight basis functions that provide the best fit to the experimental illumination. Taking a combination of the basis functions with these weightings applied gives the recovered spectrum (Figure A1.C). The quality of the fit was assessed using the R^2 statistic (the coefficient of determination, the square of Pearsons correlation coefficient between the true and recovered spectra). Averaging over all illuminations, the mean goodness of fit was $R^2 = 64.95\% \pm 12.87\%$ (mean $\pm$ SD), indicating that the CIE daylight basis functions explain a large proportion of the variation in the experimental illuminations. In Table A1, we show the mean proportion of variance explained (mean R^2) for each chromatic axis of change used in the experiment. We see that the CIE daylight model provides a better approximation of the illuminations in the conditions where the reference illumination is parametrised to fall on the Planckian locus (neutral, blue and yellow conditions).

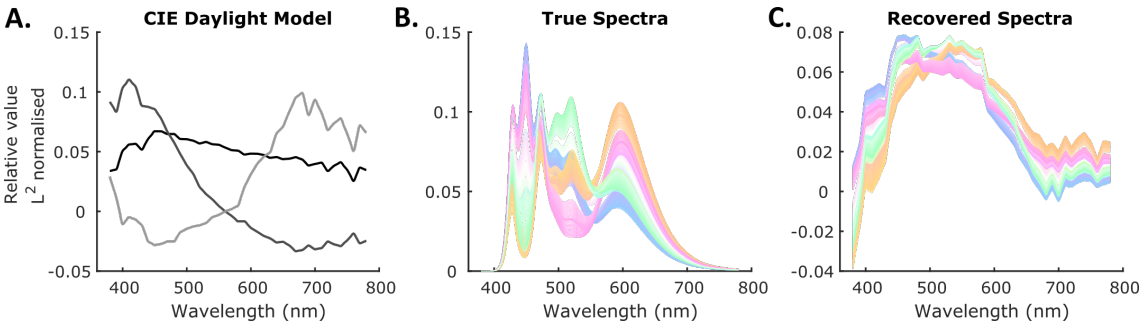


Figure A.1: Fitting the 3-component CIE daylight model to the IDT illuminations. A. The L^2 normalised CIE daylight basis functions. $S0$ is in black, $S1$ is in medium grey and $S2$ in light grey. B. The measured spectral power distributions of all illuminations used in the experiment. C. The recovered spectral power distributions of all illumination spectra obtained from fitting the 3-component CIE daylight model.

Table A.1: Table of goodness of fit values (R^2) for the CIE daylight model to each axis of chromatic change used in the experiment.

Reference	Bluer	Greener	Redder	Yellower
Neutral	76.07%	74.35%	77.80%	72.27%
Blue	68.16%	71.62%	72.47%	66.62%
Green	60.21%	55.07%	61.21%	54.76%
Red	44.78%	42.95%	49.03%	39.67%
Yellow	79.72%	76.37%	76.13%	79.68%

Appendix B

Chapter 4 appendix

B.1 Quantifying the "chromatic bias" in IDT Thresholds

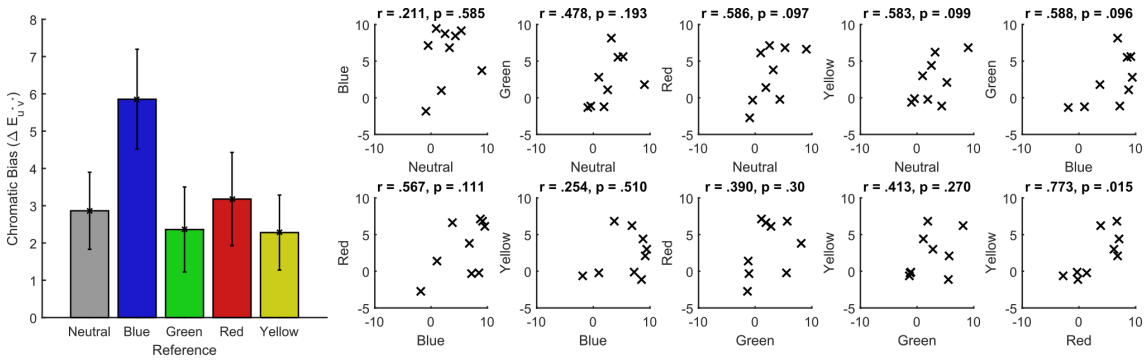


Figure B.1: Quantifying the "chromatic bias" in different reference illumination conditions. A. The mean "chromatic bias" in the five reference illumination conditions defined as the average of the difference between the chromatic direction of change in the opposite chromatic direction to the bias in the illuminations and all other directions. B-K. Scatter plots comparing the "chromatic bias" in each pair of reference illumination conditions with Pearson's correlation coefficient shown above.

B.2 Extra sources of individual differences

B.3 Extra adaptation effects plots

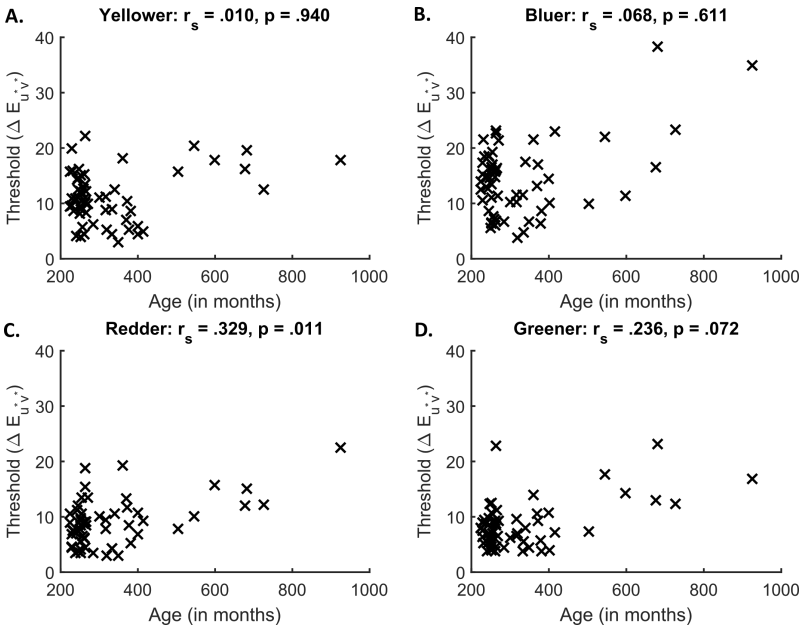


Figure B.2: Correlating the thresholds for the different directions of chromatic change in the IDT with age. A-D. Correlations of thresholds for each direction of change with age. Spearman's rank correlation coefficient between the two variables is shown above each plot.

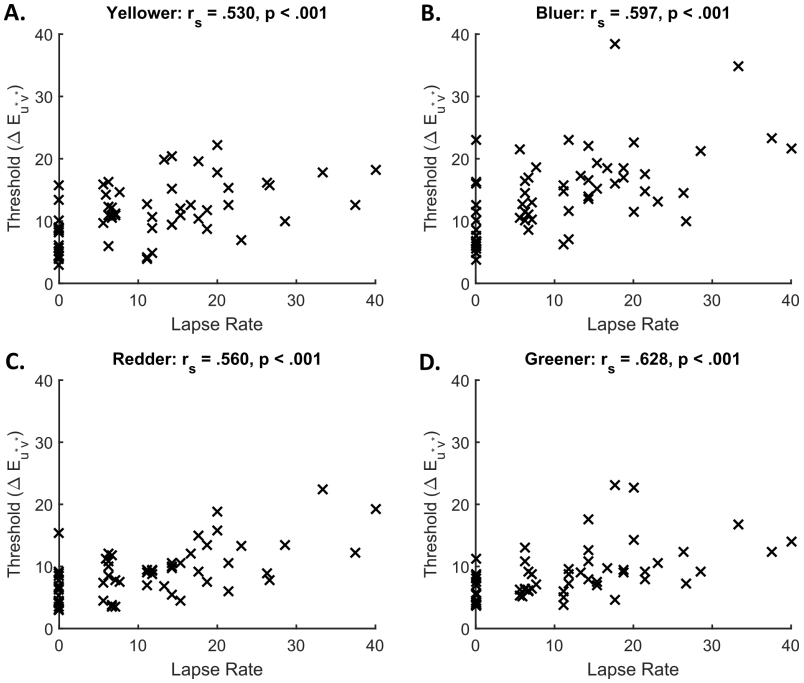


Figure B.3: Correlating the thresholds for the different directions of chromatic change in the IDT with lapse rate. A-D. Correlations of thresholds for each direction of change with lapse rate. Spearman's rank correlation coefficient between the two variables is shown above each plot.

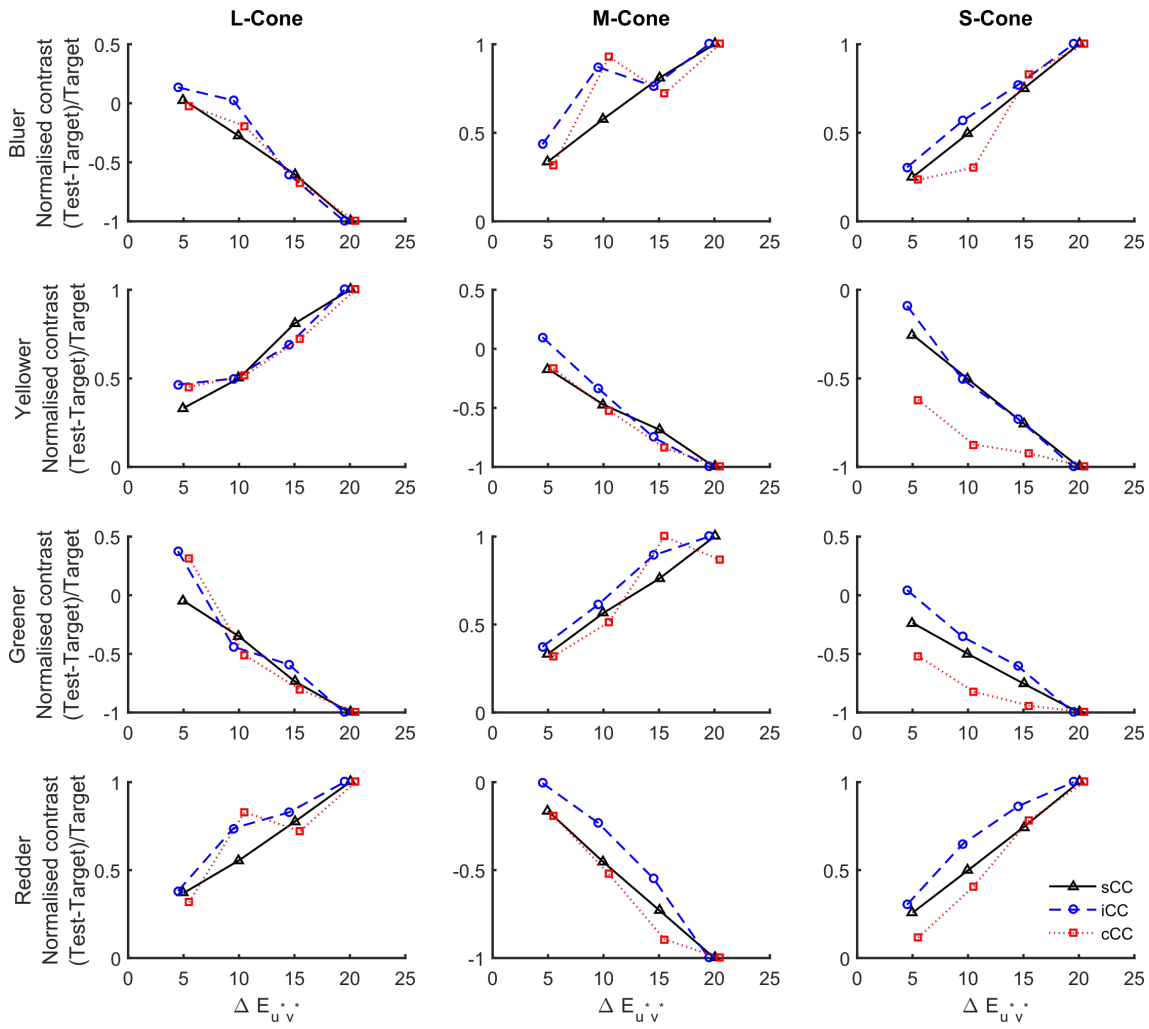


Figure B.4: Simulated current, isomerisation and standard cone contrast values for the neutral reference illumination condition when the target is in the second comparison position. Each column of plots shows the contrast values for a different cone type and each row of plots for a different direction of chromatic change in the illumination. Along the horizontal axis is plotted the distance in CIELUV between the test and target illuminations. The vertical axes show the normalised contrast values.

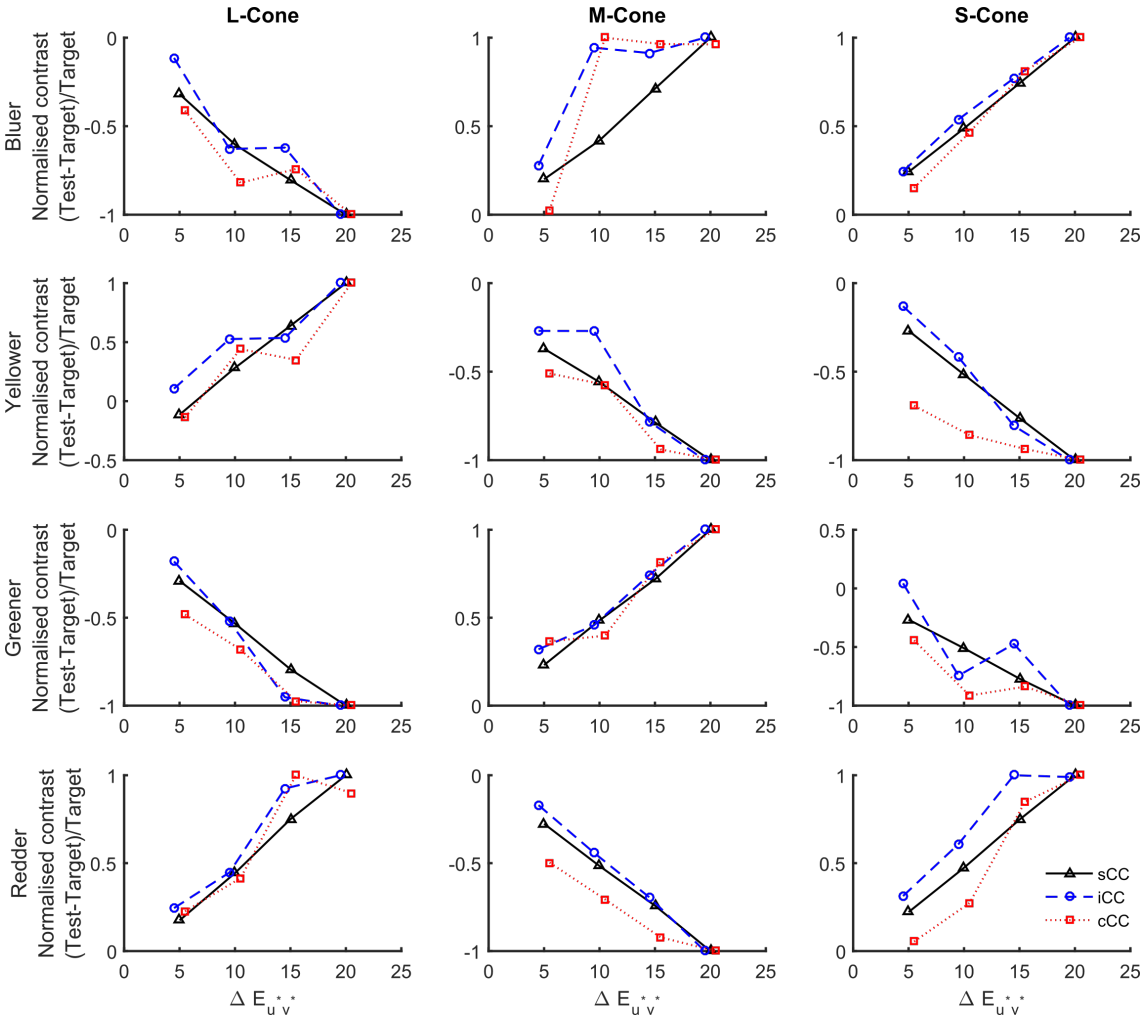


Figure B.5: Simulated current, isomerisation and standard cone contrast values for the blue reference illumination condition when the target is in the first comparison position. Each column of plots shows the contrast values for a different cone type and each row of plots for a different direction of chromatic change in the illumination. Along the horizontal axis is plotted the distance in CIELUV between the test and target illuminations. The vertical axes show the normalised contrast values.

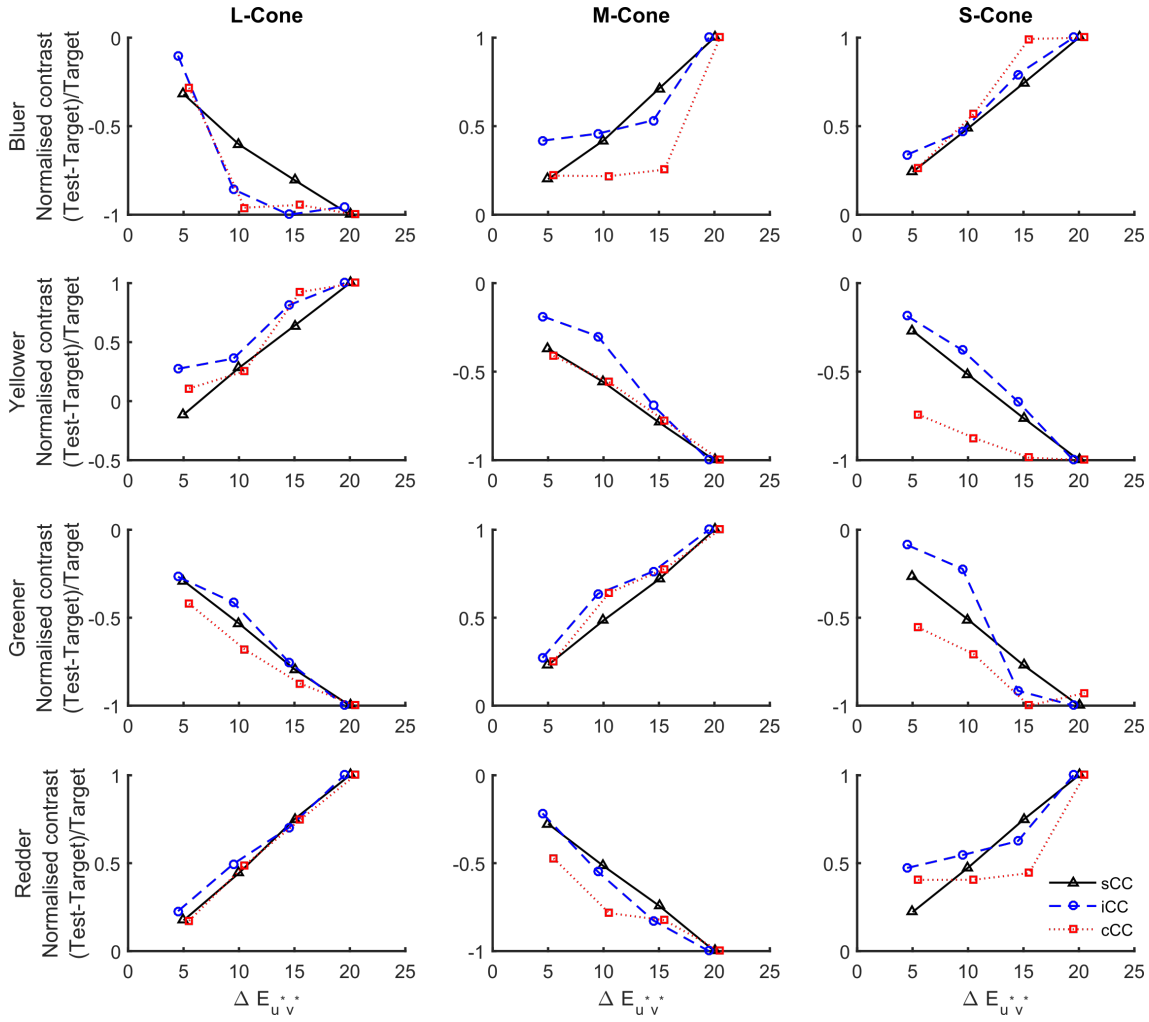


Figure B.6: Simulated current, isomerisation and standard cone contrast values for the blue reference illumination condition when the target is in the second comparison position. Each column of plots shows the contrast values for a different cone type and each row of plots for a different direction of chromatic change in the illumination. Along the horizontal axis is plotted the distance in CIELUV between the test and target illuminations. The vertical axes show the normalised contrast values.

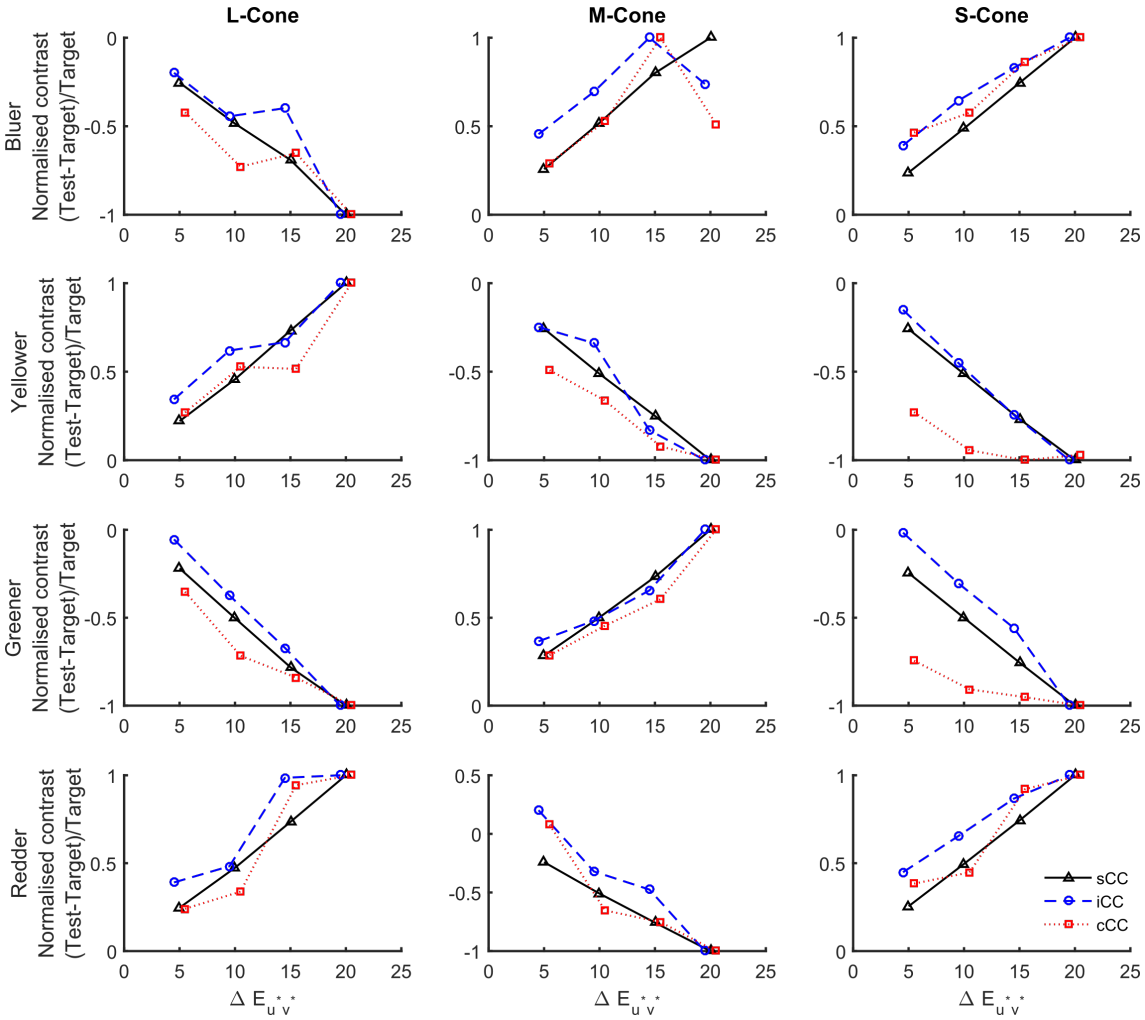


Figure B.7: Simulated current, isomerisation and standard cone contrast values for the green reference illumination condition when the target is in the first comparison position. Each column of plots shows the contrast values for a different cone type and each row of plots for a different direction of chromatic change in the illumination. Along the horizontal axis is plotted the distance in CIELUV between the test and target illuminations. The vertical axes show the normalised contrast values.

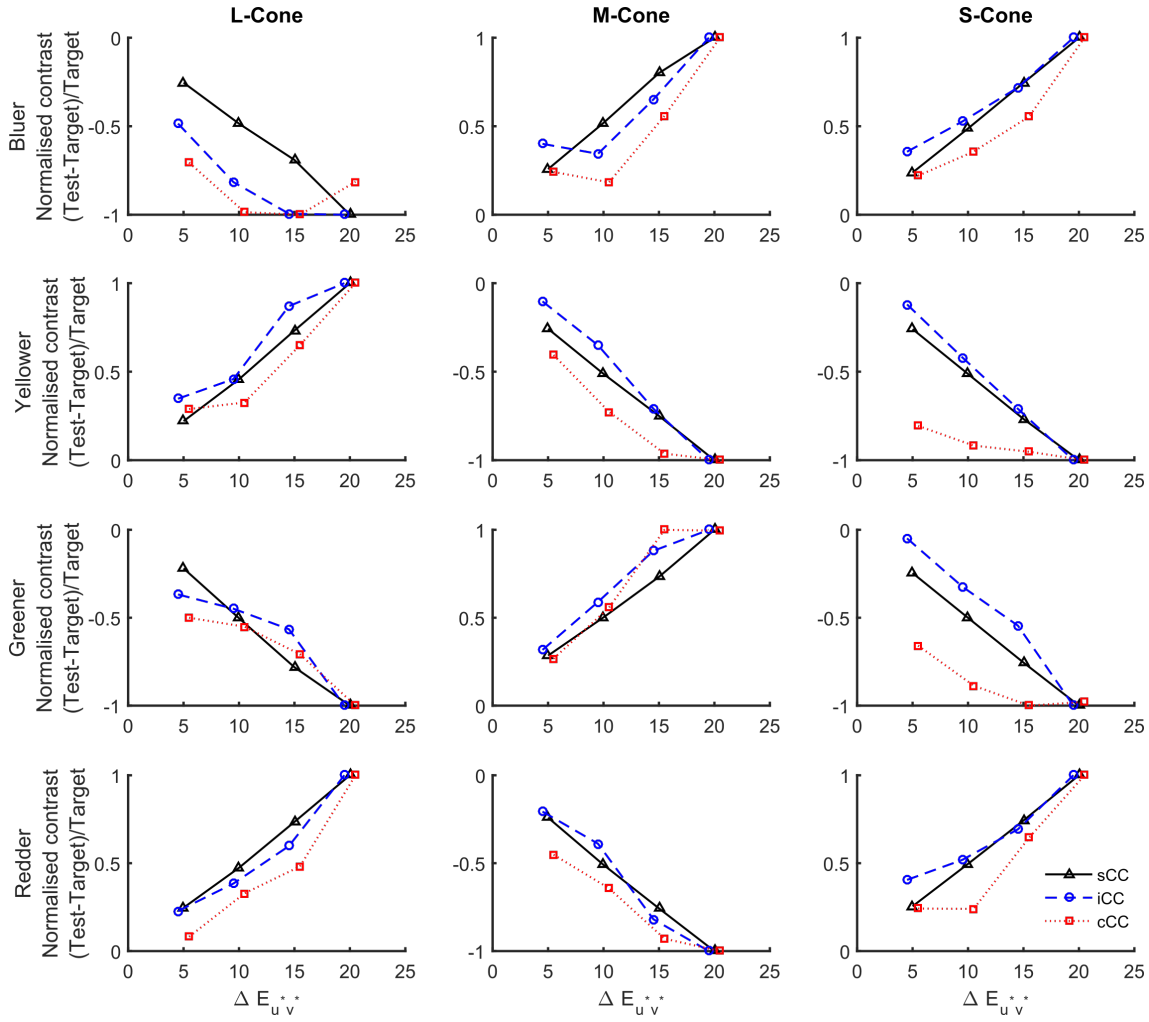


Figure B.8: Simulated current, isomerisation and standard cone contrast values for the green reference illumination condition when the target is in the second comparison position. Each column of plots shows the contrast values for a different cone type and each row of plots for a different direction of chromatic change in the illumination. Along the horizontal axis is plotted the distance in CIELUV between the test and target illuminations. The vertical axes show the normalised contrast values.

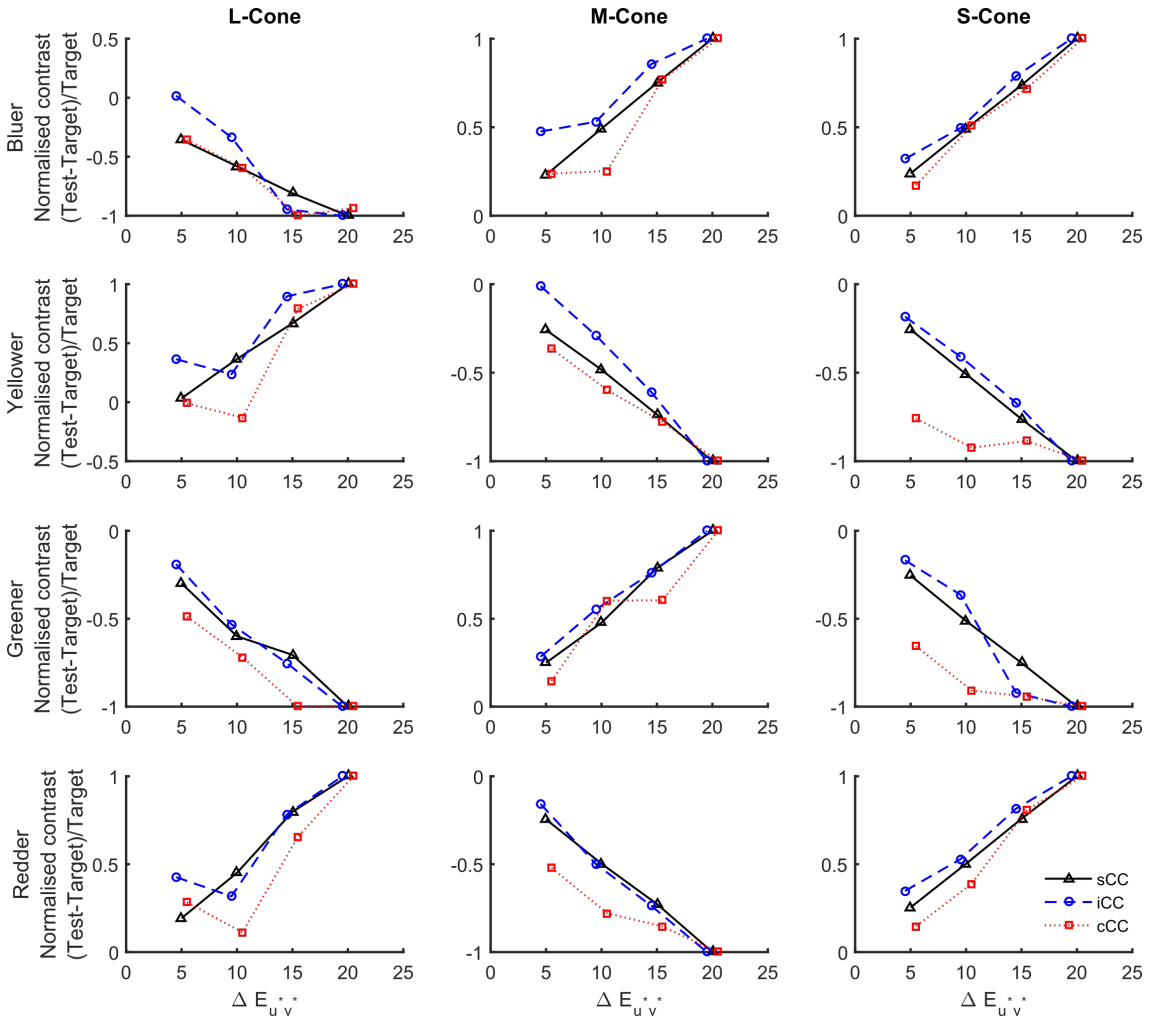


Figure B.9: Simulated current, isomerisation and standard cone contrast values for the red reference illumination condition when the target is in the first comparison position. Each column of plots shows the contrast values for a different cone type and each row of plots for a different direction of chromatic change in the illumination. Along the horizontal axis is plotted the distance in CIELUV between the test and target illuminations. The vertical axes show the normalised contrast values.

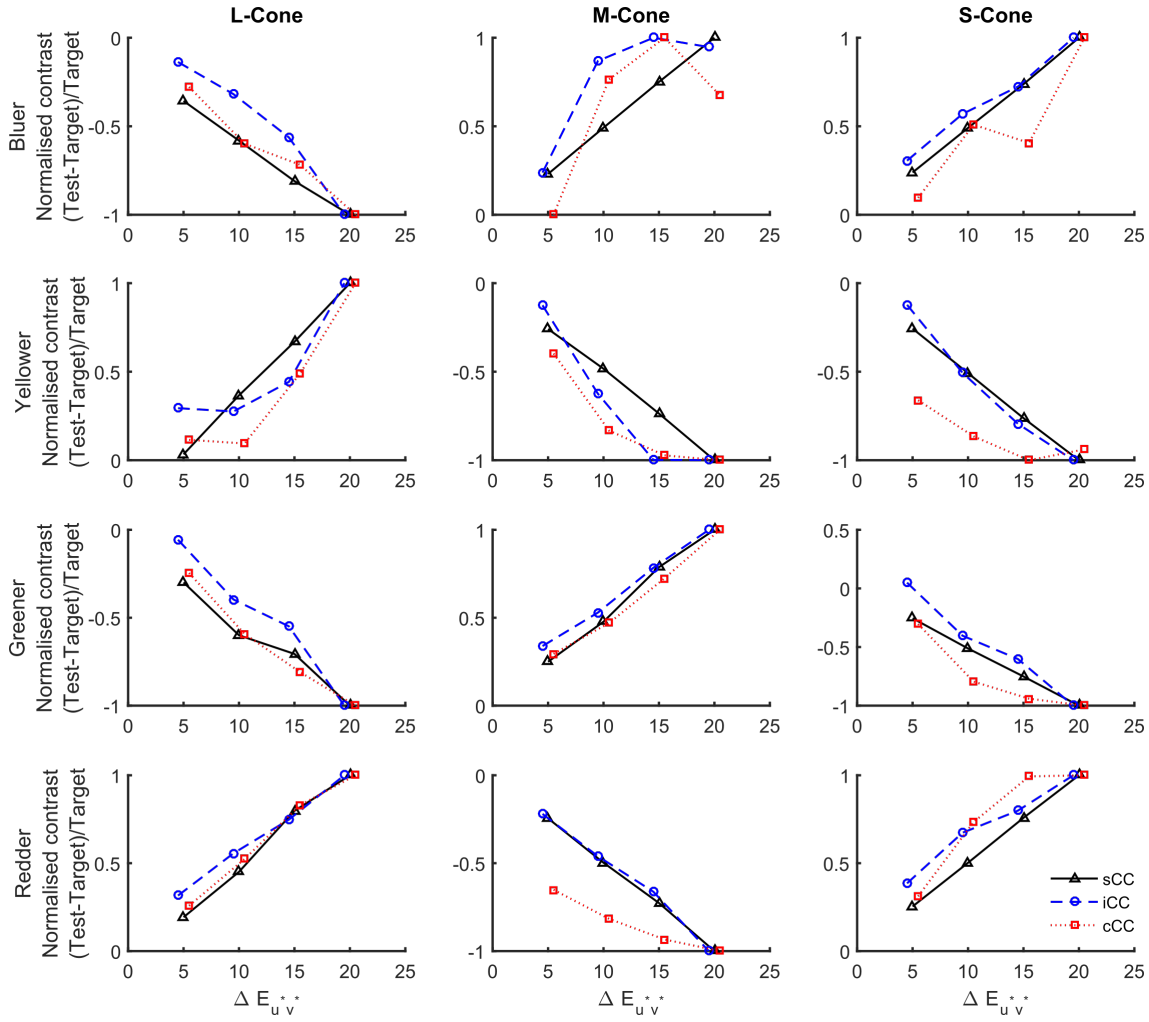


Figure B.10: Simulated current, isomerisation and standard cone contrast values for the red reference illumination condition when the target is in the second comparison position. Each column of plots shows the contrast values for a different cone type and each row of plots for a different direction of chromatic change in the illumination. Along the horizontal axis is plotted the distance in CIELUV between the test and target illuminations. The vertical axes show the normalised contrast values.

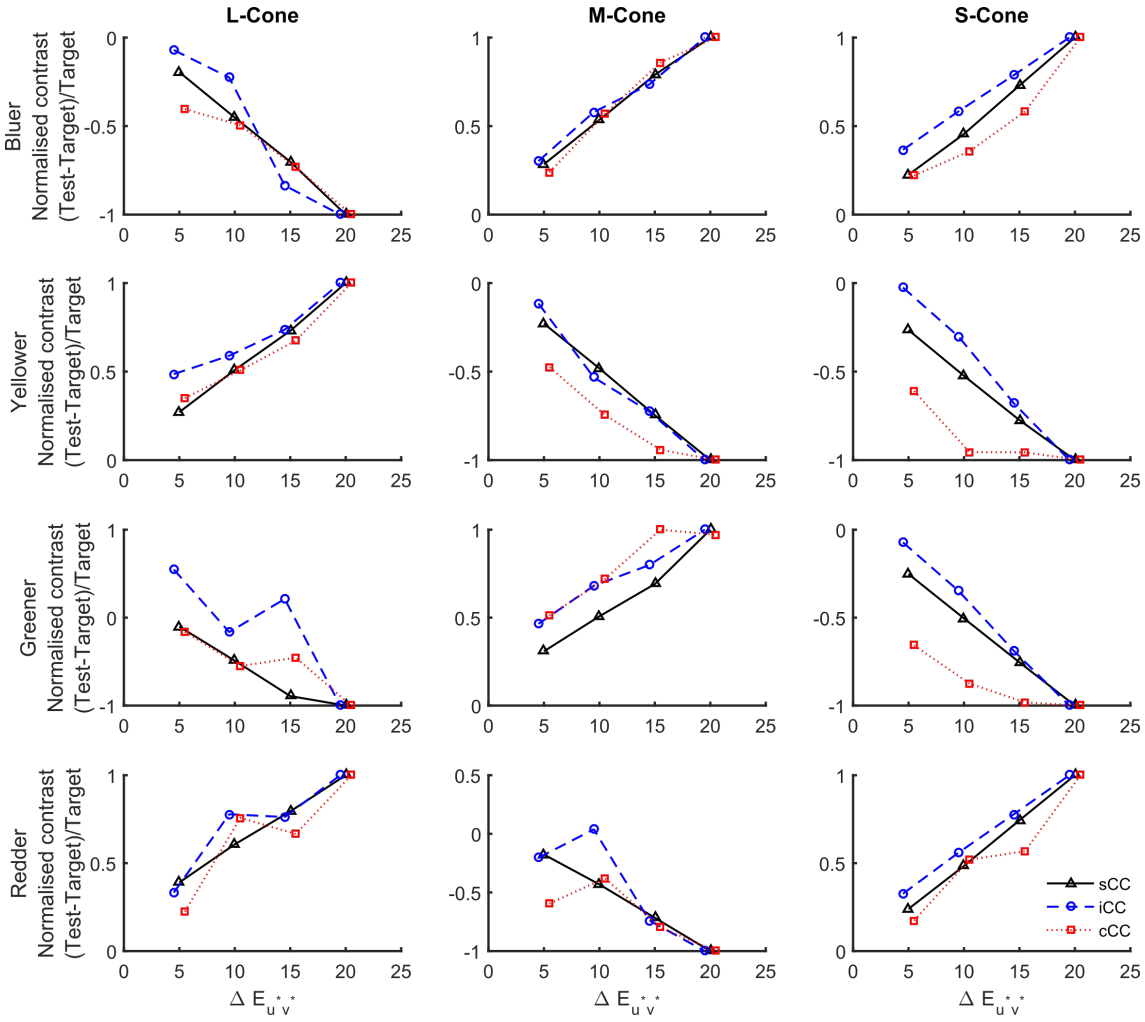


Figure B.11: Simulated current, isomerisation and standard cone contrast values for the yellow reference illumination condition when the target is in the first comparison position. Each column of plots shows the contrast values for a different cone type and each row of plots for a different direction of chromatic change in the illumination. Along the horizontal axis is plotted the distance in CIELUV between the test and target illuminations. The vertical axes show the normalised contrast values.

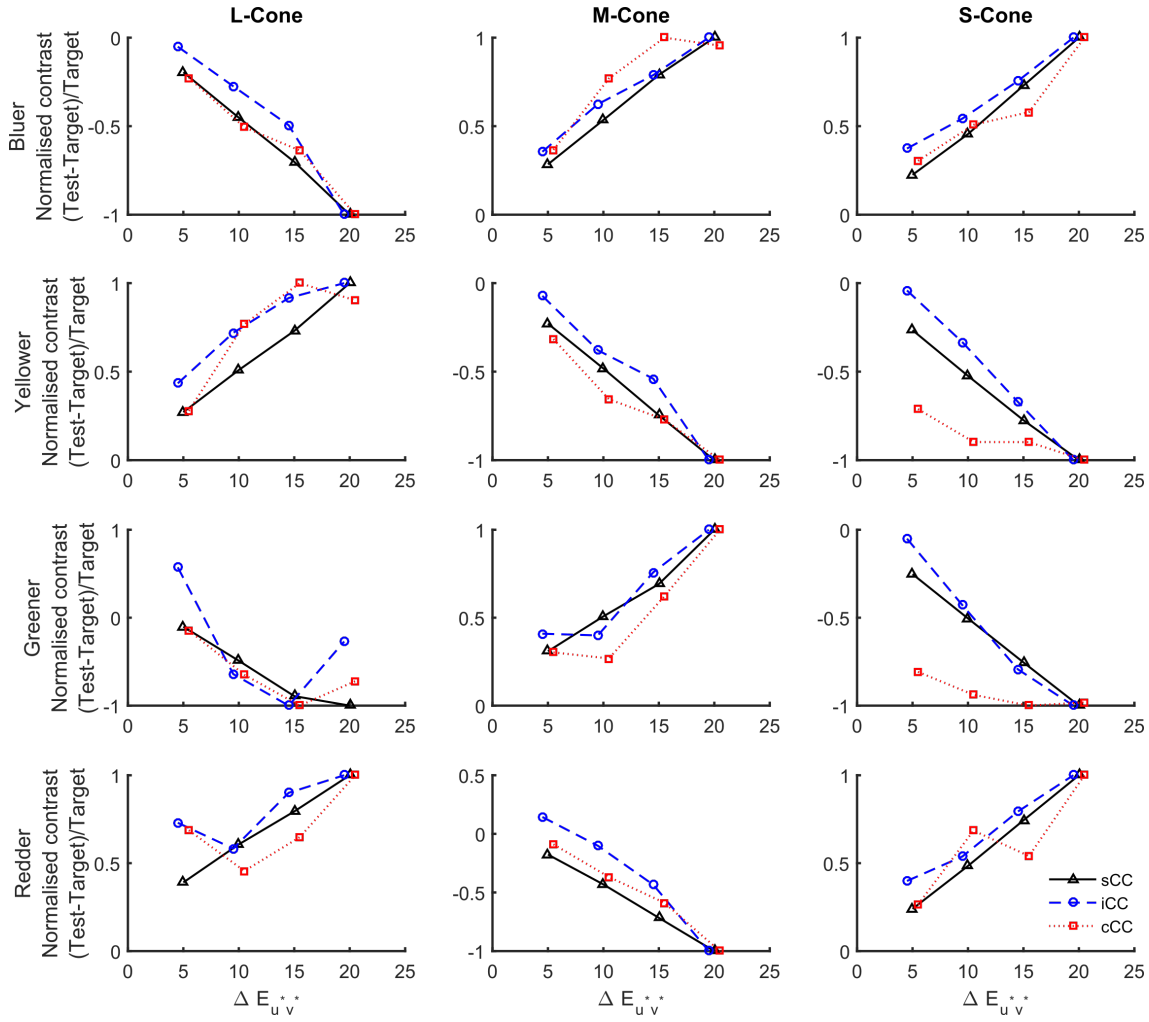


Figure B.12: Simulated current, isomerisation and standard cone contrast values for the yellow reference illumination condition when the target is in the second comparison position. Each column of plots shows the contrast values for a different cone type and each row of plots for a different direction of chromatic change in the illumination. Along the horizontal axis is plotted the distance in CIELUV between the test and target illuminations. The vertical axes show the normalised contrast values.

Appendix C

Chapter 5 appendix

C.1 Categorisation of colour names

Table C.1 shows the colour names that participants reported for the dress photograph the first time they saw it and the categorisation of these colour names into original dress colour name groups. The colour names reported by the participants were categorised independently by four members of the Hurlbert Colour Vision Laboratories at Newcastle University. The categorisations of all four lab members agreed.

Table C.2 shows the colour names that participants reported for their matches to the dress body and lace when presented in isolation. As with the original dress colour names, four members of the laboratory independently categorised these data into groups. In 29 out of the 32 cases, all four experimenters agreed. In two cases (participants 2050 and 2053), there was a 50/50 split in the categorisation (2050: blue/gold vs. white/gold; 2053: blue/black vs. blue/white). Here, the categorisation of the most experienced researcher was favoured. Similarly, for one participant (2009) all categorisations disagreed (white/gold vs. blue/green vs. purple/green vs. blue/gold). Again, the categorisation of the most experienced researcher was favoured.

C.2 ANOVA analyses of dress body and lace colour matches

With dress body matches grouped according to original dress colour names (B/K, W/G, B/G), there is no difference across groups on the lightness (L^*) or blue-yellow (v^*) axes ($F(2, 29) = 2.66$, $p = 0.087$ and $F(2, 29) = 1.73$, $p = 0.196$). There was a difference along the red-green dimension ($u^* : F(2, 29) = 5.42$, $p = 0.01$), with the B/K group matching the dress body to significantly lower u^* values (more green) than the W/G group (mean difference of 6.97, $p = 0.01$, Bonferroni corrected). However, the same grouping for dress lace matches results in significant differences along all axes of CIELUV colour space (L^*,

Table C.1: Categorisation of the colour names that participants reported for the dress photograph the first time they saw it. The original dress names column shows the colour names as reported by the participant. The categorised column shows the categorisation of these colour names.

Participant ID	Original Dress Names (Body/Lace)	Categorised (Body/Lace)
2000	white/gold	white/gold
2001	blue/black	blue/black
2002	blue/gold	blue/gold
2003	blue/black	blue/black
2005	blue/black	blue/black
2009	blue/black	blue/black
2021	white/gold	white/gold
2022	blue/(black/brown)	blue/black
2023	white/gold	white/gold
2024	white/gold	white/gold
2031	white/gold	white/gold
2033	blue/black	blue/black
2034	blue/black	blue/black
2039	blue/black	blue/black
2040	white/gold	white/gold
2041	blue/gold	blue/gold
2044	blue/gold	blue/gold
2045	blue/black	blue/black
2046	blue/black	blue/black
2047	white/gold (yellow)	white/gold
2049	blue/gold	blue/gold
2050	white/metallic gold	white/gold
2051	grey blue/gold	blue/gold
2052	light blue/chocolate brown	blue/gold
2053	blue/black	blue/black
2057	blue/black	blue/black
2059	blue/gold	blue/gold
2061	blue/gold	blue/gold
2062	white/gold	white/gold
2063	blue/black	blue/black
2064	white/gold	white/gold
2065	white/gold	white/gold

u^* and v^*: $F(2, 29) = 8.03$, $p = 0.002$; $F(2, 29) = 13.21$, $p < 0.001$ and $F(2, 29) = 6.42$, $p = 0.005$).

With dress body matches grouped according to disk colour names (B/K, W/G, B/G, P/G) there are significant differences along L^* ($F(3, 26) = 6.32$, $p = 0.002$) and v^* ($F(3, 26) = 7.56$, $p = 0.001$). Matches did not vary along the u^* dimension ($F(3, 26) = 2.47$, $p =$

Table C.2: Categorisation of the colour names that participants reported for their matches to the dress body and lace when presented in isolation. The disk names column shows the colour names as reported by the participant. The categorised column shows the categorisation of these colour names.

Participant ID	Original Dress Names (Body/Lace)	Categorised (Body/Lace)
2000	lavender/bronze	purple/gold
2001	purpley blue/brown	blue/gold
2002	blue/gold	blue/gold
2003	sky blue/dark grey brown	blue/gold
2005	blue/black	blue/black
2009	light grey with tiny hint of blue and a breath of lilac/ greeny mustard yellow ochre	white/gold
2021	lilac/fawn	purple/gold
2022	blue/brown	blue/gold
2023	very pale grey blue/dirty brown gold	blue/gold
2024	white/gold	white/gold
2031	blue/brown	blue/gold
2033	blue/black	blue/black
2034	pastel blue/reddy brown	blue/gold
2039	blue/green	blue/green
2040	pale blue/muddy yellow	blue/gold
2041	light blue/orangy brown	blue/gold
2044	blueish grey/sandy burnt	yellow/gold blue/gold
2045	duck egg/mustard	blue/gold
2046	lilac/khaki	purple/gold
2047	white/dark yellow	white/gold
2049	pale violet/fleshy tan	purple/gold
2050	grey bluey white/yellow	blue/gold
2051	grey blue/sandy orange	blue/gold
2052	light blue/dark orange	blue/gold
2053	blue/grey	blue/black
2057	blue/purple	blue/purple
2059	pale blue/mustard yellow	blue/gold
2061	light blue/gold	blue/gold
2062	blue/black	blue/black
2063	light grey/mustard	white/gold
2064	white/thorn	white/gold
2065	purple/mustard	purple/gold

0.084). Matches to the dress lace also differ significantly between disk colour names groups on the v^* ($F(3, 26) = 6.16$, $p = 0.003$) and u^* axes ($F(3, 26) = 3.73$, $p = 0.024$). Dress lace matches did not differ across disk colour name groups on the L^* axis ($F(3, 26) = 2.41$,

$p = 0.09$).

C.3 Control experiment 1: achromatic matches at different luminance levels

C.3.1 Methods

Seven participants from the main experiment returned to the laboratory at a later date to complete a control experiment. The main purpose of the control experiment was to ascertain whether the fixed luminance setting of the matching disk affects the chromaticity of the achromatic settings. In particular, we asked if an increased luminance level leads to a bluer achromatic setting by requiring that all participants adjusted the matching disk to look achromatic at each of five different fixed luminance levels (7.35 cd/m^2 , 18.20 cd/m^2, 34.46 cd/m^2, 54.78 cd/m^2 and 96.49 cd/m^2; equivalent to the minimum, lower quartile, median, upper quartile and maximum luminance settings of the dress body matches from the main experiment and later referred to by their L^* (lightness) value relative to the monitor white point: 23.93, 38.01, 50.83, 61.99, 78.19, respectively). Participants repeated the adjustment three times at each luminance level. In addition, each participant repeated their matches to the dress body and lace as well as completing three illumination matches. All matching procedures followed the same protocol as the main experiment.

C.3.2 Results

CIELUV v^* values of the achromatic settings differed significantly across the different luminance levels (Figure C.1.A; Friedman test, $\chi^2(4) = 17.257$, $p = 0.002$). Achromatic settings became bluer as luminance increased, with significantly lower v^* values at a luminance setting of 34.46 cd/m^2 ($L^* = 50.83$) compared to 7.35 cd/m^2 ($L^* = 23.93$) (mean difference of 2.739, $p = 0.024$, with a Bonferroni correction). However, CIELUV u^* settings did not differ significantly (Figure C.1.B; Friedman test, $\chi^2(4) = 3.886$, $p = 0.422$).

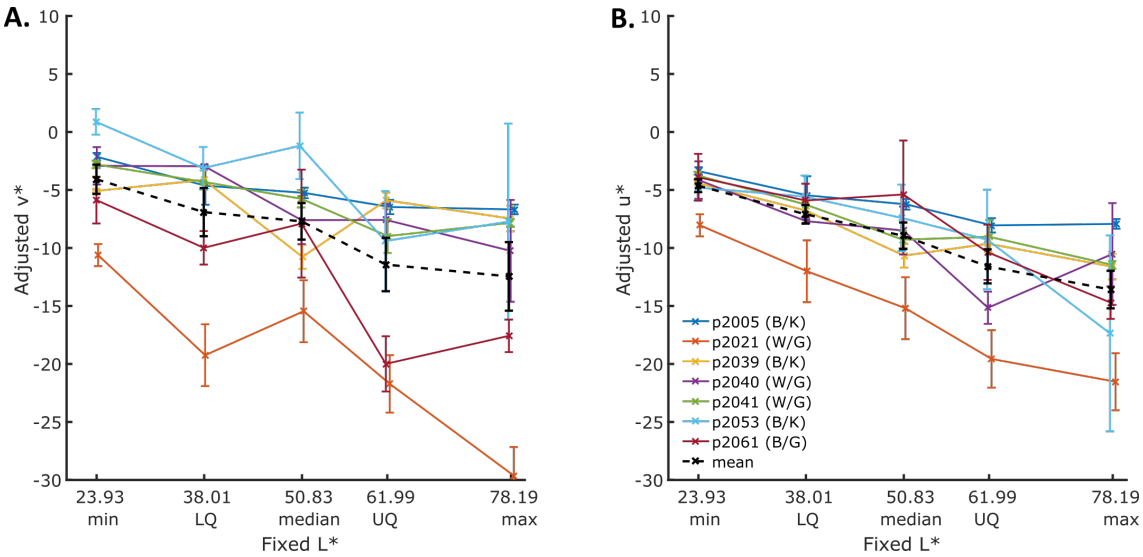


Figure C.1: The chromaticity of the achromatic settings in the control experiment. A. CIELUV u^* values of the achromatic settings across the different luminance levels, in L^* units. B. CIELUV v^* values of the achromatic settings across the different luminance levels. In both figures, each participants matches are plotted separately (solid coloured lines) as well as the average over all participants (black dashed line). The legend shows which participants took part in the control experiment and whether they were categorised in the blue and black (B/K), white and gold (W/G) or blue and gold (B/G) original dress colour groups.

The matches that participants gave to the dress body and lace did not differ significantly from their original matches along any dimension of CIELUV (Figures C.2.A and C.2.C; Wilcoxen Signed Ranks Tests, $p > 0.128$ in all cases). Mean illumination matches also did not differ from original mean illumination matches (Figures C.2.B and C.2.D; Wilcoxen Signed Ranks Tests, $p > 0.176$ in all cases).

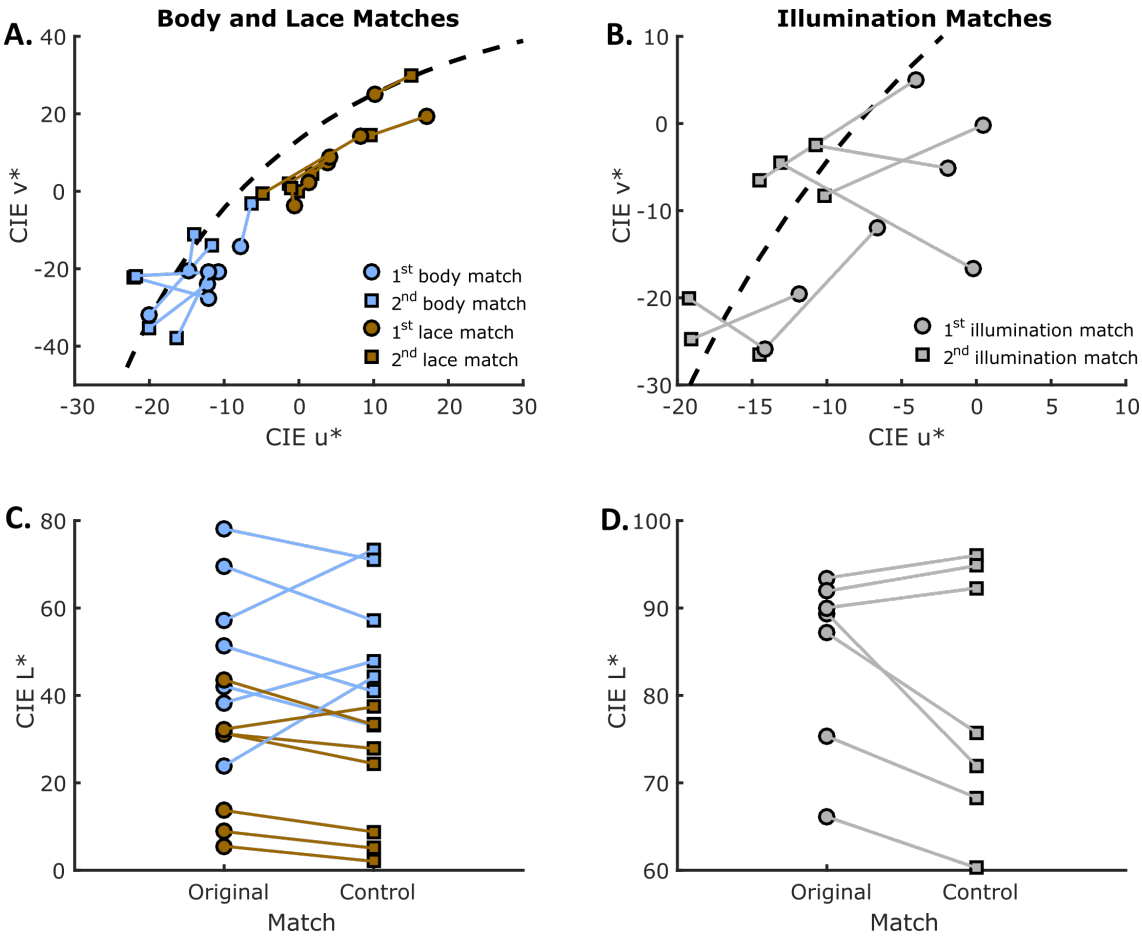


Figure C.2: Matches to the dress body, dress lace and illumination made during the control experiment. A. Original dress body (blue ∘) and lace (brown ∘) matches and those made during the control experiment (squares) plotted in the CIELUV chromaticity plane. B. Original illumination matches (∘) and illumination matches made during the control experiment (squares) plotted in the CIELUV chromaticity plane. C. Original dress body (blue ∘) and lace (brown ∘) match luminance settings and the luminance settings of those made during the control experiment (squares). D. Original illumination match luminance settings (∘) and luminance settings of those made during the control experiment (squares). In A and B, the black dashed line represents the Planckian locus.

Appendix D

Chapter 6 appendix

D.1 Experiment 1: Supplementary

The final value of each staircase was taken as the estimated PSE, with bias calculated as PSE hue value minus reference hue value. The blue staircases (starting at a hue value larger than that of the reference hue) and yellow staircases (starting at a hue value smaller than that of the reference hue) gave significantly different bias values (Wilcoxen signed-ranks test: $z = 5.048$, $p < .001$). The analysis below shows that conclusions are the same regardless of whether bias values are taken from blue staircases, yellow staircases, or are considered to be an average over the two (cf. the results in the main text).

Figures D.1.A and D.1.C show bias values calculated from the blue and yellow staircases in the different blocks of reference hues used. The overall pattern of results are the same as presented in the main text for both figures. However, the two sets of bias values differ as those calculated from yellow staircases are significantly lower on average than those form blue staircases (mean bias values of -0.89±5.54 and 0.68±5.40, respectively). For both staircase types, there is a main effect of direction of change on PSEs for the four hue values that are repeated between blocks (138, 144, 156 and 162 degrees; 2×4 repeated measures ANOVAs: $F(1, 10) = 77.95$, $p < .001$ for blue staircases and $F(1, 10) = 124.98$, $p < .001$ for yellow staircases).

Like in the main text, regression lines were fit to the bias values within each block. For bias values from both blue and yellow staircases all fitted regression slopes are negative (Figure D.1.B and D.1.D). In addition, the average slope values are significantly different from zero in all three blocks for bias values from both blue ($t(10) = -7.22$, $p < .001$, $t(10) = -7.05$, $p < .001$ and $t(10) = -6.10$, $p < .001$, all Bonferroni corrected) and yellow staircases ($t(10) = -7.12$, $p < .001$, $t(10) = -10.02$, $p < .001$ and $t(10) = -9.13$, $p < .001$, all Bonferroni corrected). Finally, a repeated measures ANOVAs with block

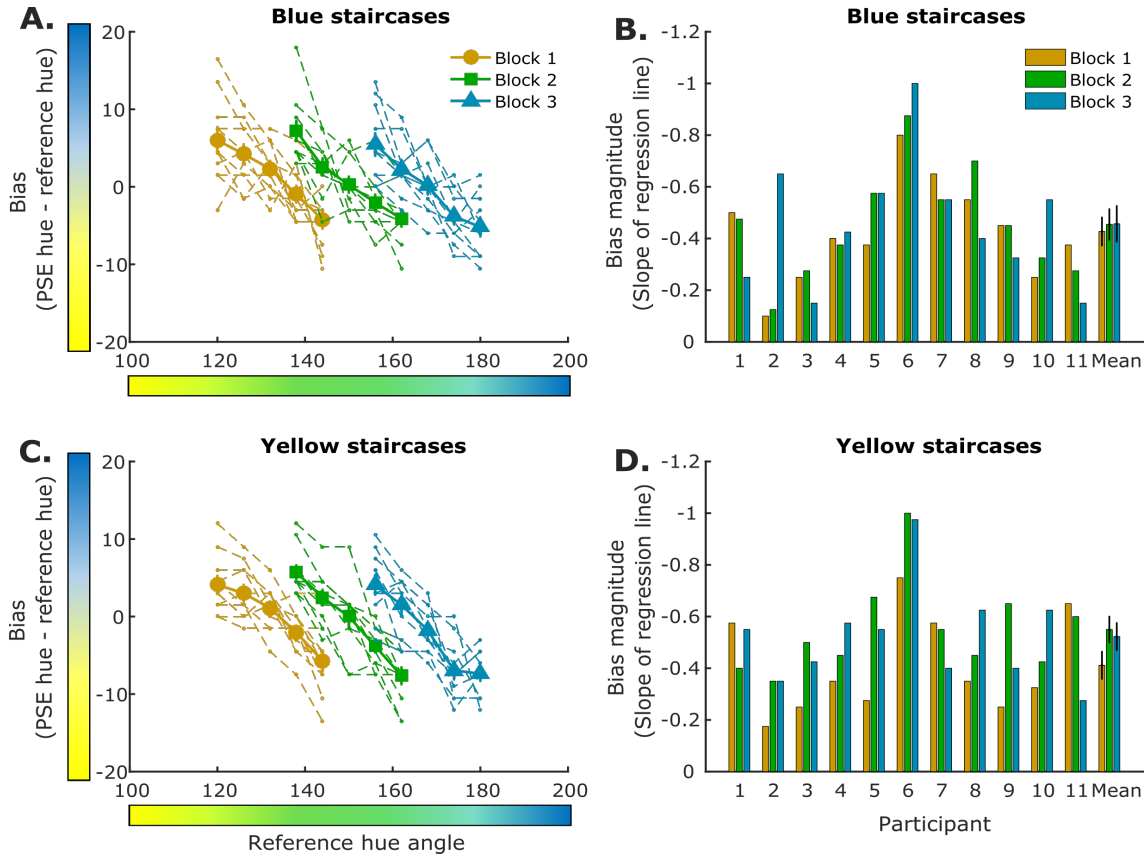


Figure D.1: Central tendency bias in experiment 1 for both blue and yellow staircases. A. The bias for each reference hue angle with respect to block (1-3) calculated from just the blue staircases. Bias is calculated as PSE hue angle minus reference hue angle in $CIE\ L^*a^*b^*$, with a positive bias indicating that the PSE for a particular reference was "bluer" than the true hue and a negative bias indicating that the PSE was "yellower". Bold lines with circle, square or triangular markers are the mean bias trend lines over all participants. Thin dashed lines are individual bias trend lines for each participant. B. The slopes of the regression lines fit to each individual bias trend line calculated from only blue staircases for each participant and the mean over these (error bars are $\pm$ 1 SEM). C. The bias for each reference hue angle calculated from just the yellow staircases. D. The slopes of the regression lines fit to each individual bias trend line calculated from only yellow staircases Note that as all slopes were negative, values on the y-axis have been flipped in B and D.

as the independent variable and fitted slope as the dependent variable shows that slope values did not differ across blocks for either blue staircase ($F(2, 20) = 0.14$, $p = .868$) or yellow staircase bias values ($F(2, 20) = 3.23$, $p = .061$).

D.2 Convergence of central tendency estimates for draws from a uniform distribution over five values

Figure D.2.A shows the estimated expected value of samples from a discrete uniform distribution over the values 1 to 5 with respect to sample size. We generated 10,000 sets of 100 samples from the distribution and the traces in Figure D.2.A show the cumulative estimates of the expected value. In Figure D.2.B we show the mean estimated expected value across the 10,000 sets for each number of samples along with the 95% confidence intervals. We have marked the values 2.5 and 3.5 on the plots for an easier comparison of the change in the confidence interval (and also because values in this interval would round to 3, the true expected value).

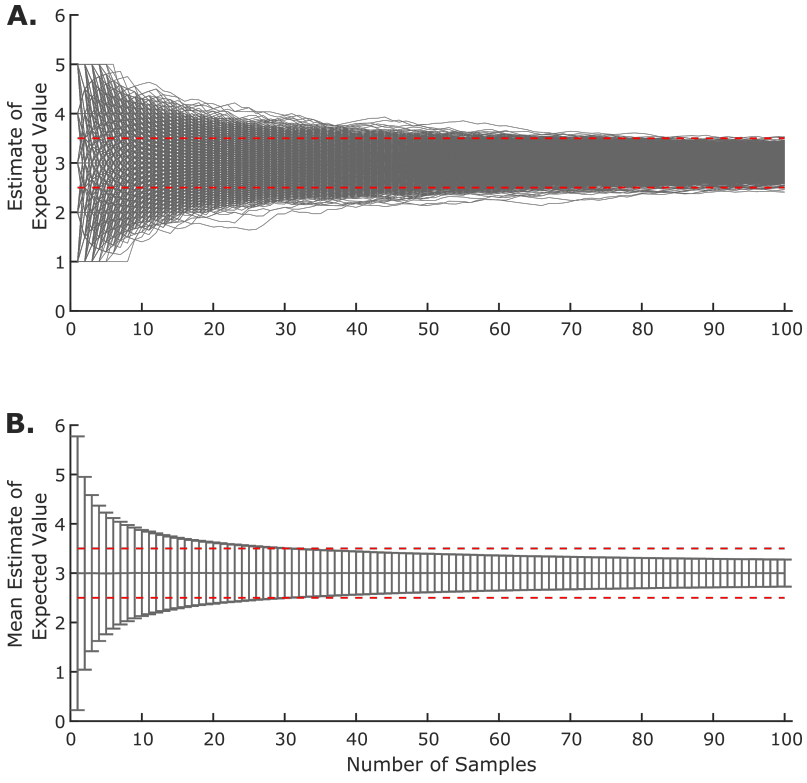


Figure D.2: Convergence of random samples from a discrete uniform distribution. Samples are from a discrete uniform distribution such that $X \sim DU(1, 5)$. A. The trace of the estimated expected value (the mean) for 10,000 sets of 100 samples (the cumulative mean). B. The mean estimated expect value over the 10,000 sets of sample. Error bars are 95% confidence intervals. Red dotted lines mark the values 2.5 and 3.5 in each plot.

D.3 Experiment 4 Questionnaire

Visual Memory and Serial Dependence in Colour Vision

Questionnaire

Thank you for completing our study. We would be very grateful if you could answer this quick questionnaire and return it to us. Many thanks, Colour Lab.

You were asked to do a task which involved remembering the colour of a reference light and adjusting the colour of a test light to match it.

1. Which of the following best describes how you remembered the reference light:

☐ I took a mental snapshot of the light and visualised the colour

☐ I used colour names to describe the light and remembered them

☐ I associated the colour with a familiar object and used that to keep it in mind

☐ Other, please explain:

2. Which of the following best describes how you made your adjustments to the test light:

☐ I gave the A and B buttons a colour name label and used these to help me know how to change the light

- How did you label the buttons?
- A – more Yellow
- B - more Blue

☐ I used A and B to adjust the colour of the test light to match a mental snapshot/visualised image of the reference light

☐ I used A and B to adjust the test light to something that I would call the same colour as I verbally labelled the reference light

☐ Other, please explain:

Figure D.3: The questionnaire used in Expeirment 4.

Bibliography

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

Adan, A., Archer, S. N., Hidalgo, M. P., Milia, L. D., Natale, V. & Randler, C. 2012 Circadian Typology : A Comprehensive Review. *Chronobiology international* **29** (9), 1153–1175.

Allred, S. R. 2012 Approaching Color with Bayesian Algorithms. In *Visual Experience: Sensation, Cognition, and Constancy*, 1st edn. (ed. G. Hatfield & S. R. Allred). Oxford University Press.

Alvaro, L., Linhares, J. M. M., Moreira, H., Lillo, J. & Nascimento, S. M. C. 2017 Robust colour constancy in red-green dichromats. *PloS one* **12** (6), 1–17.

Angueyra, J. M. & Rieke, F. 2013 Origin and effect of phototransduction noise in primate cone photoreceptors. *Nature Neuroscience* **16** (11), 1692–1700.

Arend, L. & Reeves, A. 1986 Simultaneous color constancy. *Journal of the Optical Society of America. A, Optics and image science* **3** (10), 1743–1751.

Arend, L. E., Reeves, A. & Goldstein, R. 1991 Simultaneous color constancy : papers with diverse Munsell values. *Journal of the Optical Society of America A* **8** (4), 661–672.

Ashourian, P. & Loewenstein, Y. 2011 Bayesian Inference Underlies the Contraction Bias in Delayed Comparison Tasks. *PloS one* **6** (5).

Barnard, K., Cardei, V. & Funt, B. 2002 A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data. *IEEE Transactions on Image Processing* **11** (9), 972–984.

Bejjanki, V. R., Knill, D. C. & Aslin, R. N. 2016 Learning and inference using complex generative models in a spatial localization task. *Journal of Vision* **16** (2016), 1–13.

Brainard, D. & Hurlbert, A. 2015 Colour Vision: Understanding #TheDress. *Current Biology* **25** (13), R551–R554.

Brainard, D. H. 1996 Part IV: Cone contrast and opponent modulation color spaces. In *Human Color Vision*, 2nd edn. (ed. P. Kaiser & R. M. Boynton), pp. 563–579. Optical Society of America.

Brainard, D. H. 1997 The Psychophysics Toolbox. In *Spatial Vision*, , vol. 10, pp. 10:433–436.

Brainard, D. H. 1998 Color constancy in the nearly natural image. 2. Achromatic loci. *Journal of the Optical Society of America. A, Optics, image science, and vision* **15** (2), 307–325.

Brainard, D. H. 2009 Bayesian Approaches To Color Vision. In *The Cognitive Neurosciences IV*. MIT Press, Cambridge, MA.

Brainard, D. H., Brunt, W. a. & Speigle, J. M. 1997 Color constancy in the nearly natural image. I. Asymmetric matches. *Journal of the Optical Society of America. A, Optics, image science, and vision* **14** (9), 2091–2110.

Brainard, D. H. & Freeman, W. T. 1997 Bayesian color constancy. *Journal of the Optical Society of America. A, Optics, image science, and vision* **14** (7), 1393–1411.

Brainard, D. H., Jiang, H., Cottaris, N. P., Rieke, F., Chichilnisky, E., Farrell, Joyce, E. & Wandell, B. A. 2015 ISETBIO: Computational tools for modeling early human vision. In *Imaging and Applied Optics*. OSA Technical Digest.

Brainard, D. H., Longère, P., Delahunt, P. B., Freeman, W. T., Kraft, J. M. & Xiao, B. 2006 Bayesian model of human color constancy. *Journal of vision* **6** (11), 1267–81.

Brainard, D. H. & Maloney, L. T. 2011 Surface color perception and equivalent illumination models. *Journal of vision* **11** (2011), 1–18.

Brainard, D. H. & Radonjić, A. 2014 Color constancy. *The New Visual Neurosciences* **1**, 545–556.

Brainard, D. H. & Stockman, A. 2009 Colorimetry. In *Handbook of Optics, Volume III, Vision and Optics*, 3rd edn. McGraw Hill: New York.

Bramwell, D. I. & Hurlbert, A. C. 1996 Measurements of colour constancy by using a forced-choice matching technique. *Perception* **25**, 229–241.

Brill, M. & West, G. 1981 Contributions to the theory of invariance of color under the condition of varying illumination. *Journal of Mathematical Biology* **11** (3), 337–350.

Brown, A. M., Lindsey, D. T. & Guckes, K. M. 2011 Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *Journal of Vision* **11** (12), 2.

Buchsbaum, G. & Gottschalk, A. 1983 Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal Society of London. Series B, Biological Sciences* **220** (1218), 89–113.

Chauhan, T., Perales, E., Hird, E. & Wuerger, S. 2014 The achromatic locus : Effect of navigation direction in color space. *Journal of vision* **14** (2014), 1–11.

Chetverikov, A. & Ivanchei, I. 2016 Seeing "the Dress" in the Right Light: Perceived Colors and Inferred Light Sources. *Perception* **0** (0), 1–21.

Churma, M. E. 1994 Blue shadows: physical, physiological, and psychological causes. *Applied optics* **33** (21), 4719–4722.

Cicchini, G. M., Anobile, G. & Burr, D. C. 2014 Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *PNAS* **111** (21).

CIE 2004 *Colorimetry*, 3rd edn. Bureau Central de la CIE, Paris.

Corbett, J. E., Fischer, J. & Whitney, D. 2011 Facilitating Stable Representations : Serial Dependence in Vision. *PloS one* **6** (1).

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.

Craven, B. J. & Foster, D. H. 1992 An operational approach to colour constancy. *Vision research* **32** (7), 1359–1366.

Crick, F. & Koch, C. 1995 Are we aware of neural activity in primary visual cortex? *Nature* **375**, 121–123.

De Valois, R. L., Abramov, I. & Jacobs, G. H. 1966 Analysis of Response Patterns of LGN Cells. *Journal of the Optical Society of America* **56** (7), 966–977.

Delahunt, P. B. & Brainard, D. H. 2004 Does human color constancy incorporate the statistical regularity of natural daylight ? *Journal of Vision* (4), 57–81.

Derrington, A., Krauskopf, J. & Lennie, P. 1984 Chromatic mechanisms in lateral geniculate nucleus of macaque. *Journal of physiology* (357), 241–265.

Dicarlo, J. J. & Cox, D. D. 2007 Untangling invariant object recognition. *Trends in cognitive sciences* **11** (8).

Duffy, S., Huttenlocher, J., Hedges, L. V. & Crawford, L. E. 2010 Category effects on stimulus estimation: shifting and skewed frequency distributions. *Psychonomic Bulletin and Review* **17** (2), 224–230.

Ernst, M. O. & Banks, M. S. 2002 Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415** (6870), 429–433.

Eskew, R. T. 2009 Higher order color mechanisms: A critical review. *Vision Research* **49** (22), 2686–2704.

Evans, R. M. 1951 Method for correcting photographic color prints.

Fairchild, M. D. 2005 Human Color Vision. In *Color Appearance Models*, 2nd edn., , vol. 1, pp. 1–34. Wiley.

Fine, I. & Jacobs, R. A. 2002 Comparing perceptual learning tasks: a review. *Journal of Vision* **2** (2), 190–203.

Finlayson, G., Mackiewicz, M., Hurlbert, A., Pearce, B. & Crichton, S. 2014 On calculating metamer sets for spectrally tunable LED illuminators. *Journal of the Optical Society of America. A, Optics, image science, and vision* **31** (7), 1577–1587.

Fischer, J. & Whitney, D. 2014 Serial dependence in visual perception. *Nature Publishing Group* **17** (5), 738–743.

Forsyth, D. A. 1990 A Novel Algorithm for Color Constancy. *International Journal of Computer Vision* **5** (1), 5–36.

Foster, D. H. 2003 Does colour constancy exist? *Trends in Cognitive Sciences* **7** (10), 439–443.

Foster, D. H. 2011 Color constancy. *Vision research* **51** (7), 674–700.

Foster, D. H., Amano, K. & Nascimento, S. M. 2003 Tritanopic colour constancy under daylight changes? In *Normal & Defective Colour Vision* (ed. J. D. Mollon), pp. 218–224. Oxford: Oxford University Press.

Foster, D. H., Amano, K. & Nascimento, S. M. C. 2015 Time-lapse ratios of cone excitations in natural scenes. *Vision Research* **120** (Supplement C), 45–60.

Foster, D. H. & Nascimento, S. M. C. 1994 Relational Colour Constancy from Invariant Cone-Excitation Ratios. *Proceedings: Biological Sciences* **257** (1349), 115–121.

Gegenfurtner, K. R., Bloj, M. & Toscani, M. 2015 The many colours of 'the dress'. *Current Biology* **25** (13), R543–R544.

Gegenfurtner, K. R. & Kiper, D. C. 2003 Color vision. *Annual review of neuroscience* **26**, 181–206.

Gegenfurtner, K. R. & Rieger, J. 2000 Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology* **10** (13), 805–808.

Geisler, W. S. 2011 Contributions of ideal observer theory to vision research. *Vision Research* **51** (7), 771–781.

Giesel, M. & Gegenfurtner, K. R. 2010 Color appearance of real objects varying in material, hue, and shape. *Journal of Vison* **10** (9), 1–21.

Hansen, T. & Gegenfurtner, K. R. 2013 Higher order color mechanisms: Evidence from noise-masking experiments in cone contrast space. *Journal of Vision* **13** (1), 26.1–21.

Hansen, T., Olkkonen, M., Walter, S. & Gegenfurtner, K. R. 2006 Memory modulates color appearance. *Nature neuroscience* **9** (11), 1367–8.

Hatfield, G. & Allred, S. 2012 *Visual Experience: Sensation, Cognition, and Constancy*. Oxford University Press.

Hellström, Å. 1985 The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin* **97** (1), 35–61.

Helson, H. 1964 *Adaptation-level theory: an experimental and systematic approach to behaviour*. Harper & Row.

Hernández-Andrés, J., Romero, J., Nieves, J. L. & Lee, R. L. 2001 Color and spectral analysis of daylight in southern Europe. *Journal of the Optical Society of America. A, Optics, image science, and vision* **18** (6), 1325–35.

Hofer, H., Carroll, J., Neitz, J., Neitz, M. & Williams, D. R. 2005 Organization of the Human Trichromatic Cone Mosaic. *The Journal of Neuroscience* **25** (42), 9669 LP – 9679.

Hohwy, J. 2013 *The Predictive Mind*. Oxford University Press UK.

Hollingworth, H. L. 1910 The Central Tendency of Judgment. *The Journal of Philosophy, Psychology and Scientific Methods* **7** (17), 461–469.

Horne, J. A. & Ostberg, O. 1976 A Self-Assessment Questionnaire to Determine Morningness-Eveningness in Human Circadian Rhythms. *International Journal of Chronobiology* **4**, 97–110.

Hubert-Wallander, B. & Boynton, G. M. 2015 Not all summary statistics are made equal : Evidence from extracting summaries across time. *Journal of Vision* **15** (4), 1–12.

Hunt, R. W. G. 1998 *Measuring Colour*, 3rd edn. Fountain Press.

Hurlbert, A. C. 1998 Computational models of color constancy. In *Perceptual Constancy: Why things look as they do*, 1st edn. (ed. V. Walsh & J. Kulikowski), pp. 283–232. Cambridge University Press.

Huttenlocher, J., Hedges, L. V. & Vevea, J. L. 2000 Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General* **129** (2), 220–241.

Jaaskelainen, T., Parkkinen, J. & Toyooka, S. 1990 Vector-subspace model for color representation. *Journal of the Optical Society of America* **7** (4), 725–730.

Jamieson, D. G. 1977 Two presentation order effects. *Canadian Journal of Psychology* **31** (4), 184–194.

Jamieson, D. G. & Petrusic, W. M. 1975 The dependence of time-order error direction on stimulus range. *Canadian Journal of Psychology* **29** (3), 175–182.

Jazayeri, M. & Shadlen, M. N. 2010 Temporal context calibrates interval timing. *Nature Neuroscience* **13** (8), 1020–1026.

Jordan, G., Deeb, S. S., Bosten, J. M. & Mollon, J. D. 2010 The dimensionality of color vision in carriers of anomalous trichromacy. *Journal of vision* **10** (2010), 1–19.

Judd, D. B., MacAdam, D. L. & Wyszecki, G. 1964 Spectral Distribution of Typical Daylight as a Function of Correlated Color Temperature. *Journal of the Optical Society of America* **54** (8), 1031–1040.

Kaernbach, C. 1991 Simple adaptive testing with the weighted up-down method. *Perception & psychophysics* **49** (3), 227–9.

Kingdom, F. & Prins, N. 2010 *Psychophysics: A Practical Introduction*. Elsevier Ltd.

Kleiner, M., Brainard, D. H. & Pelli, D. 2007 What's new in Psychtoolbox-3?

Knill, D. C. & Pouget, A. 2004 The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences* **27** (12), 712–719.

Kondo, A., Takahashi, K. & Watanabe, K. 2012 Sequential effects in face-attractiveness judgment. *Perception* **41** (1), 43–49.

Krauskopf, J. 1980 Discrimination and detection of changes in luminance. *Vision Research* **20** (8), 671–677.

Krauskopf, J., Williams, D. R. & Heeley, D. W. 1982 Cardinal Directions of Colour Space. *Vision research* **22**, 1123–1131.

Krauskopf, J., Williams, D. R., Mandler, M. B. & Brown, A. M. 1986 Higher order color mechanisms. *Vision Research* **26** (1), 23–32.

von Kries, J. 1878 Beitrag zur physiologie der gesichtsempfindung. *Arch. Anat. Physiol.* **2**, 505–524.

Kuriki, I. 2015 Lightness dependence of achromatic loci in color-appearance coordinates. *Frontiers in Psychology* **6** (67), 1–10.

Lafer-Sousa, R., Hermann, K. L. & Conway, B. R. 2015 Striking individual differences in color perception uncovered by 'the dress' photograph. *Current Biology* **25** (13), R545–R546.

Land, E. H. & McCann, J. J. 1971 Lightness and Retinex Theory. *Journal of the Optical Society of America* **61** (1), 1–11.

Laquitaine, S. & Gardner, J. L. 2017 A Switching Observer for Human Perceptual Estimation. *Neuron* pp. 1–13.

Li, C. & Luo, M. R. 2001 The Estimation of Spectral Reflectances Using the Smoothness Constraint Condition. *Proceedings of the IS&T/SID Ninth Color Imaging Conference* pp. 62–67.

Liberman, A., Fischer, J. & Whitney, D. 2014 Serial dependence in the perception of faces. *Current Biology* **24** (21), 2569–2574.

MacAdam, D. L. 1942 Visual Sensitivities to Color Differences in Daylight*. *Journal of the Optical Society of America* **32** (5), 247–274.

Maloney, L. T. 1986 Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America A* **3** (10), 1673.

Maloney, L. T. 1999 Physics-Based Approaches to Modeling Surface Color Perception. In *Color Vision: From Genes to Perception* (ed. K. R. Gegenfurtner & L. T. Sharpe), pp. 387–422. Cambridge University Press.

Maloney, L. T. & Wandell, B. a. 1986 Color constancy: a method for recovering surface spectral reflectance. *Journal of the Optical Society of America. A, Optics and image science* **3** (1), 29–33.

Maloney, L. T. & Yang, J. N. 2003 The Illuminant Estimation Hypothesis and Surface Colour Perception. In *Colour: Connecting the Mind to the Physical World* (ed. R. Mausfeld & D. Heyer), pp. 335–358. Oxford University Press.

Mamassian, P., Landy, M. & Maloney, L. T. 2003 Bayesian Modelling of Visual Perception. In *Probabalistic Models of the Brain: Perception and Neural Function* (ed. R. P. N. Rao, B. A. Olshausen & M. S. Lewicki), pp. 13–36. MIT Press.

Marimont, D. H. & Wandell, B. A. 1992 Linear models of surface and illuminant spectra. *Journal of the Optical Society of America* **9** (11), 1905–1913.

Marimont, D. H. & Wandell, B. A. 1994 Matching color images: the effects of axial chromatic aberration. *Journal of the Optical Society of America A* **11** (12), 3113–3122.

Mattar, M. G., Kahn, D. A., Thompson-Schill, S. L. & Aguirre, G. K. 2016 Varying Timescales of Stimulus Integration Unite Neural Adaptation and Prototype Formation. *Current Biology* **26** (13), 1669–1676.

McKenna, F. P. 1984 Assimilation and contrast in perceptual judgments. *The Quarterly journal of experimental psychology. A, Human experimental psychology* **36** (3), 531–548.

Mollon, J. D. 1989 Tho' she kneel'd in that place where they grew... The uses and origins of primate colour vision. *Journal of Experimental Biology* **146** (1), 21 LP – 38.

Nascimento, S. M. C., Amano, K. & Foster, D. H. 2016 Spatial distributions of local illumination color in natural scenes. *Vision Research* **120** (Supplement C), 39–44.

Nascimento, S. M. C., Ferreira, F. P. & Foster, D. H. 2002 Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America A* **19** (8), 1484–1490.

Olkkonen, M. & Allred, S. R. 2014 Short-Term Memory Affects Color Perception in Context. *PlosOne* **9** (1), 1–11.

Olkkonen, M., McCarthy, P. F. & Allred, S. R. 2014 The central tendency bias in color perception : Effects of internal and external noise. *Journal of Vision* **14** (11), 1–15.

Pearce, B., Crichton, S., Mackiewicz, M., Finlayson, G. D. & Hurlbert, A. 2014 Chromatic Illumination Discrimination Ability Reveals that Human Colour Constancy Is Optimised for Blue Daylight Illuminations. *PLoS ONE* **9** (2), e87989.

Pelli, D. G. 1997 The VideoToolbox software for visual psychophysics: Transforming numbers into movies. In *Spatial Vision*, pp. 10:437–442.

Poirson, A. B. & Wandell, B. A. 1993 Appearance of colored patterns : pattern-color separability. *Journal of the Optical Society of America* **10** (12), 2458–2470.

Pokorny, J., Smith, V. C. & Lutze, M. 1987 Aging of the human lens. *Applied optics* **26** (8), 1437–1440.

Pouget, A., Beck, J. M., Ma, W. J. & Latham, P. E. 2013 Probabilistic brains : knowns and unknowns. *Nature neuroscience* **16** (9).

Radonjić, A. & Brainard, D. H. 2016 The Nature of Instructional Effects in Color Constancy. *Journal of Experimental Psychology: Human Perception and Performance* **42** (6), 847–865.

Radonjić, A., Cottaris, N. P. & Brainard, D. H. 2016*a* Color constancy in a naturalistic, goal-directed task. *Journal of Vision* **15** (2015), 1–21.

Radonjić, A., Pearce, B., Aston, S., Krieger, A., Cottaris, N. P., Brainard, D. H. & Hurlbert, A. C. 2016*b* Illumination discrimination in real and simulated scenes. *Journal of Vision* **16**, 1–18.

Rieke, F. 2014 Model for primate cone responses. *ISETBIO DOCS* **1** (1), 1–3.

Riskey, D. R., Parducci, A. & Beauchamp, G. K. 1979 Effects of context in judgments of sweetness and pleasantness. *Perception & Psychophysics* **26** (3), 171–176.

Roberson, D., Damjanoviv, L. & Pilling, M. 2007 Categorical perception of facial expressions : Evidence for a category adjustment model. *Memory and Cognition* **35** (7), 1814–1829.

Ruderman, D. L., Cronin, T. W. & Chiao, C.-C. 1998 Statistics of cone responses to natural images: implications for visual coding. *Journal of the Optical Society of America A* **15** (8), 2036–2045.

Rushton, W. A. H. 1972 Review Lecture. Pigments and signals in colour vision. *The Journal of Physiology* **220** (3), 1–31.

Rust, N. C. & Stocker, A. A. 2010 Ambiguity and invariance : two fundamental challenges for visual processing. *Current Opinion in Neurobiology* **20**, 382–388.

Rüttiger, L., Mayser, H., Sérey, L. & Sharpe, L. T. 2001 The color constancy of the red-green color blind. *Color Research & Application* **26** (S1), S209–S213.

Ryan, L. J. 2011 Temporal context affects duration reproduction reproduction. *Journal of Cognitive Psychology* **23** (1), 157–170.

Schifferstein, H. N. 1995 Contextual effects in difference judgments. *Perception & psychophysics* **57** (1), 56–70.

Shepard, T. G., Swanson, E. A., McCarthy, C. L., Eskew, R. T. & Eskew Jr. ,Rhea T. 2016 A model of selective masking in chromatic detection. *Journal of Vision* **16** (9), 3.

Smithson, H. E. 2005 Sensory, computational and cognitive components of human colour constancy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **360** (1458), 1329–1346.

Soo, F. S., Detwiler, P. B. & Rieke, F. 2008 Light Adaptation in Salamander L-Cone Photoreceptors. *Journal of Neuroscience* **28** (6), 1331–1342.

Spence, I., Wong, P., Rusan, M. & Rastegar, N. 2006 How Color Enhances Visual Memory for Natural Scenes. *Psychological Science* **17** (1), 1–6.

Spitschan, M., Aguirre, G. K., Brainard, D. H. & Sweeney, A. M. 2016 Variation of outdoor illumination as a function of solar elevation and light pollution. *Nature Publishing Group* pp. 1–13.

Stocker, A. A. & Simoncelli, E. P. 2006 Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience* **9**, 578.

Stockman, A. & Sharpe, L. T. 2000 The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research* **40** (13), 1711–1737.

Tassinari, H., Hudson, T. E. & Landy, M. S. 2006 Combining Priors and Noisy Visual Cues in a Rapid Pointing Task. *Journal of Neuroscience* **26** (40), 10154–10163.

Taubert, J., Alais, D. & Burr, D. 2016*a* Different coding strategies for the perception of stable and changeable facial attributes. *Nature: Scientific Reports* **6**, 2–8.

Taubert, J., Van der Burg, E. & Alais, D. 2016*b* Love at second sight: Sequential dependence of facial attractiveness in an on-line dating paradigm. *Nature: Scientific Reports* **6** (March), 22740.

Toscani, M., Gegenfurtner, K. R. & Doerschner, K. 2017 Differences in illumination estimation in #thedress. *Journal of vision* **17** (1), 22.

Troost, J. M. & Weert, C. M. M. D. E. 1991 Naming versus matching in color constancy. *Perception & psychophysicsn* **50** (6), 591–602.

Uchikawa, K., Fukuda, K., Kitazawa, Y. & Macleod, D. I. A. 2012 Estimating illuminant color based on luminance balance of surfaces. *Journal of the Optical Society of America* **29** (2), 133–143.

Uchikawa, K., Morimoto, T. & Matsumoto, T. 2016 Prediction for individual differences in appearance of the dress by the optimal color hypothesis. *Journal of Vision* **16** (12), 745.

Wallisch, P. 2017 Illumination assumptions account for individual differences in the perceptual interpretation of a profoundly ambiguous stimulus in the color domain: The dress. *Journal of Vision* **17** (4), 5.

Webb, A. R. 2006 Considerations for lighting in the built environment: Non-visual effects of light. *Energy and Buildings* **38** (7), 721–727.

Webster, M. A. & Mollon, J. D. 1991 Changes in colour appearance following post-receptoral adaptation. *Nature* **349** (6306), 235–238.

Westland, S. 2002 Functional Colour Vision. In *Signals and Perception: The Fundamentals of Human Sensation*, pp. 133–146. Palgrave Macmillan.

Wiesel, T. N. & Hubel, D. H. 1966 Spatial and chromatic interactions in the lateral geniculate body of the rhesus monkey. *Journal of Neurophysiology* **29** (6), 1115 LP – 1156.

Winkler, A. D., Spillmann, L., Werner, J. S. & Webster, M. A. 2015 Asymmetries in blue-yellow color perception and in the color of 'the dress'. *Current Biology* **25** (13), R547–R548.

Witzel, C., Racey, C. & O'Regan, J. 2016 Perceived colors of the color-switching dress depend on implicit assumptions about the illumination. *Journal of Vision* **16** (12), 223.

Witzel, C., Valkova, H., Hansen, T. & Gegenfurtner, K. R. 2011 Object knowledge modulates colour appearance. *i-Perception* **2**, 13–49.

Wyszecki, G. & Stiles, W. 1967 *Color Science*. John Wiley & Sons, Inc.

Xia, Y., Leib, A. Y. & Whitney, D. 2016 Serial dependence in the perception of attractiveness. *Journal of Vision* **16** (15), 1–8.

Zaidi, Q. 1997 Decorrelation of L- and M-cone signals. *Journal of the Optical Society of America A* **14** (12), 3430–3431.

Zaidi, Q. & Shapiro, A. G. 1993 Adaptive orthogonalization of opponent-color signals. *Biological Cybernetics* **69** (5), 415–428.